



THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

Research Commons

<http://researchcommons.waikato.ac.nz/>

Research Commons at the University of Waikato

Copyright Statement:

The digital copy of this thesis is protected by the Copyright Act 1994 (New Zealand).

The thesis may be consulted by you, provided you comply with the provisions of the Act and the following conditions of use:

- Any use you make of these documents or images must be for research or private study purposes only, and you may not make them available to any other person.
- Authors control the copyright of their thesis. You will recognise the author's right to be identified as the author of the thesis, and due acknowledgement will be made to the author where appropriate.
- You will obtain the author's permission before publishing any material from the thesis.

Structural and enzymatic characterisation of
nucleoside triphosphate diphosphohydrolases
from *Trifolium repens* and *Dolichos biflorus*

A thesis submitted in fulfilment
of the requirements for the degree of

Doctor of Philosophy in Biological Sciences

at

The University of Waikato

by

Mathew Hoani Cumming

2011



THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

This thesis is dedicated to our little boy Anaru.

You gave us so much pleasure

Abstract

Extracellular nucleoside triphosphate diphosphohydrolases (NTPDases) are enzymes that reduce the extracellular nucleotide signal and inactivate the purinogenic signalling pathway. These enzymes, in the presence of a divalent cation sequentially hydrolyses the γ - and β - phosphoanhydride bond from a range of nucleotide di- or triphosphates to the corresponding nucleotide monophosphate.

There is evidence that purinogenic signalling is present in higher plants and it is becoming clear that NTPDases play a role in the early stages of rhizobium infection during nodulation in legumes. To further our understanding of NTPDases, this study investigated the biochemical and structural characteristics of two legume NTPDases believed to be involved in nodulation. The first, 7WC, was isolated from the roots of white clover and the second, DbLnP, from the roots of *Dolichos biflorus*. DbLnP has been characterised as a carbohydrate binding NTPDase that is directly associated with the perception of rhizobia.

Using X-ray crystallography a number of crystal structures of 7WC and DbLnP were determined at resolutions between 1.9 Å and 2.9 Å. For 7WC, structures were determined for an apo- form, an AMP-bound and also bound with the nonhydrolysable ATP analogue AMPPNP. For DbLnP structures were solved with phosphate and Mn^{2+} bound and another with AMPPNP and Mn^{2+} bound.

Kinetic analysis of a range of substrates together with the analysis of the binding modes of 7WC and DbLnP explained substrate preference for each of the NTPDases.

These analyses showed that NTPDases can adopt two conformations depending on substrate and co-factor binding. The central hinge region creates a 'butterfly' motion of the domains that reduces the width of the active site cleft. This phenomenon has been previously hypothesised but has not been observed for NTPDases.

A model of catalysis is proposed whereby the 'open' form first binds substrate in an inactive orientation. Binding of the metal ion induces a conformational change that brings the domains together and allows movement of the phosphate tail deeper into the active site cleft – a reorganisation that is required for catalysis. Hydrolysis occurs via nucleophilic attack on the terminal phosphate. Finally, release of the metal ion allows the 'open' conformation to be restored for subsequent catalysis to occur.

Acknowledgments

There are numerous people that without their love and support this study would not be possible. Firstly, my work colleagues: Ali, Emma L, Emma S, Ian, Jo H, Jo M, John, Judith, Line and of course Marisa. We really did this together. Your consistent friendship made my time in Hamilton a real pleasure. And to all the Agresearch team, particularly Richard, and ChungHong, thanks you for the help and kindness.

Svend, I had to give you your own paragraph as you were amazing. You got me through some really tough times and without your love and support I would have not have gotten through.

To my co-supervisor Nick Roberts; it was your belief in me that got me into this mess in the first place and it's your belief in me that helped me get where I am. I whole heartily appreciate this. Your friendship and guidance I helped me indefinitely.

Many thanks to my supervisor Vic. thank you for your guidance throughout my research (even though I didn't always listen) and for making it possible that I could work with you. You have made an independent researcher with the tools to take on the world.

To Mum, Dad and whanau, thank you so much for you support in whatever I wanted to do throughout without doubt or judgment. It is this that has allowed me to follow these paths without constraints. I love you all.

Lastly, Sacha, you have had to endure this "Damn PhD!!!" the most. I don't know how you did it (you should be getting an award). I will never forget the never-ending understanding and support that you have given me over the years and for that I am truly in debt to you. I love you so very very much.

Aroha nui to you all.

Table of Contents

Abstract	iii
Acknowledgments.....	v
Table of Contents.....	vi
List of Figures	xvi
List of Tables	xxi
List of Equations	xxii
List of Abbreviations	xxiii
1. Introduction.....	1
1.1 NTPDases.....	1
1.2 Physiological mechanisms of NTPDases.	1
1.2.1 Purinogenic signalling and physiological impact of NTPDases in plants 1	
1.2.1.1 Physiological role of NTPDases in legumes.....	4
1.2.2 Purinogenic signalling and physiological impact of NTPDases in mammalian systems.....	5
1.2.3 NTPDases pathogenic organisms	6
1.3 The primary sequence of NTPDases	7
1.3.1 The apyrase conserved regions (ACR)	7
1.3.1.1 Membrane and soluble NTPDases.....	9
1.3.2 Disulfide bonds within NTPDases	10
1.4 The NTPDase structure	10
1.4.1 Early structural determination.....	10
1.4.2 Overall published structures of NTPDases.....	11
1.4.3 Substrate binding mechanism.....	12
1.4.4 Trans-domain movement upon substrate binding in exopolyphosphatase and actin.....	14

1.5	NTPDase kinetics.....	16
1.5.1	Early work on the catalytic function.....	16
1.5.2	Catalytic mechanism of NTPDase.....	16
1.5.3	Mechanism of hydrolysis.....	17
1.5.4	Residues involved in catalysis-site directed mutagenesis.....	19
1.5.5	NTPDase activity-Substrate specificity	22
1.5.6	Michaelis–Menten kinetics of NTPDases	24
1.6	Research objectives	26
2	General methods.....	27
2.1	Reagents.....	27
2.2	Bioinformatic analysis	27
2.3	DNA manipulation.....	27
2.3.1	Gene constructs	27
2.3.2	DNA plasmid extraction.....	27
2.3.3	DNA electrophoresis gel.....	27
2.3.4	DNA fragment gel extraction.....	28
2.3.5	DNA quantification.....	28
2.3.6	Restriction enzyme digest of DNA plasmid.....	28
2.3.7	DNA ligation	28
2.3.8	DNA sequencing	29
2.3.9	DNA transformation.....	29
2.3.10	<i>E. coli</i> storage stocks.....	29
2.4	Protein expression.....	29
2.4.1	Growth of cells for protein expression	29
2.4.2	Harvesting the cells of test cultures	30

2.4.3	Harvesting of large scale expression culture	30
2.5	Large scale purification using fast liquid chromatography (FPLC).....	30
2.5.1	IMAC in denaturing conditions.....	30
2.5.2	Size exclusion chromatography	31
2.5.3	Cation exchange	31
2.6	Refolding of denatured protein	31
2.6.1	Isolation and preparation of inclusion bodies	31
2.6.2	High throughput method for refolding.....	32
2.6.3	Small scale method for refolding	33
2.6.4	Large scale refolding using the dilution method.....	33
2.7	Protein analysis.....	33
2.7.1	Native-PAGE.....	33
2.7.2	SDS-PAGE	34
2.7.3	Staining PAGE gels	35
2.7.3.1	Coomassie stain.....	35
2.7.3.2	Silver stain.....	35
2.7.4	Protein concentration measurement.....	35
2.7.4.1	Nanodrop method.....	35
2.7.4.2	Bradford assay	36
2.7.5	Concentrating protein.....	36
2.7.5.1	Using protein spin concentrators	36
2.7.5.2	Using stir-cell concentrator.....	36
2.7.6	Buffer exchange via dialysis tubing.....	37
2.7.7	Protein activity analysis.....	37
2.7.7.1	Protein activity analysis for calcium saturation curve.	37

2.7.7.2	Protein activity analysis for optimal pH.....	38
2.7.7.3	Determination of the initial rate.....	38
2.7.7.4	Michaelis-Menten kinetic analysis.....	38
2.7.8	In-gel activity assay.....	39
2.7.9	Small scale thrombin cleavage trials of DbLnP.....	39
2.7.10	Large scale thrombin cleavage.....	39
2.7.10.1	Purification of thrombin cleaved DbLnP.....	39
2.7.10.2	Site directed mutagenesis.....	40
2.8	Protein crystallography.....	40
2.8.1	General methods.....	40
2.8.2	Initial crystallisation trials-robot.....	40
2.8.3	Optimisation of crystallisation conditions.....	41
2.8.3.1	Hanging drop method for crystallography.....	41
2.8.3.2	Sitting drop method for crystallography.....	41
2.8.4	Seeding.....	41
2.8.4.1	Micro-seeding.....	41
2.8.4.2	Streak-seeding.....	42
2.8.4.3	Macro-seeding.....	42
2.8.5	Additive screening.....	42
2.8.6	Izit dye.....	42
2.8.7	Co-crystallisation with substrate.....	42
2.8.8	Substrate soaks.....	43
2.9	Preparation for data collection.....	43
2.9.1	Cryo-protectants.....	43
2.9.2	Flash freezing.....	43

2.10	Diffraction data collection.....	43
2.10.1	Home source data collection	43
2.10.2	Synchrotron data collection.....	44
2.11	X-ray data processing.....	44
2.11.1	Indexing with MOSFLM.....	44
2.11.2	Beam centre estimation.....	44
2.11.3	Scaling and merging of reflections	44
2.11.4	Analysis of unit cell using Matthew's coefficient	45
2.11.5	Phase determination by molecular replacement	45
2.11.6	Model building and refinement.....	45
2.11.6.1	Automated building.....	45
2.11.7	Manual building.....	45
2.11.7.1	Model refinement.....	45
2.11.8	Structural analysis.....	46
2.11.8.1	B-average analysis.....	46
2.11.8.2	Ramachandran analysis.....	46
2.11.8.3	TLS analysis.....	46
2.11.8.4	Secondary structure comparison.....	47
2.11.8.5	Hinge region analysis	47
2.12	Mass spectrometry.....	47
3	Expression, refolding and purification of 7WC and DbLnP.....	48
3.1	Introduction	48
3.2	Cloning and expression of 7WC.....	48
3.2.1	Cloning of mature 7WC peptide.....	48
3.3	Cloning and expression of DbLnP.....	49

3.3.1	Cloning of mature DbLnP peptide	49
3.3.2	Initial expression trials of DbLnP.....	50
3.4	Large scale expression and solubilisation of 7WC and DbLnP	51
3.4.1	Considerations for large scale expression and solubilisation	51
3.4.2	IMAC chromatography in denaturing conditions of 7WC and DbLnP	52
3.5	Refolding	54
3.5.1	Introduction.....	54
3.5.2	Refolding matrix.....	56
3.5.2.1	Selection of the refolding buffers for 7WC	56
3.5.2.2	Determining the efficiency of refolding 7WC based on activity..	57
3.5.2.3	Determining the efficiency of refolding 7WC based on aggregation.....	58
3.5.2.4	Fine screening of the refolding conditions for 7WC	59
3.5.3	Selection of the refolding buffers for DbLnP.....	60
3.5.3.1	Determining the efficiency of refolding DbLnP based on activity 60	
3.5.3.2	Determining the efficiency of refolding based on aggregation for DbLnP	61
3.5.3.3	Fine screening of the refolding conditions for DbLnP.....	62
3.5.4	Additional parameters for optimal refolding for 7WC and DbLnP. 62	
3.5.5	Large scale refolding preparation of 7WC and DbLnP	64
3.5.6	Analysis and purification of refolded 7WC	64
3.5.6.1	In-gel activity assay.....	64
3.5.6.2	Removing the soluble aggregate.....	65

3.5.6.3	Cation exchange of 7WC.....	65
3.5.6.4	Purification of 7WC by Size exclusion chromatography.....	66
3.5.7	Analysis and purification of refolded DbLnP.....	68
3.5.7.1	Purification of active form of DbLnP by Size exclusion chromatography.	69
3.6	Discussion and concluding remarks	72
4	Crystallisation and crystallography 7WC.....	75
4.1	Introduction	75
4.2	Initial crystallisation experiments.	75
4.2.1	Further crystallisation trials.	76
4.3	Data collection.....	77
4.4	Data processing.....	78
4.5	Molecular replacement.....	79
4.6	Overall structure of the 7WC enzyme.....	83
4.7	Comparisons with other NTPDases	90
4.8	Substrate soaking.....	93
4.9	X-ray data collection of the AMP soaked crystals.....	93
4.10	Molecular replacement for the AMP structure.....	94
4.11	AMP binding mode in the active site	97
4.12	Overall structure and comparison of the AMP bound with the apo 7WC structure.....	99
4.13	Comparison of the AMP bound 7WC with other homologous structures.	100
4.14	X-ray data collection of the AMPPNP soaked crystals	102
4.15	Data processing of the AMPPNP soaked crystals.....	102
4.16	Molecular replacement for the AMPPNP-bound structure.....	103

4.17	The active site of the 7WC containing AMPPNP	106
4.18	Conformation of AMPPNP based on LpNTPDase1 structure	107
4.18.1	Conformation of AMPPNP based on the NTPDase2 structure	110
4.19	Discussion.....	111
5	DbLnP crystallisation	113
5.1	Introduction.....	113
5.2	Crystallisation of DbLnP	113
5.2.1	Thrombin cleavage.....	113
5.2.2	Crystallisation of thrombin treated DbLnP	116
5.2.3	Further crystallisation trials.....	116
5.2.3.1	Crystal form 1 of thrombin-treated DbLnP	117
5.2.3.2	Crystal form 2 of thrombin-treated DbLnP	117
5.2.4	Data collection	118
5.2.4.1	Crystal form 1 of thrombin-treated DbLnP	118
5.2.4.2	Crystal form 2 of thrombin-treated DbLnP	118
5.2.5	Data processing of the AMP and AMPPNP co-crystallised DbLnP 119	
5.2.6	Molecular replacement of the DbLnP co-crystallised with AMP ..	120
5.2.7	Model refinement of the DbLnP co-crystallised with AMP model 121	
5.2.8	The overall structure of DbLnP+PO ₄	124
5.2.9	Comparison of the DbLnP model with apo-7WC.....	127
5.2.10	Metal ion is present in the active site of DbLnP.....	129
5.2.11	Water is coordinated for nucleophilic attack	132
5.2.12	Structural analysis of the DbLnP+PO ₄ structure.....	133
5.3	DbLnP co-crystallised with AMPPNP	135

5.3.1	Molecular replacement of DbLnP co-crystallised with AMPPNP	135
5.3.2	Model refinement of the DbLnP co-crystallised with AMPPNP ...	136
5.3.3	Structural analysis of the two DbLnP structures	139
5.4	Substrate induced trans-domain movement of 7WC and DbLnP	140
5.4.1	Hinge domain analysis using apo-7WC and 7WC with AMP bound. 141	
5.4.2	Hinge domain analysis using apo-7WC and DbLnP co-crystallised with AMPPNP	143
5.4.3	The active site of the AMPPNP co-crystallised DbLnP.....	146
5.4.3.1	Interactions that AMPPN makes with DbLnP side-chains	147
5.4.3.2	Position of the free phosphate in the DbLnP+ AMPPN model .	150
5.4.3.3	Hydrolysis of the AMPPNP in the DbLnP crystal.....	151
5.4.3.4	Metal ion binding in the active site.....	152
5.4.3.5	DbLnP co-crystallised with AMPPNP does not contain the nucleophilic water.	154
5.4.3.6	Superimposition of the AMPPN with the rat NTPDase2 model 154	
5.5	Discussion	155
6	Kinetic parameters of 7WC and DbLnP.....	158
6.1	Introduction	158
6.2	Kinetic analysis of 7WC.....	158
6.2.1	Divalent cation co-factor dependent 7WC catalysis.....	158
6.2.2	Influence of pH on the activity of 7WC.....	159
6.2.3	Substrate specificity	160
6.2.4	Determination of the initial rate.....	162
6.2.5	Michaelis-Menten kinetics for 7WC.....	165

6.3	Kinetic parameters of DbLnP.....	168
6.3.1	Divalent cation co-factor dependent DbLnP catalysis.....	168
6.3.2	Influence of pH profile on the activity of DbLnP.	169
6.3.3	Substrate specificity of DbLnP	172
6.3.4	Determination of the initial rate.....	172
6.3.5	Michaelis-Menten kinetics for DbLnP	173
6.4	Discussion.....	177
7	General discussion.....	182
7.1	General background	182
7.2	Expression, refolding and crystallisation.....	182
7.3	Structural elucidation	183
7.4	Kinetic characterisation	187
7.5	Future research.....	188
7.6	Conclusion.....	189
8	Appendices	190
8.1	Nucleotide and protein sequences.....	190
8.2	Refolding screens	192
	References.....	194

List of Figures

Figure 1-1 Signal transduction pathway of extracellular ATP in plants.....	3
Figure 1-2 Amino acid sequence alignment for NTPDases.	8
Figure 1-3 Surface representation of published NTPDase structures.	12
Figure 1-4 Active site of the rat NTPDase2 illustrating the adenine and the first phosphate binding site.	13
Figure 1-5 Active site of the rat NTPDase2 illustrating the metal binding site.	14
Figure 1-6 Catalytic mechanism of NTPDase.	18
Figure 3-1 Protein sequence alignment of soluble plant NTPDases.	48
Figure 3-2 Agarose electrophoresis of the restriction digested DNA.	49
Figure 3-3 Small scale expression trial of pet28:DbLnP.	50
Figure 3-4 IMAC purification of 7WC in denaturing conditions.	53
Figure 3-5 IMAC purification of DbLnP in denaturing conditions.....	54
Figure 3-6 Schematic of how protein aggregates.....	55
Figure 3-7 Activity assay of 7WC after refolding.	57
Figure 3-8 Aggregation assay of 7WC after refolding.....	59
Figure 3-9 Evaluation of protein activity after refolding DbLnP.	60
Figure 3-10 Aggregation of DbLnP after refolding	61
Figure 3-11 Activity analysis of refolding at various temperatures for 7WC and DbLnP.	63
Figure 3-12 Activity analysis of refolding with increasing time for 7WC and DbLnP.	63
Figure 3-13 Native gel analysis of 7WC after cation exchange.....	66
Figure 3-14 A) S200 chromatography of 7WC. B) 10 % native and C) 10 % SDS-PAGE.	68
Figure 3-15 Native gel analysis of DbLnP after refolding.....	69
Figure 3-16 16/60 S200 Size exclusion chromatography of refolded DbLnP.	70
Figure 3-17 A) Native-PAGE and B) In-gel activity assay of the size exclusion purified DbLnP.	71

Figure 3-18 10/300 S200 Size exclusion chromatography of refolded DbLnP after two weeks.	72
Figure 4-1 Crystallisation of 7WC.....	77
Figure 4-2 Diffraction pattern collected on the home source x-ray generator and MAR detector.	78
Figure 4-3 The result from the TLS analysis of apo-7WC.	81
Figure 4-4 Ramachandran analysis of apo 7WC.....	83
Figure 4-5 Overall structure of 7WC.....	85
Figure 4-6 Topology model of 7WC.....	86
Figure 4-7 Superposition of apo-7WC (light blue) with rat NTPDase2 with AMPPNP bound (magenta).	87
Figure 4-8 Surface model of apo-7WC demonstrating the three regions of 7WC that interact with substrate.....	88
Figure 4-9 Cartoon representation of 7WC highlighting the ACRs.	89
Figure 4-10 Overlay of apo 7WC and NTPDase2.	92
Figure 4-11 Ramachandran analysis of AMP bound 7WC.	95
Figure 4-12 Stereo image of the active site of 7WC with AMP bound.	98
Figure 4-13 Superimposition of the apo-7WC and AMP bound 7WC using SSM.	99
Figure 4-14 Pymol B-factor putty.....	100
Figure 4-15 Stereo image of the active site comparison of <i>R. novogicus</i> NTPDase2 and 7WC.	101
Figure 4-16 Ramachandran analysis of AMPPNP bound 7WC.....	104
Figure 4-17 SSM superimposition of apo (pink), AMP bound (blue) and AMPPNP (green) 7WC structures.	106
Figure 4-18 Stereo image of a comparison of the binding mechanism of AMPPNP in LpNTPDase1 and NTPDase2.....	106
Figure 4-19 Close up of the AMPPNP with electron density map.	108
Figure 4-20 Stereo image of the binding mechanism of AMPPNP in 7WC based on the conformation of LpNTPDase1.....	109

Figure 4-21 Stereo image of the active site of 7WC with AMPPNP bound in conformation 1.	111
Figure 5-1 Thrombin cleavage site of the N-terminus of DbLnP	113
Figure 5-2 SDS-PAGE of the small scale optimisation of thrombin cleavage of DbLnP	114
Figure 5-3 Size exclusion of thrombin-treated DbLnP.	115
Figure 5-4 Crystal type 1 of DbLnP co-crystallised with AMPPNP.....	117
Figure 5-5 Crystals of DbLnP.....	118
Figure 5-6 X-ray diffraction of crystal form 2 of DbLnP co-crystallised with AMP.	119
Figure 5-7 The result from the TLS analysis of DbLnP+PO ₄	122
Figure 5-8 Ramachandran analysis of DbLnP +PO ₄	124
Figure 5-9 Cartoon representation of DbLnP+PO ₄	125
Figure 5-10 Topology model of DbLnP+PO ₄	126
Figure 5-11 Stereo diagram of the active site of DbLnP+PO ₄	127
Figure 5-12 Stereo diagram of phosphate binding site 1 of apo-7WC (magenta) and DbLnP+PO ₄ (yellow).	128
Figure 5-13 Stereo diagram of an overlay of the C-terminal phosphate binding site.	129
Figure 5-14 X-ray fluorescence scan of DbLnP+ PO ₄ crystal outside of the loop.	130
Figure 5-15 Stereo diagram of manganese coordination in the DbLnP+PO ₄ structure.....	131
Figure 5-16 Stereo diagram of an overlay of DbLnP+PO ₄ (green) and rat NTPDase2 with AMPPNP bound (pink).	132
Figure 5-17 Overlay of DbLnP+PO ₄ and rat NTPDase2 with AMPPNP bound that illustrates the position of the nucleophilic water.....	132
Figure 5-18 The results from the TLS analysis of DbLnP+AMPPNP.	137
Figure 5-19 Thermal ellipsoid diagrams of NTPDases.....	138
Figure 5-20 Overlay of DbLnP structures co-crystallised in AMP (green) and AMPPNP (cyan).....	140

Figure 5-21 SSM overlay of apo-7WC (blue) and AMPPNP co-crystallised DbLnP (red).	140
Figure 5-22 Visualisation of domain movement and hinge domains of 7WC upon AMP binding.	142
Figure 5-23 Visualisation of the hinge region using apo-7WC and AMPPNP co-crystallised with DbLnP.	144
Figure 5-24 AMPPN and phosphate ion conformation in the active site of DbLnP.	146
Figure 5-25 Stereo diagram of the nucleobase binding site within the active site of DbLnP+AMPPN.	147
Figure 5-26 Stereo diagram of an overlay of 7WC and DbLnP + AMPPN.	148
Figure 5-27 Stereo diagram of the active site of DbLnP+AMPPN.	150
Figure 5-28 Overlay of the free phosphate of the apo-7WC and DbLnP+AMPPN structures.	151
Figure 5-29 Mass spectrometry plot of AMPPNP.	152
Figure 5-30 Stereo diagram of manganese (black sphere) bound in the active site of DbLnP+AMPPN.	153
Figure 5-31 Stereo diagram of the superposition of rat bound AMPPNP (purple) and DbLnP bound AMPPN and phosphate (green).	154
Figure 6-1 Calcium saturation curve for 7WC with ADP and ATP as substrate.	159
Figure 6-2 7WC enzyme activity over a range of pH with various nucleotides.	162
Figure 6-3 Specific activity of various substrates for 7WC	162
Figure 6-4 Time course activity assay of 7WC with various nucleotides.	164
Figure 6-5 Michaelis-Menten plots for 7WC for various nucleotides.	167
Figure 6-6 Calcium saturation curve for DbLnP with ADP and ATP as substrate.	169
Figure 6-7 DbLnP enzyme activity over a range of pH with various nucleotides.	170
Figure 6-8 Time course activity assay of DbLnP with various nucleotides.	171
Figure 6-9 Specific activity of various substrates for DbLnP.	172
Figure 6-10 Michaelis-Menten plots for DbLnP for various nucleotides.	175

Figure 7-1 Schematic of the binding and catalytic mechanism of NTPDases.187

Figure 8-1 Gene and protein sequence of the 7WC expression construct..... 190

Figure 8-2 Gene and protein sequence of the DbLnP expression construct.191

List of Tables

Table 1-1 Site mutagenesis of NTPDases.....	21
Table 1-2 Specific activities of naturally sourced NTPDases.	23
Table 1-3 Michaelis-Menten kinetics of various NTPDases.	25
Table 2-1 Components of the native PAGE gels.	34
Table 3-1 Refolding conditions that showed the highest activity for 7WC.	58
Table 3-2 The top refolding conditions for DbLnP based on activity.	61
Table 3-3 Summary of the top condition by aggregation for refolding DbLnP.....	62
Table 4-1 Data collection statistics for home source data for apo-7WC.....	79
Table 4-2 Refinement and validation statistics of apo-7WC.	82
Table 4-3 Residues involved in the substrate binding regions within the active site cleft.	88
Table 4-4 DaliLite structural alignment.....	91
Table 4-5 Data collection statistics for AMP bound 7WC	94
Table 4-6 Refinement and model statistics for AMP bound 7WC.....	97
Table 4-7 Data collection statistics for AMPPNP bound 7WC.....	103
Table 4-8 Refinement and model statistics for AMPPNP bound 7WC.	105
Table 5-1 Data processing statistics for AMP and AMPPNP co-crystallised DbLnP	120
Table 5-2 Model and refinement statistics of the final model of DbLnP+PO ₄	123
Table 5-3 DALI-Lite structural alignment.	134
Table 5-4 Model and refinement statistics of the final model of DbLnP co-crystallised with AMPPNP.....	136
Table 6-1 Summary of kinetic parameters for 7WC.....	168
Table 6-2 Summary of the kinetic parameters of DbLnP	176
Table 7-1 Table of structures determined in this project.	184
Table 8-1 7WC refolding screen.	192
Table 8-2 Refolding plate for DbLnP.	193

List of Equations

Equation 1-1 One step hydrolysis.....	17
Equation 1-2 Two step hydrolysis	17
Equation 1-3 Three step hydrolysis	17
Equation 4 Beer's law.....	36
Equation 5 Determination of the time required for 10 % of the total substrate to be hydrolysed.....	38
Equation 6. Proposed method of hydrolysis.....	110

List of Abbreviations

Unless mentioned, SI units (Système international d'unités) and standard chemical formulae are used throughout this thesis. Commonly used abbreviations are listed below.

AMP	Adenosine-5'-monophosphate
AMPPNP	Adenosine 5'-(β -imido)diphosphate
AMPPNP	Adenosine 5'-(β,γ -imido)triphosphate
ADP	Adenosine-5'-diphosphate
APS	Ammonium persulfate
ATP	Adenosine-5'-triphosphate
bp	Base pair(s)
BSA	Bovine serum albumin
°C	Degrees Celsius
cm	Centimetres
COS	<i>Cercopithecus aethiops, origin-defective SV-40 cellline</i>
CV	Column volume
Da	Dalton
DNA	Deoxyribonucleic acid
DSMO	Dimethyl sulfoxide
DTT	Dithiolthreitol
EDTA	Ethylenediaminetetraacetic acid
FPLC	Fast liquid chromatography
x·g	Multiples of gravity (9.81 m/s ²)
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
His-tag	Poly histidine tag
IMAC	Immobilised metal affinity chromatography
IPTG	Isopropyl β -D-1-thiogalactopyranoside
Kan	Kanamycin
kDa	KiloDalton
K	Kelvin

L	Litre
LB	Luria broth
MES	2-(N-morpholino)ethanesulfonic acid
m	Meters
MPD	2-Methyl-2,4-pentanediol
min	Minutes
MilliQ	Milli-Q water, purified water that has been passed through a ion-exchange filter
MAPK	Mitogen-activated protein kinase
MWCO	Molecular weight cut-off
nm	Nano meter
N-terminus	amino terminus of peptide chain
PAGE	Polyacrylamide gel electrophoresis
PDB	Protein data base
PEG	Polyethylene glycol
rms	Root mean square
RNase	Ribonuclease
rpm	Revolutions per minute
RO	Reverse osmosis purified water
ROS	Reactive oxygen species
sec	Second
SDS	Sodium dodecyl sulfate
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SE	Standard error
SSRL	Stanford Synchrotron Radiation Laboratory
TAE	Tris-acetate-EDTA
TEMED	Tetramethylethylenediamine
Tris-HCl	2-Amino-2-hydroxymethyl-propane-1,3-diol-HCl
TBS	50 mM Tris-HCl, 150 mM NaCl, pH 7.5
UV	Ultraviolet light (280 nm)
V	Volts

1. Introduction

1.1 NTPDases

The nucleoside triphosphate diphosphohydrolases or NTPDases are a family of enzymes that catalyse the hydrolysis of the phosphoanhydride bond of nucleotide di- or tri-phosphate liberating a phosphate ion and nucleotide di- or monophosphates (EC 3.6.1.5). The reaction requires the presence of a divalent metal ion for efficient hydrolysis to occur.

The NTPDases are present in Animalia, Plantae, Fungi, Protista and Prokaryota (Knowles, 2011) and are involved in diverse physiological functions including thromboregulation, vascular inflammation, neuronal transmission and cell adhesion in humans, nodulation and cell division in plants and are associated with pathogenesis of bacteria and fungi (Geigenberger, et al., 2009; Robson, 2006; Sansom, et al., 2008).

Due to the extracellular localisation and the ability to hydrolyse nucleotides, the NTPDases play key roles in regulating the purinogenic signalling pathway in humans and higher plants (Clark, et al., 2009; Yegutkin, 2008).

1.2 Physiological mechanisms of NTPDases.

1.2.1 Purinogenic signalling and physiological impact of NTPDases in plants

Over the past 10 years there has been an influx of evidence demonstrating that ATP and other nucleotides present in the extracellular matrix can act as signalling molecules that control a myriad of physiological processes (Clark, et al., 2009; Geigenberger, et al., 2009; Jeter, et al., 2006; Roux, et al., 2007). This phenomenon was first observed after the application of exogenous ATP to the Venus fly trap (*Dionaea muscipula*) and observing the closure of the leaves (Jaffe, 1973). Since then extracellular signalling by ATP has been attributed to

membrane depolarisation, production of reactive oxygen species (ROS) and an influx of cytosolic calcium ions which have been shown to promote plant growth. The control of the extracellular nucleotides by extracellular NTPDases therefore influences the growth of the plant. For example, silencing of the apoplastic apyrase in potato and DNA microarray showed an increased level of cell wall transcripts and other plant growth genes (Riewe, et al., 2008b). Further evidence was reported in *Arabidopsis thaliana* where the knockout of two NTPDase genes, *AtAPY1* and *AtAPY2*, or by suppressing the NTPDase activity using antibodies, resulted in the inhibition of pollen tube germination (Steinebrunner, et al., 2003; Wu, et al., 2007). Furthermore, RNA interference of *AtAPY2* in an *AtAPY1* knock-out background resulted in 12-40 % reduction in the hypocotyl and root length when compared to the wild type (Wu, et al., 2007). Conversely, the over expression of the *AtAPY2* or the *AtAPY1* genes led to 10-15 % longer hypocotyls and pollen tubes grew significantly faster. If the extracellular ATP dropped below an optimal concentration, growth was inhibited. This was demonstrated with the application of substantial amounts of potato apyrase to *Medicago truncatula* which significantly reduced root growth (Kim, et al., 2006). In addition, cell death in *A. thaliana* cell cultures was attributed to low ATP after treatment with apyrase (Chivasa, et al., 2005).

The current model, proposed by Roux and Steinebrunner, 2007, is that a positive physiological impact of extracellular nucleotides for growth is optimal at a specific nucleotide concentration and that deviations from this, either higher or lower results in less growth (Roux, et al., 2007).

At the present time, the extracellular nucleotide signal transduction pathway in plants has not been completely resolved, however based on research thus far a schematic was produced by Tanaka, et al., 2010 (Figure 1-1).

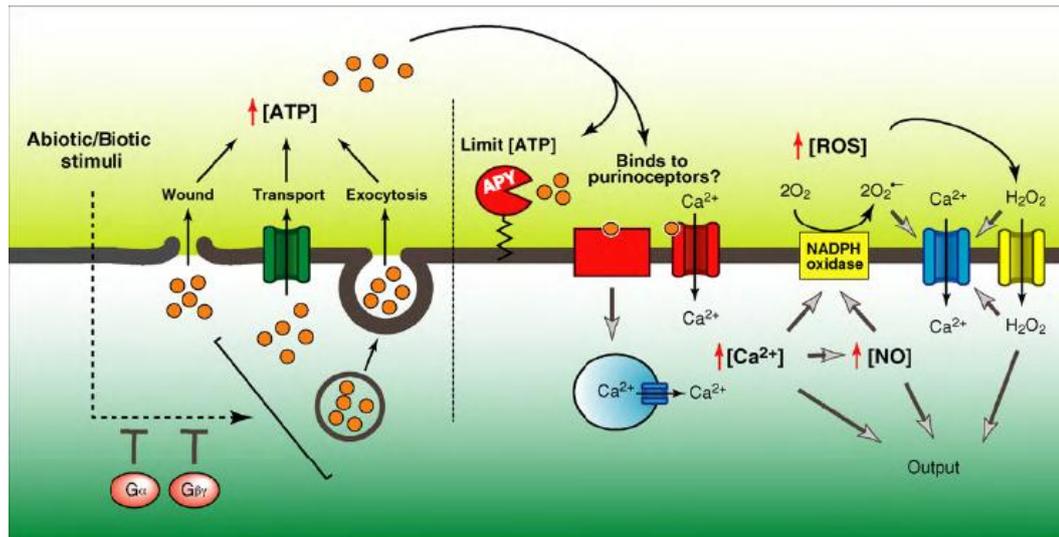


Figure 1-1 Signal transduction pathway of extracellular ATP in plants. This figure has been taken from (Tanaka, et al., 2010).

With reference to Figure 1-1: The plant responds to a stimulus (either abiotic or biotic) and ATP is released either by a physical wound, exocytosis, or active transport through multidrug resistance transporters into the apoplast possibly facilitated by heterotrimeric G proteins (Jeter, et al., 2004; Lazarowski, et al., 2003; Thomas, et al., 2000; Weerasinghe, et al., 2009). If the plant transduction pathway is similar to the mammalian system (section 1.2.2) then the extracellular ATP activates a purinoreceptor. However, purinoreceptors have not yet been identified (Geigenberger, et al., 2009; Jeter, et al., 2006; Roux, et al., 2007). Extensive sequence based investigations with rice and *A. thaliana* have found no plant proteins that exhibit substantial sequence similarity to the animal purinoreceptors (Gookin, et al., 2008; Moriyama, et al., 2006). A recent study has revealed a plasma membrane localised purinoreceptor that contains < 30 % sequence identity with the human receptors in the green algae *Ostreococcus tauri* (Fountain, et al., 2008). This sequence is not present in the genome of higher plants (Tanaka, et al., 2010). Although, with the detection of an adenosine nucleoside in the extracellular matrix it cannot be ruled out that an adenosine receptor is responsible for transmission of the signal (not shown in Figure 1-1) (Demidchik, et al., 2003; Riewe, et al., 2008a; Zhang, et al., 2000).

The downstream impact of the nucleotide signalling pathway is an increase of the cytosolic Ca^{2+} concentration either through the release of Ca^{2+} from internal stores or by the opening of the plasma membrane Ca^{2+} channels (Demidchik, et al., 2009; Jeter, et al., 2004). The increased Ca^{2+} in turn, increases cytosolic nitric oxide (NO) which both increases the NADPH oxidase activity and increases ROS in the apoplast (Corpas, et al., 2004; Demidchik, et al., 2009; Kim, et al., 2006; Reichler, et al., 2009; Song, et al., 2006). This increase in ROS promotes additional influx of Ca^{2+} ions into the cytoplasm by opening the Ca^{2+} ion gated channels.

These molecules (Ca^{2+} , ROS and NO) represent the initial steps of the ATP signalling pathway; however the subsequent steps in the transduction pathway are yet to be determined. It has been proposed that the initial steps of the adenylate pathway act on transcription by increasing the hormone ethylene that mediates the MAPK signal transduction pathway and changes gene expression (Jeter, et al., 2006).

1.2.1.1 *Physiological role of NTPDases in legumes*

In addition to the effect of extracellular nucleotides on plant growth, some NTPDases have been identified that play a role in the symbiosis between legumes and rhizobium during nodulation. The lectin-nucleotide phosphohydrolase isolated from the surface of the roots of the legume *Dolichos biflorus* (DbLnP) was found to bind to the lipochitin-oligosaccharides or NOD factors that increase the DbLnP activity (Etzler, et al., 1999). The NOD factors are produced by the rhizobium as early signalling molecules that trigger nodulation. The relationship between DbLnP and nodulation was further reinforced by the redistribution of DbLnP to the tips of the root hairs upon infection with either the NOD factor or the correct symbiotic rhizobium (Kalsi, et al., 2000). The importance for NTPDases in early nodulation was further emphasised with comprehensive work on the soybean (*Glycine soja*) NTPDase, termed GS52, which was observed to be expressed during nodulation (Day, et al., 2000). Over-expression of GS52 in *Lotus japonicus* resulted in increased numbers of nodules

and increased length of the infection thread upon infection with rhizobium (McAlvin, et al., 2005). Further studies showed that NTPDases are involved very early in the establishment of the symbiosis as RNA interference of GS52 in *L. japonicus* developed immature nodules but did not establish infection of rhizobium nor produce mature nodules (Govindarajulu, et al., 2009). Consistent with the proposed signal transduction pathway, the rhizobium infection in legumes results in an increase in cellular Ca²⁺ ions and ROS (Hamel, et al., 2010).

1.2.2 Purinogenic signalling and physiological impact of NTPDases in mammalian systems

The extracellular nucleotide pathway in mammals has been well characterised largely due to the overwhelming evidence that nucleotide signalling plays an essential role in a number of physiological processes including pain perception, sympathetic neurotransmission, vascular inflammation, cystic fibrosis and vascular thrombosis (Abbracchio, et al., 2006; Burnstock, 2007). During purinogenic signalling nucleotides are released from the cell by exocytosis and activate purinoreceptors. The membrane bound purinergic receptors are a class of proteins that bind to nucleotides to initiate a signal cascade which initially increases the levels of cytosolic Ca²⁺ ions before inferring a physiological response. The purinogenic receptors are classed as P2X, P2Y or P1. The P2X are ligand-gated ion channels that open when ATP is bound to form a channel letting Ca²⁺ and Na⁺ ions into the cytosol (Kawate, et al., 2009; Surprenant, et al., 2009). The P2Y are G-coupled receptors that initiates a cascade of signals when either ATP or ADP is bound resulting in release of cytosolic Ca²⁺ from cytosolic stores (Abbracchio, et al., 2006). The P1 receptor is a G-coupled receptor that specifically binds to adenosine that regulates the adenylate cyclase.

The control of the signal transduction pathway rests with the membrane bound ecto-NTPDases which regulate the concentration of ATP, ADP, AMP and 5'-nucleotidases that hydrolyse AMP further to adenosine (Colgan, 2006; Yegutkin, 2008; Zimmermann, 2000).

The best characterised pathway involving NTPDases is the regulation of vascular ischemia in humans by NTPDase1 (CD39). Upon damage to the vascular system blood platelets release ADP that promotes the accumulation of platelets to form an occlusion through interactions with the platelet purinoreceptors. CD39 is an integral component of the endothelial cell surface of the vascular system that hydrolyses the nucleotides which inhibits platelet aggregation and maintains vascular fluidity (Marcus, et al., 2005; Marcus, et al., 2001; Robson, 2006). The intravenous application of recombinant soluble CD39 to mice inhibited platelet aggregation (Gayle, et al., 1998). Furthermore, mice deficient in CD39 or over-expression of human CD39 showed that the CD39 had a direct impact on the homeostasis of platelet aggregation (Dwyer, et al., 2004; Enjyoji, et al., 1999). The importance of CD39 in vascular thrombosis has made it a target for the treatment of strokes.

1.2.3 NTPDases pathogenic organisms

NTPDases in various pathogenic organisms received significant attention due to their importance in the infection process, notably *Toxoplasma gondii*, *Schistosoma mansoni*, *Trypanosoma cruzi*, *Trichomonas vaginalis* and *Legionella pneumophila* (Asai, et al., 1983; de Jesus, et al., 2002; DeMarco, et al., 2003; Fietto, et al., 2004; Sansom, et al., 2008). A detailed mechanism for purinogenic signalling has yet to be elucidated. But experimental evidence has shown that the NTPDases in parasitic organisms either work towards establishing an infection by subverting the host defence mechanisms or by maintaining the infection through purine scavenging (Sansom, et al., 2008). For example, *T. gondii* is unable to synthesise purines, and so relies on the scavenging of purines from the host. It is thought that NTPDases play an integral role in hydrolysing the hosts ATP prior to the transportation of purines by the parasite for the anabolism of purine nucleotides (de Koning, et al., 2000; Silverman, et al., 1998).

1.3 The primary sequence of NTPDases

1.3.1 The apyrase conserved regions (ACR)

The apyrase conserved regions are highly conserved motifs in the primary sequence that identifies the protein as an NTPDase. Handa and Guidotti in 1996 first identified four conserved regions after the molecular characterisation of the *S. tuberosum* and penned the acronym ACR (Handa, et al., 1996). Later that year a fifth was added to the nomenclature after the identification of an NTPDase in *S. mansoni* (Vasconcelos, et al., 1996). These conserved regions are shown in Figure 1-2 highlighted in the red boxes. Each of these regions has been denoted a function after comprehensive mutagenesis and structural studies (section 1.5 and 1.4).

1.3.1.1 Membrane and soluble NTPDases

The NTPDases in various biological systems can exist as soluble or membrane bound enzymes where the latter have transmembrane domains either at the N-terminus or at both the N- and C-terminus. The soluble NTPDases which are typically extracellular include the NTPDases in this study 7WC (NTPDase7 from *Trifolium repens*), DbLnP (*Dolichios biflorus* lectin nucleotide phosphohydrolase), and the potato apyrase and the mammalian endoplasmic reticulum NTPDase5. Membrane bound extracellular NTPDases in mammals include NTPDase1-4 and 7 which are anchored by two transmembrane domains whereas NTPDase5 and 6 are anchored by one N-terminal transmembrane domain. The NTPDase5 is present in either the membrane bound or soluble form.

The importance of the trans-membrane domains have been demonstrated in studies on human NTPDase1 where the removal of either one or two of the two transmembrane domains by detergent treatment not only significantly reduced activity and altered the substrate specificity but also resulted in the release of the intermediate during catalysis (Chen, et al., 2001; Grinthal, et al., 2002). A similar impact has been observed for other cell surface NTPDases including the rat NTPDase1 (Wang, et al., 1998), human NTPDase2 (Chiang, et al., 2008) and human and chicken NTPDase8 (Knowles, et al., 2006; Li, et al., 2010).

Furthermore, by adding a variety of amphiphilic compounds that impacted the elasticity of the lipid bilayer the enzyme activity was altered (Grinthal, et al., 2007). This work has illustrated that a high amount of movement is inextricably linked to the enzymes catalytic function and it has been suggested that this is a mode of regulation or involvement with the substrate-induced conformational change between the two domains (section 1.4.4) (Grinthal, et al., 2006, 2007; Zebisch, et al., 2008).

The transmembrane domains have also been illustrated to play a role in the oligomerisation in mammalian NTPDases. The membrane bound NTPDases form oligomers either as a dimer or as a tetrameric complexes and that in some cases the formation of protein complexes was shown to have an impact on the catalytic function (Caldwell, et al., 2001; Failer, et al., 2003; Wang, et al., 1998).

However, it is debated whether the oligomers, particularly the tetrameric state, are stable within the plasma membrane as there is not a consistent number of monomers observed within each oligomer (Failer, et al., 2003; Grinthal, et al., 2006; Knowles, 2011).

1.3.2 Disulfide bonds within NTPDases

The family of NTPDases contain a number of cysteines that form disulfide bonds. The elucidation of the rat NTPDase2 structure confirmed the formation of 5 cystine bridges. With reference to Figure 1-2; two disulfides (#3 and #5) are conserved within the larger GDA1/CD39 superfamily with one additional cystine (#4) in the plant NTPDases including *Trifolium repens* NTPDase7 (7WC), DbLnP, and potato NTPDase (Kirley, et al., 2006; Roberts, et al., 1999).

The recent elucidation of the *Legionella pneumonia* NTPDase1 structure shows only 2 disulfides that are analogous to #3 and #5 (Vivian, et al., 2010).

The disulfides #1 and #2 are specific to the membrane bound NTPDase1, 2 and 8 (Ivanenkov, et al., 2005; Zebisch, et al., 2008).

1.4 The NTPDase structure

1.4.1 Early structural determination

Key research by Smith and Kirley, 1999, demonstrated that the NTPDases were related to the actin/hsp70 superfamily (Smith, et al., 1999b). This study identified the key DXG phosphate binding motif that is strictly conserved within members of this family. (Bork, et al., 1992; Hurley, 1996; Smith, et al., 1999b). Also, point mutations of these residues significantly reduced the activity of NTPDase3. With this discovery the first insight into the tertiary structure of an NTPDase3 was produced using computational modelling based on the exopolyphosphatase/guanosine pentaphosphate phosphohydrolase (PPX/GPPA) X-ray structure (Ivanenkov, et al., 2005; Kristensen, et al., 2004). Later, using similar techniques a model for the potato apyrase was determined and the

binding mechanism of the ATP and metal ion were proposed based on earlier structures of actin (Kozakiewicz, et al., 2008b; Schuler, 2001).

1.4.2 Overall published structures of NTPDases

It wasn't until recently that the first experimental crystal structure of an NTPDase was revealed. The rat NTPDase2 structure was determined from bacterially expressed protein with the transmembrane domains removed. This was crystallised and the apo structure solved associated with: AMPPNP and Ca²⁺, AMP and Ca²⁺ and AMP with a phosphate and a Ca²⁺ ion at resolutions between 1.7-2.1 Å (Zebisch, et al., 2008). Following the rat NTPDase2 structure, the crystal structure of *Legionella pneumonia* apo-NTPDase (LpNTPDase1) was determined and also solved bound to AMPPNP and the NTPDase inhibitor ARL 67156 at a resolution of 1.6-2.0 Å (Vivian, et al., 2010).

The NTPDase2 and the LpNTPDase1 structures confirmed the results of the computational modelling demonstrating that the general structure of NTPDase2 and LpNTPDase1 is made up of N- and C-terminal domains each exhibiting a modified RNase-H fold that is characteristic of the actin/hsp70 superfamily (Aravind, et al., 1999; Koonin, 2010; Smith, et al., 1999b). Each domain has an overall topology of a mixed β -sheet flanked by a number of α -helices on one side and a single α -helix on the other. The domains mirror each other with the single α -helix of each domain mutually perpendicular at the bottom of a large cleft (Figure 1-3). The active site is located at the bottom of the large cleft between the two domains, where residues from both domains contribute to the binding of the substrate.

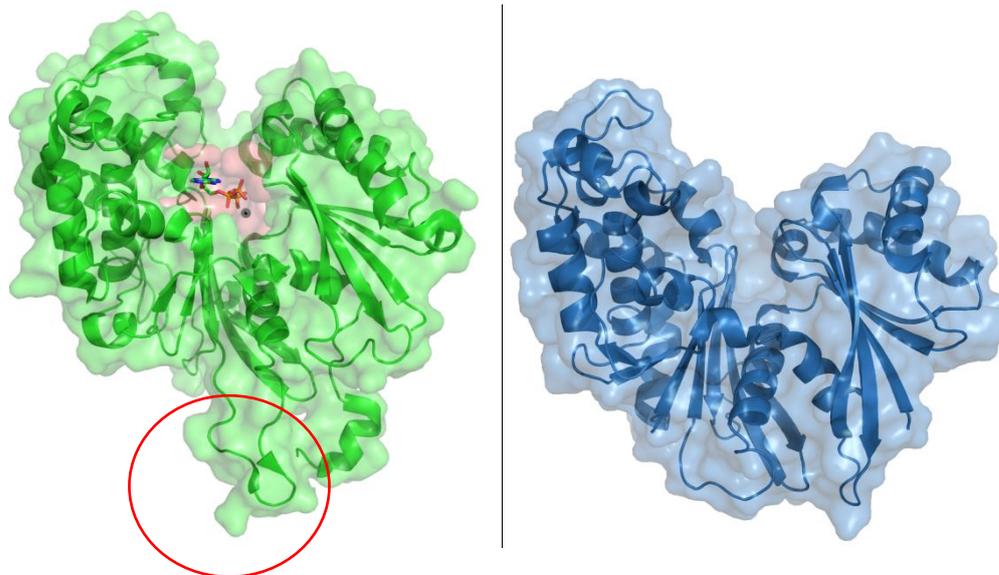


Figure 1-3 Surface representation of published NTPDase structures.

A) rat NTPDase2 (green) containing AMPPNP and Ca^{2+} ion (grey) and B) *L. pneumonia* NTPDase1 (blue). Pink surface in A) represents the active site of the enzyme. Circled is a loop extension of NTPDase2.

Comparing the overall structures of the rat NTPDase2 and LpNTPDase1 shows that despite the low sequence identity (22.5 %) both structures have the same overall fold comprising two domains. They differ in a number of extensions that are present in the rat NTPDase2, particularly in the C-terminal domain. These additional regions include a long loop (circled in Figure 1-3) that is thought to interact with the cell membrane and a larger helical bundle in the C-terminal domain (Figure 1-3).

1.4.3 Substrate binding mechanism

The binding of AMPPNP in the rat NTPDase2 was used to infer the binding mechanism of ATP and the metal ion in NTPDase2 (Zebisch, et al., 2008). Interactions with the substrate and the metal ion are facilitated by highly conserved residues that make up the ACRs. The adenine base associates with the enzyme only through π -stacking associations with the side chains of Tyr350 and Arg394. Such an association explains the low substrate specificity of NTPDase2.

The ribose of AMPPNP is in a C2'-endo conformation and the 2' hydroxyl forms an H-bond with Arg394.

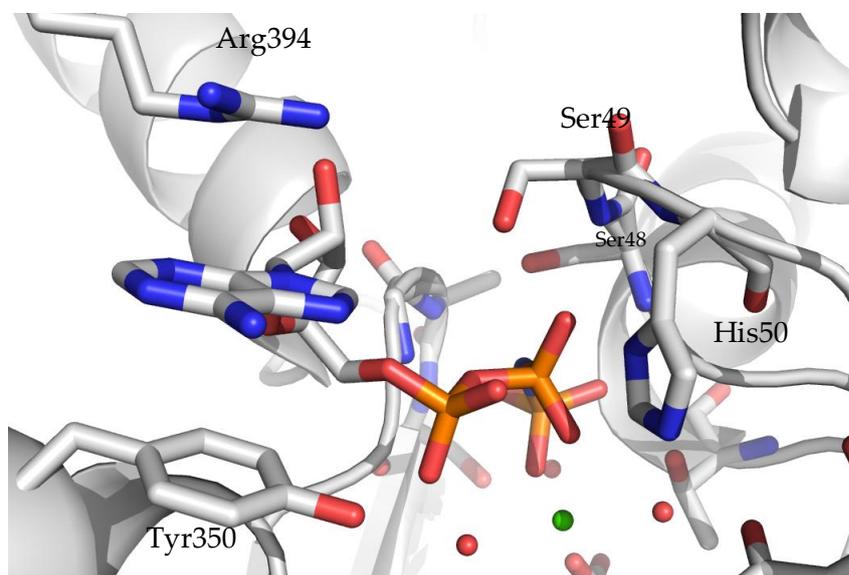


Figure 1-4 Active site of the rat NTPDase2 illustrating the adenine and the first phosphate binding site. Waters and metal ion are represented as red and green spheres, respectively.

The ACR1 contains Ser48, Ser49, and His50 that associate directly with the α - and β -phosphates and creates the phosphate binding site 1 (Figure 1-4). The second phosphate binding site consists of Ala205 and Ser205 that associates with the γ -phosphate of ATP (Figure 1-5).

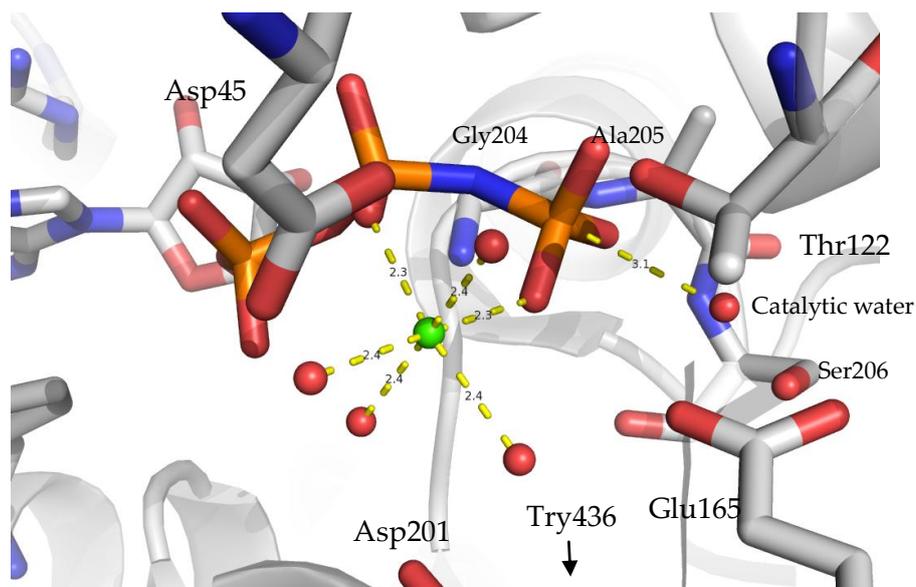


Figure 1-5 Active site of the rat NTPDase2 illustrating the metal binding site. Waters and metal ion are represented as red and green spheres, respectively.

The divalent metal ion is octahedrally coordinated with 4 waters and oxygen atoms from the β - and γ -phosphates of ATP. The waters are coordinated by Asp45 in ACR1, Thr122 in ACR2, Glu165 in ACR3, Asp201 in ACR4 and Trp436 in ACR5.

The *L. pneumonia* NTPDase reported the AMPPNP bound in a different conformation than the NTPDase2 (Vivian, et al., 2010). Comparable to that of the rat NTPDase2 structure, the LpNTPDase1 shows that the adenosine of the AMPPNP associated through π -stacking with the analogous Tyr346 and the ribose with the hydroxyl of the Tyr346 side-group. The major difference is that the γ -phosphate associates with the phosphate binding site 1 (consisting of Ser52, Thr53, Arg56) rather than with the second phosphate binding site (consisting of Gly189 and Gly190). Another major difference is that the LpNTPDase1 structure contains no metal ion bound in the active site.

1.4.4 Trans-domain movement upon substrate binding in exopolyphosphatase and actin

For both actin and *A. aeolicus* exopolyphosphatase it has been demonstrated that upon binding of both the metal ion and the substrate, a trans-domain

conformational change is induced (Kristensen, et al., 2004; Kristensen, et al., 2008; Schuler, 2001). Due to the structural relationship of the NTPDases with the *Aquifex aeolicus* exopolyphosphatase and actin (which are both members of the actin/hsp70 superfamily) a number of authors have suggested that NTPDases have a substrate bound conformational change that brings the domains closer, reducing the width of the inter-domain cleft (Gaddie, et al., 2010; Kirley, et al., 2006; Vivian, et al., 2010; Zebisch, et al., 2008). In the case of exopolyphosphatase the open state of the *Escherichia coli* exopolyphosphatase was compared with the closed conformation of the substrate bound *A. aeolicus* exopolyphosphatase that demonstrated a rotational movement of 22.5° that was required for the binding of the substrate. Kristensen, et al., revealed a region of amino acids that would act as a hinge domain between Arg121 and Tyr124 that are located in a stretch of residues that bridge the two domains. Extensive analysis has been carried out on the domain movement of actin. Reports of actin analysis has identified three possible ligand induced conformation changes; 1. The movement of the two domains that close the inter-domain shift brought about by shearing of the internal helix; 2. A clam-like rotation of the inter-domain cleft about a hinge at the base of the cleft, or 3. Through a number of smaller movements in the regions at the outer edges of the active site cleft. Generally speaking the N- and C-terminal domains act as rigid bodies. The current hypothesis for nucleotide binding is that the substrate with a Ca²⁺ ion initially binds to the phosphate binding site which results in the bridging of the substrate between the domains (Kabsch, et al., 1990). Subsequently the exchange of divalent cations from Ca²⁺ to the higher affinity Mg²⁺ coordinates the substrate towards the second phosphate binding site which in-turn induces a clam-like rotation of the domains (Frieden, 1983; Schuler, 2001). This rotation appears to be controlled by the coordination of the higher affinity metal ion and the binding of the substrates (Frieden, et al., 1988). The binding of the nucleotide, with the divalent cation, has also been seen to induce a rotation about more than one axis mediated by the shearing of the inter-domain helix (Frieden, et al., 1988; Goddette, et al., 1985; Kinoshian, et al., 1993; Schuler, 2001).

1.5 NTPDase kinetics

1.5.1 Early work on the catalytic function

The family of NTPDases enzymes was first described in the pioneering research on the *Solanum tuberosum* NTPDase as early as the 1940s which identified a family of enzymes that could hydrolyse the γ - and β - phosphate of nucleotides (Mayerhoff, 1945). Following the purification from *S. tuberosum* tubers the NTPDase was further characterised demonstrating the importance of a divalent cation cofactor for enzymatic function (Cohn, et al., 1956; Kettlun, et al., 1982; Krishnan, 1948; Liebecq, et al., 1962; Traversocori, et al., 1965).

Early mammalian research mistook NTPDase for the better characterised ATPase due to the extracellular localisation and the ability to hydrolyse ATP. The ATPases have a high specificity and hydrolyse ATP to ADP and phosphate, yet a subset of ATPases, later termed NTPDases, emerged that had low specificity for a range of nucleotide di- and triphosphates with the final product of hydrolysis being nucleotide monophosphate and two phosphate ions. In addition, the NTPDases were insensitive to ATPase specific F-, P-, and V-type inhibitors and alkaline phosphatases.

It wasn't until the molecular characterisation of the potato apyrase in 1966 by Handa and Guidotti that identified sequence homology between the apyrases and the recently cloned mammalian NTPDase1 that reclassification of the NTPDases occurred (Handa, et al., 1996; Kansas, et al., 1991; Maliszewski, et al., 1994; Plesner, et al., 1997).

1.5.2 Catalytic mechanism of NTPDase

The hydrolysis of ATP to AMP by the NTPDases was first proposed to operate via three possible mechanisms. It is widely agreed that a single step mechanism, shown in Equation 1-1, does not occur due to the accumulation of phosphate ions and not pyrophosphate produced during catalysis. This was first observed by the detection of radio labelled β - and γ -phosphates from ATP during the catalytic

reaction (Tognoli, et al., 1981). Other modes of catalysis are based on two step mechanisms; either ADP is not released before being hydrolysed (Equation 1-2) as shown for human and mouse membrane bound NTPDase1 (Kukulski, et al., 2005), or ADP is released before associating with the NTPDase again (Equation 1-3) which is observed for potato NTPDase, soluble mammalian NTPDase1-3 and -8 (Chen, et al., 2001; Kukulski, et al., 2005; Laliberte, et al., 1983; Vorhoff, et al., 2005; Zebisch, et al., 2007).



Equation 1-1 One step hydrolysis



Equation 1-2 Two step hydrolysis



Equation 1-3 Three step hydrolysis

1.5.3 Mechanism of hydrolysis

Early attempts to determine the catalytic mechanism for NTPDase (before the determination of the NTPDase structure) suggested a number of catalysis mechanisms based on point mutation studies and the mechanism determined for actin (Kirley, et al., 2006; Kozakiewicz, et al., 2008a; Vorobiev, et al., 2003). However it wasn't until the structural determination of the *R. novogicus* NTPDase2 bound to the non-hydrolysable substrate, AMPPNP was an acceptable mechanism agreed upon (Zebisch, et al., 2008). The catalytic mechanism of ATP hydrolysis by NTPDases is based on evidence from a number of studies on other members of the actin/hsp70 super-family including actin and hsp70 and the NTPDase *Solanum tuberosum* apyrase (Cohn, et al., 1956; Flaherty, et al., 1994; Vorobiev, et al., 2003; Zebisch, et al., 2008). Hydrolysis is thought to occur by the nucleophilic attack on the terminal phosphate by a 'catalytically active' water molecule. In the NTPDase2 structure a water molecule is coordinated by a serine side chain hydroxyl, the carboxyl of a glutamate side chain and backbone

amide of alanine. The catalytically active water is 3.11 Å from the terminal phosphate.

The function of the glutamate, 2.5 Å from the nucleophilic water is to act as a general base which activates the water before hydrolysis occurs. The active site glutamate is conserved within all NTPDases and the importance of these residues is supported by mutagenic studies (Kirley, et al., 2006). An analogous glutamate was suggested to play an equivalent role in the structurally related *A. aeolicus* exopolyphosphate phosphatase/guanosine pentaphosphate phosphatase (PPX/GPPA)(Kristensen, et al., 2004; Kristensen, et al., 2008).

The active site contains a divalent metal ion such as Mg^{2+} , Mn^{2+} or Ca^{2+} that is essential for maximal activity. It is thought to position the terminal phosphate closer to the nucleophilic water and contributes to the catalysis by polarising the phosphoanhydride bond of the terminal phosphate.

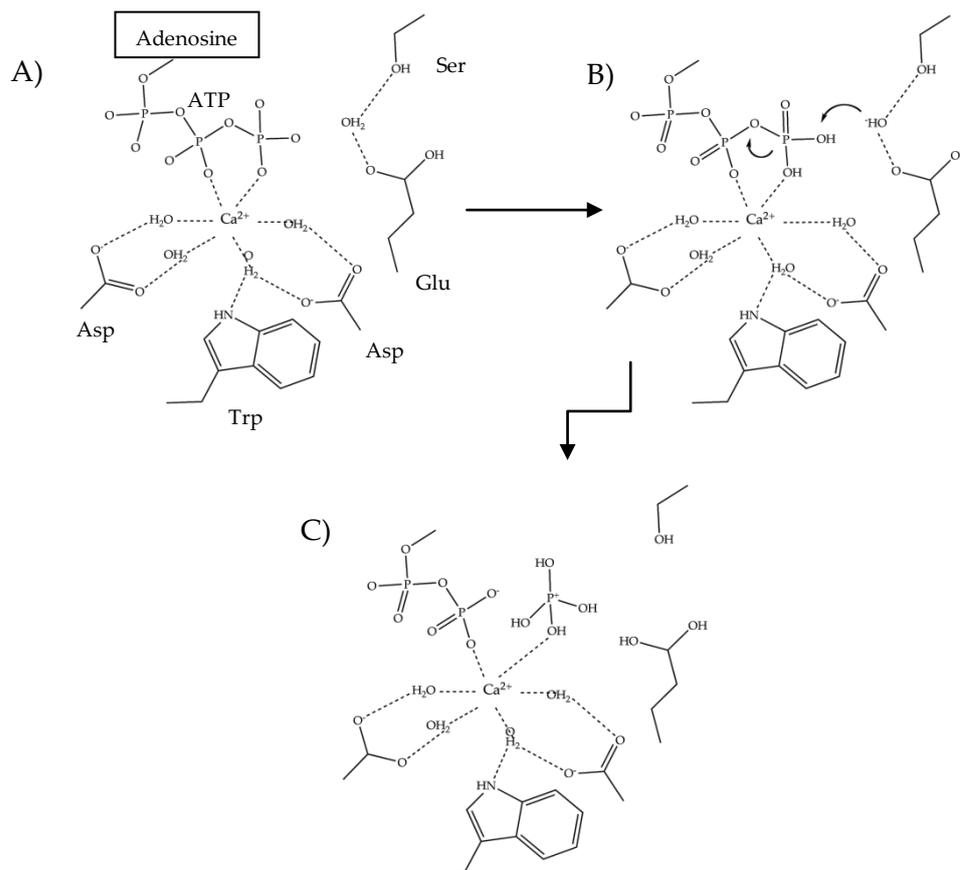


Figure 1-6 Catalytic mechanism of NTPDase.
Adapted from (Cohn, et al., 1956; Vorobiev, et al., 2003).

The hydrolysis of the ATP first requires the reorientation of the terminal phosphate to associate with the metal ion that creates an octahedral coordination of the metal ion with 4 waters and oxygens of the β - and γ -phosphate which withdraws the electrons from the phosphate (Figure 1-6-A). The glutamate activates the catalytic water to form the nucleophile hydroxide ion which attacks the γ -phosphate forming a pentacoordinate transition state. This nucleophilic attack on the terminal phosphate results in the release of the γ -phosphate from the ATP (Figure 1-6-B, C). The negative charge of the transition state of the product (ADP) after the nucleophilic attack is then stabilised through the electron withdrawing character of the metal ion by a series of hydrogen bonds from the NH backbone, side chain hydroxyls and an histidine imidazole ring (in the case of mammalian NTPDase1-3) or by arginine (in the case of 7WC and DbLnP). The hydrolysis of ADP is thought to be achieved by the same hydrolytic site that is used for ATP. The same residues that are involved in ATP catalysis are also essential for ADP hydrolysis since point mutations of the serine that orientates the catalytic water or the proton accepting glutamate abolish both ATP and ADP activities (Drosopoulos, et al., 2000; Tanaka, et al., 2011a; Yang, et al., 2001). The ADP hydrolysis mechanism has yet to be clarified through structural biology but it is thought that for ADP hydrolysis to occur the substrate would need to move closer into the binding pocket either through extension of the phosphate tail or global movement of the domains and/or localised movements of side chains (Zebisch, et al., 2008).

1.5.4 Residues involved in catalysis-site directed mutagenesis.

A large number of studies have been performed, particularly on the mammalian NTPDases1-3, to elucidate the residues essential for catalysis through point mutations.

Table 1-1 shows a list of the residues that have the greatest impact on catalytic activity. An exhaustive list of point mutations have been collated in recent reviews (Javed, et al., 2007; Kirley, et al., 2006). Point mutation studies have

generally targeted residues that occupy the apyrase conserved regions (ACRs) (section 1.4.3) and have been tested for the mutations' impact on catalytic function or affinity to metal ions. The residues that have the greatest impact on catalytic activity in human NTPDases are: for NTPDase1; Ser57, His59, Ser59, Tyr127, Glu174, Asp213 and Ser218 (Drosopoulos, 2002; Drosopoulos, et al., 2000), for NTPDase2; Trp58, Lys62, Thr66, Ile68 and Val69 (Javed, et al., 2006; Javed, et al., 2007) and for NTPDase3; Asp62, Gly64, S224, Trp187, Gly141, Gly222, Asn226 (Kirley, et al., 2001; Smith, et al., 1999a; Yang, et al., 2001).

Mutations on mammalian NTPDase1			
D54A (human)	ACR1	Lower affinity for Ca ²⁺ and nucleotides	(Drosopoulos, 2002)
H59G, S, R (rat)	ACR1	Convert to an ADPase	(Grinthal, et al., 2000)
S57A (human)	ACR1	Increase in Ca ²⁺ -ADPase activity	(Drosopoulos, et al., 2000)
H59A (human)	ACR1	Increase in activity	
S61A (human)	ACR1	Near WT activity	
K66A (human)	Near ACR1	Increase in activity	
Y127A (human)	ACR2	Decrease in activity, Mg ²⁺ -ADPase	
R135A (human)	ACR2	Small decrease in Mg ²⁺ -ADPase	
E174A (human)	ACR3	Inactive	
S218A (human)	ACR4	Mostly inactive	
D213A (human)	ACR4	Lower affinity for Ca ²⁺ and nucleotides	(Drosopoulos, 2002)
Mutations on mammalian NTPDase2			
C399S (human)	Disulfide bond 5	Inactive	(Mateo, et al., 2003)
N443	ACR5	Mostly inactive	
W58A	ACR1	Inactive	(Javed, et al., 2007)
K62A, R	ACR1	Inactive	
T66A	ACR1	Inactive	
I68A	ACR1	Inactive	
V69A	ACR1	Inactive	
Mutations on mammalian NTPDase3			
D62A, G64A	DXG1/ACR1	Inactive	(Smith, et al., 1999a)
D219A, G221A	DXG2/ACR4	Inactive	
W187A	ACR3	Inactive	
W459A	ACR5	Increase in ATPase/ADPase	
R143A	ACR2	Inactive	(Yang, et al., 2001)
E182D, Q	ACR3	Inactive	
N191A	ACR3	Decrease in ATPase/ADPase	
S224A	ACR4	Inactive	
Q226A	ACR4	Mostly inactive	
H135A	ACR2	Mostly inactive	
G98A	ACR1	~60% wt activity	(Kirley, et al., 2006)
G141A	ACR2	~20% wt activity	
G183A	ACR3	~65% wt activity	
G222A	ACR4	~10% wt activity	
G263A	ACR4	Inactive	
G462A	ACR5	~15% wt activity	
C92S/C116S	Disulfide bond 1	~20% wt activity	(Ivanenkov, et al., 2005)
C261S/C308S	Disulfide bond 2	~10% wt activity	
C289S/C343S	Disulfide bond 3	Inactive	
C347S/C353S	Disulfide bond 4	~90% wt activity	
C399S/C422S	Disulfide bond 5	Inactive	
Mutation on <i>G. soja</i> NTPDase (GS52)			
E182A	ACR3	Inactive	(Tanaka, et al., 2011a)
D209A	ACR3	Inactive	
S214A	ACR4	Inactive	
Q216A	ACR4	Inactive	

Table 1-1 Site mutagenesis of NTPDases.
Taken from (Kirley, et al., 2006; Tanaka, et al., 2011a)

Mutagenesis studies in plants and parasitic microbes have been largely based on the mammalian findings which show the same residues are fundamental for catalytic activity. Residues that have been identified in the *Glycine soya* GS52: Glu182, Asp209, Ser214, Trp216 (Tanaka, et al., 2011a), and for the *Legionella pneumophila* NTPDase1: Arg122, Glu159, Asn168, Gln193 and Trp384 (Sansom, et al., 2008).

Generally characterised plant and parasitic microbe NTPDases have 2-3 disulfide pairs rather than the 5 identified in mammals and point mutations of the conserved 10 cysteines in the human NTPDase3 identified 3 disulfides essential for correct folding and catalytic activity (Cys261S/Cys308S, Cys289S/Cys343S, Cys399S/Cys422S). These correspond to those conserved in plant and microbe NTPDases (Ivanenkov, et al., 2005; Roberts, et al., 1999).

1.5.5 NTPDase activity-Substrate specificity

Generally the NTPDase family of enzymes have the ability to hydrolyse a broad range of nucleotide di- and triphosphates. Although the efficiency at which hydrolysis occurs for each nucleotide differs. When a range of nucleotides are tested, the NTPDases show moderate specificity for the nucleobase (as shown for the soluble rat NTPDase1 and *L. pneumophila* NTPDase1) or, nucleotide di- or triphosphate (as shown for human and mouse NTPDase2). In some cases there is no discernable preference for nucleotides as observed for membrane bound human and mouse NTPDase1 (Kukulski, et al., 2005; Sansom, et al., 2008; Zebisch, et al., 2007). The specific activity range of host purified NTPDases varies considerably depending on the species and the tissue from which the sample was collected. For example the rat NTPDase2 from skeletal muscle T-tubules has specific activities as high as 6,000 $\mu\text{mol}\cdot\text{min}^{-1}\cdot\text{mg}^{-1}$ whereas the same enzyme from the chicken gizzard has a catalytic specificity as 88.6 $\mu\text{mol}\cdot\text{min}^{-1}\cdot\text{mg}^{-1}$ (Stout, et al., 1994; Treuheit, et al., 1992). Other reported specific activities for ATP from various species include; plant purified pea and potato NTPDase with a specific activity of 428.6 $\mu\text{mol}\cdot\text{min}^{-1}\cdot\text{mg}^{-1}$ and 10,000 $\mu\text{mol}\cdot\text{min}^{-1}\cdot\text{mg}^{-1}$ (Chen, et al., 1987;

Handa, et al., 1996), respectively. *Toxoplasmosa* NTPDase1 and 2 has reported specific activities of 2540 $\mu\text{mol}\cdot\text{min}^{-1}\cdot\text{mg}^{-1}$ and 446 $\mu\text{mol}\cdot\text{min}^{-1}\cdot\text{mg}^{-1}$, respectively (Asai, et al., 1983).

NTPDase	Tissue	Substrate	Specific activity ($\mu\text{mol}\cdot\text{min}^{-1}\cdot\text{mg}^{-1}$)	Reference
NTPDase1	Human placenta	CaATP	15-20	*Makita
	Porcine brain cortex	CaATP	69.4	Kul
NTPDase2	Rabbit skeletal muscle	MgATP	6,600	(Treuheit, et al., 1992)
	Chicken gizzard	MgATP	88.6	(Stout, et al., 1994)
	Porcine brain cortex	CaATP	79.6	(Kukulski, et al., 2003)
NTPDase8	Rat liver	CaATP	19.1	
	Rat liver	CaATP	454	
Pea nuclei NTPase		MgATP	428.6	(Chen, et al., 1987)
Potato apyrase		CaATP	10,000	(Handa, et al., 1996)
<i>Toxoplasma</i> NTPDase1		MgATP	2,540	(Asai, et al., 1983)
<i>Toxoplasma</i> NTPDase2		MgATP	446	(Asai, et al., 1983)

Table 1-2 Specific activities of naturally sourced NTPDases.
Adapted from (Knowles, 2011)

The specific activity of a number of the cell surface membrane bound NTPDases is intrinsically linked to the transmembrane domains as the application of detergents to human NTPDase 2 and 8 (which both contain 2 transmembrane domains) reduces the specific activity to as low as 3 % and 9 % compared with the wild-type, respectively . Also, the removal of the transmembrane domains in the chicken NTPDase8 and the human NTPDase1 both reduced the specific

activity (Li, et al., 2010; Wang, et al., 1998). In addition to lowering the activity the removal of the transmembrane domains from cell-surface NTPDases also alters several biochemical features such as the ADPase/ATPase ratios, Michaelis-Menten kinetics for substrates, pH-activity-curves, and whether the product is released from the enzyme (Chiang, et al., 2008; Grinthal, et al., 2002; Li, et al., 2010; Wang, et al., 1998).

1.5.6 Michaelis–Menten kinetics of NTPDases

The Michaelis–Menten kinetics have been studied for better characterised NTPDases including human and mouse NTPDase1-3, 8, potato apyrase and *L. pneumophila* NTPDase1. The Michaelis constant, K_M for ATP and ADP varies between various reports. For example in the case of the CHO cell expressed full-length rat NTPDase1-3 the K_M for ATP has been reported to range from 76 μM to 311 μM , whereas the K_M for ATP for the COS cell expressed mouse and human NTPDase1-3 is 17 to 75 μM and 21-35 μM (Iqbal, et al., 2006; Kukulski, et al., 2005). Using the prokaryotically expressed and refolded rat NTPDase1-3 with the extracellular domain cleaved showed a K_M for ATP of between 22.6 μM and 125 μM . The Michaelis constant for the potato NTPDase for ATP varied between 25 μM and 60 μM depending on which variety the NTPDase was isolated from (Desirée or Pimpernel). Prokaryotically expressed LpNTPDase1 has a reported K_M of 400 μM for ATP. It is noteworthy that LpNTPDase1 has a very high K_M of 1,000 μM for ADP.

The k_{cat} values for NTPDases are often not reported. The k_{cat} constant describes the amount of substrate that is turned over per catalytic site per unit time (typically per second). The k_{cat} values that have been reported are for prokaryotically expressed rat NTPDase1-3 which ranged from 34.3 s^{-1} to 255.2 s^{-1} for ATP hydrolysis and 24.2 s^{-1} to 95.9 s^{-1} for ADP hydrolysis (Zebisch, et al., 2007). However more commonly reported is the k_{cat}/K_m ratio which describes the overall catalytic efficiency of the enzyme typically reported in units of $\text{M}^{-1}\cdot\text{s}^{-1}$. The values that have been reported include NTPDase1 and -2 isolated from

porcine brain cortex synaptosomes, which have k_{cat}/K_m ratio of 9.0×10^6 and 3.0×10^6 for ATP, that are comparable to the prokaryotically expressed rat NTPDase1-3 values 1.1×10^6 and 4.9×10^6 for ATP, respectively (Kukulski, et al., 2003; Zebisch, et al., 2007). NTPDases isolated from potato have a large range of k_{cat}/K_m ratios, depending on from what potato variety the enzyme was isolated; 0.3×10^6 to as high as 67.4×10^6 for ATP (Kettlun, et al., 2005; Kettlun, et al., 1982). The prokaryotically expressed *L. pneumonia* exhibited a high k_{cat}/K_m ratio for ATP hydrolysis of 1.2×10^8 and for ADP hydrolysis of 7.9×10^7 demonstrating that this enzyme is very efficient at hydrolysing ATP and ADP (Sansom, et al., 2008).

Species		K_M (μM)	k_{cat} (s^{-1})	k_{cat}/K_M ($\text{M}^{-1}\cdot\text{s}^{-1}$)	Reference
Rat (<i>E.coli</i>)	NTPDase1	125	255.2	2.0×10^6	(Zebisch, et al., 2007)
	NTPDase2	30.3	34.3	1.1×10^6	
	NTPDase3	22.6	111.5	4.9×10^6	
Rat (COS)	NTPDase1	76	N/R	N/R	(Iqbal, et al., 2006)
	NTPDase2	203	N/R	N/R	
	NTPDase3	311	N/R	N/R	
Mouse	NTPDase1	12	N/R	N/R	(Kukulski, et al., 2005)
	NTPDase2	37	N/R	N/R	
	NTPDase3	11	N/R	N/R	
Human	NTPDase1	17	N/R	N/R	
	NTPDase2	70	N/R	N/R	
	NTPDase3	75	N/R	N/R	
Potato	Desirée	70	1439	20.6×10^6	(Kettlun, et al., 1982)
apyrase	Pimpernel	250	79	0.3×10^6	
<i>Legionella pneumonia</i>	LpNTPDase1	400	N/R	1.2×10^8	(Sansom, et al., 2008)

Table 1-3 Michaelis-Menten kinetics of various NTPDases.
N/R represents data not reported.

1.6 Research objectives

In legumes the NTPDases have been associated with the early stages of rhizobium symbiosis. Nodules form and are placed into different types: indeterminate or determinate. During development, indeterminate nodules generally form in temperate legumes including white clover and pea. Nodules form in the root cortex and induce the division of the meristematic outer cortical cells. The determinate nodules generally form in tropical and sub-tropical legumes (including soy-bean and *Dolichos*). These nodules are formed in the outer root cortex and growth occurs by cell expansion rather than cell division.

In this project two NTPDases were chosen from each nodulation type in legumes, white clover (7WC) and *Dolichos biflorus* (DbLnP). 7WC has yet to be shown to be involved in nodulation, but is localised in the root where nodulation occurs. DbLnP for years has been known to be directly involved in nodulation.

Previous research showed that these enzymes could only be expressed as inactive inclusion bodies in *E.coli*. To overcome this problem refolding trials were conducted to produce a high yield of active enzyme that could be used for further research.

To better understand the function of the legume NTPDases a catalytic mechanism was investigated utilising enzyme activity assays to determine the biochemical profile including substrate specificity and Michaelis-Menten kinetics. A third objective was to determine high resolution structures of NTPDases from both legumes. Attempts to solve the protein structures were performed using X-ray crystallography. In addition, to aid the understanding of the binding mechanism of substrates to NTPDases, attempts to solve the structures with the non-hydrolysable substrate or AMP were performed. The determination of the protein structures would provide a better understanding of the catalytic and binding mechanism of NTPDases in legumes and would give essential clues as to the physiological function of these enzymes.

2 General methods

2.1 Reagents

All reagents if not stated were obtained from either Merck KGaA (Germany), BDH Chemicals Ltd (UK), Becton, Dickson and Co. (France), Scharlau (Spain) or Fluka (Germany).

2.2 Bioinformatic analysis

Bioinformatic analysis, including sequence alignment was carried out using either Vector NTi® (Invitrogen, USA) or geneious PRO™ (New Zealand).

2.3 DNA manipulation

2.3.1 Gene constructs

Plasmid DNA of 7WC in the expression vector pET28 was supplied by Chung Hong Chen (Agresearch Grasslands, New Zealand)(Chen, 2008). The gene construct of DbLnP (Accession number: AF1398007) was supplied by Genart® (Germany) and truncated based on the mature peptide (Etzler, et al., 1999).

2.3.2 DNA plasmid extraction

Plasmids were isolated using the QIAprep miniprep (QIAGEN, Netherlands) procedure that uses the principles of a modified alkaline lysis method (Birnboim, et al., 1979). Single colonies were isolated by streaking *Escherichia coli*, from a glycerol storage stock on to LB-agar (5 % NaCl, 10 % yeast extract, 10 % Bacto-tryptone, and 7.5 % agar) and incubating over night at 37 °C. Single colonies were transferred to 5 ml of LB broth and incubated overnight at 37 °C with mixing. Plasmid DNA was extracted using the manufacturer's protocol. The plasmid DNA was eluted in 50 µl MilliQ.

2.3.3 DNA electrophoresis gel

DNA fragments were prepared for electrophoresis by resuspension in 10 x DNA loading dye (50 % glycerol, 0.25 % xylene cyanol FF and 0.25 % bromophenol blue) and were separated via electrophoresis in 1 % agarose gel in a TAE

electrophoresis buffer (40 mM Tris-acetate, pH 8.5, 2 mM EDTA). DNA fragment size was determined by comparing the bands with a 1 kB plus DNA ladder (Invitrogen, USA). The gel was visualised under UV light after incubation with 0.5 $\mu\text{g}\cdot\text{ml}^{-1}$ ethidium bromide after electrophoresis.

2.3.4 DNA fragment gel extraction

DNA fragments were isolated using the QIAquick gel extraction kit as per the manufacturer's protocol. Gel fragments were eluted using 50 μl of 10 mM Tris-HCl, pH 8.6.

2.3.5 DNA quantification

DNA concentration was determined by two methods. Generally the DNA was quantified using the Nanodrop ND-1000 spectrophotometer (Analytical technologies, Biolab, UK) that measures the DNA absorbance at 230 nm. The nanodrop method was not appropriate after DNA fragment extraction so DNA quantification was determined by comparing the band intensities after electrophoresis with DNA mass ladders (Invitrogen, USA) at various dilutions.

2.3.6 Restriction enzyme digest of DNA plasmid

Restriction enzyme digest was performed using Invitrogen *Eco* R1 or *Nco* 1 according to manufacturer's protocol for 4 hr at 37 °C. The restriction enzymes were inactivated by incubation at 65 °C for 20 min.

2.3.7 DNA ligation

Ligation of isolated DNA fragments was performed using T4 DNA ligase (Invitrogen, USA) according to the manufacturer's protocol with the addition of 0.5 μl of 10 mM ATP per 20 μl reaction mixture. Prior to ligation the plasmid fragment was dephosphorylated to reduce self ligation of the plasmid by incubating the plasmid DNA with shrimp alkaline phosphatase (Invitrogen, USA) according to the manufacturer's protocol.

2.3.8 DNA sequencing

DNA sequencing was performed at the Alan Wilson Centre sequencing facilities (Massey University, Palmerston North, New Zealand).

2.3.9 DNA transformation

DNA was transformed into *E. coli* DH5 α TOP10 competent cells (Invitrogen, USA) or *E. coli* RosettaTM(DE3) competent (Novagen, Germany) using the heat shock method with 10 μ l ligation mix per 50 μ l aliquot of competent cells. To select for correctly ligated DNA fragments, the cells were plated onto LB-agar containing 60 mg \cdot L⁻¹ kanomycin. Plasmids from colonies that grew were analysed by DNA sequencing to confirm that cloning was successful (section 2.3.8).

2.3.10 *E. coli* storage stocks

For long time storage of *E. coli*, 70 μ l of DMSO was added to a 1 ml of an overnight culture and stored at -80 °C until required.

2.4 Protein expression

2.4.1 Growth of cells for protein expression

E. coli was streaked onto a LB agar plate containing an appropriate antibiotic and incubated overnight at 37 °C. A single colony from a freshly streaked LB plate was used to inoculate a 10 ml starter culture 50 ml Falcon tube (BD Bioscience, USA) containing an appropriate antibiotic and incubated at 37 °C with shaking. For small scale expression the 1 ml of starter culture was used to inoculate a 10 ml culture whereas for large scale protein expression, 10 ml of the starter culture was transferred to a 1 L LB broth containing in a 5 L conical flask. In both cases the LB contained an appropriate antibiotic and were shaken at 200 rpm (small scale) or 140 rpm (large scale) over night.

To induce protein expression, the culture was induced with 1 mM IPTG when the culture reached an absorbance of 0.6 at 600 nm.

2.4.2 Harvesting the cells of test cultures

Pelleted cells were resuspended in small scale lysis buffer which consisted of 25 mM Tris-HCl, pH 7.5, 1.0 mM EDTA, 2.0 mM MgCl₂, 0.1 % Triton X-100, 0.1 mg·ml⁻¹ Lysozyme, 0.01 mg·ml⁻¹ RNase and 0.05 mg·ml⁻¹ DNase. The suspension was placed on ice for 1 hr before centrifugated at 15,000 x·g for 20 min. The supernatant was collected as the soluble fraction. The recovered pellet was washed as in 2.6.1.

2.4.3 Harvesting of large scale expression culture

Cells were recovered by centrifugation and resuspended in resuspension buffer and sonicated. Careful consideration was made to ensure that resuspended cells didn't get too hot during sonication by keeping the suspension on ice.

2.5 Large scale purification using fast liquid chromatography (FPLC)

All FPLC was conducted on either the AKTA™ Prime or the AKTA™ Basic and controlled using the UNICORN™ interface (GE Healthcare, Sweden).

2.5.1 IMAC in denaturing conditions

Denatured immobilised metal affinity chromatography was carried out using a 5 ml HisTrap HP column (GE Healthcare, Sweden). The IMAC column was equilibrated with 5 CV of MilliQ and 5 CV of buffer A containing 100 mM Tris-HCl, pH 7.5, 150 mM NaCl, 8 M urea and 10 mM imidazole before use.

The prepared solubilised inclusion bodies (2.6.1) were filtered through a Minisart 0.45 µm syringe filter (Sartorius stedimbiotech, Germany) before loading onto an equilibrated HisTrap HP column. The column is then washed with 3-5 CV of buffer A until the absorbance reached zero. The His-tagged protein was eluted from the column by passing a gradient over 5-10 CV of buffer B that comprised of 100 mM Tris-HCl, pH 7.5, 150 mM NaCl, 8 M Urea and 300 mM imidazole. Examination of the UV trace indicated which fractions to collect.

The column was routinely stripped of the Ni²⁺ ions with 5 CV of TBS containing 20 mM EDTA and charged with 1 CV of 100 mM NiCl₂.

2.5.2 Size exclusion chromatography

Size exclusion chromatography was performed with either the HiLoad™ Superdex 200 16/60 or the Superdex 200 10/300 GL (GE Healthcare, Sweden). Columns were equilibrated with 1 CV MilliQ and 1 CV run buffer. The samples were concentrated to 0.5 ml or 5 ml and loaded onto the Superdex 200 16/60 or the Superdex 200 10/300, respectively and added to a pre-equilibrated column. The protein was eluted with 20 mM Tris, 150 mM NaCl, pH 7.5 at 0.5 ml·min⁻¹ - 1 ml·min⁻¹ for 1 CV. The size of the eluted protein was estimated by examination of the UV trace and comparing the elution volume with a standard curve produced by reference proteins (GE Healthcare, Sweden).

2.5.3 Cation exchange

Cation exchange chromatography was carried out using a 5 ml HiTrap SP HP column (GE Healthcare, Sweden). The column was equilibrated with 10 CV MilliQ followed by 5 CV of buffer A containing 50 mM sodium acetate, pH 5.0, 5 CV of buffer B containing 50 mM sodium acetate, pH 5.0, 2 M NaCl then 10 CV of buffer A. The sample was loaded onto the column at 1 ml/min. The column was washed with 2 CV of 5 % buffer B before a two step with 5-45 % buffer B gradient over 10 ml to elute the target protein, then a 45-100 % buffer B gradient over 5 ml to remove all bound protein.

2.6 Refolding of denatured protein

2.6.1 Isolation and preparation of inclusion bodies

A large scale expression culture was pelleted by centrifugation for 30 min at 5,000 x·g and the cells were resuspended in 40 ml resuspension buffer containing 100 mM Tris-HCl, pH 7.5 and 150 mM NaCl. Cell contents were extracted using a sonicator ultra sonic processor XL 2020 (Misonix incorporated, USA) for 5 x 1 min with a 1 min rest between each pulse, using a power level of 7. The inclusion bodies were isolated by centrifugation 30 min at 5,000 x·g, the supernatant containing the soluble fraction was disposed of and the pellet containing the inclusion bodies was recovered. The inclusion bodies were subsequently washed

to remove any contaminating soluble proteins by resuspending in the same volume of resuspension buffer containing TBS and 0.1 % Tween-100 and centrifuged for 30 min at 5,000 x-g. This process was repeated 3 times with the last resuspension in TBS. The washed inclusion bodies were resuspended in solubilisation buffer (100 mM Tris-HCl, 150 mM NaCl, 8 M Urea, 2 mM DTT, 10 mM imidazole) and mixed end-over-end for 16 hr at 25 °C. Solubilised inclusion bodies were isolated by centrifugation at 40,000 x-g for 30 min and the supernatant recovered and routinely purified by IMAC before further experiments (section 2.5.1).

2.6.2 High throughput method for refolding

A flat bottom greiner bio-one (greiner bio-one, Germany) 96 well plate with a well depth of 300 µl was used. Each well contained 100 µl of one of the 92 buffers (four wells are reserved as negative controls containing MilliQ or are empty wells). Each well had a final volume of 200 µl (plus the added protein). For the compounds of the refolding plate see appendix (section 8.2).

To each well, 2 µl of solubilised protein was added to a final concentration of 0.05 mg·ml⁻¹. A negative control was performed by adding 2 µl of solubilisation buffer without protein to a refolding plate. The plates were incubated at 15 °C. Correct refolding was evaluated in two ways; the first was evaluating protein aggregation, the other was to measure the activity of the potentially refolded protein. Aggregation of unfolded protein can give an indication of incorrect folding. This was done so by simply measuring the absorbance at 355 nm and comparing with a negative control plate containing no protein.

To measure the activity, 20 µl of a 10 x buffer (1 M Tris-HCl, 750 µM MgCl₂) and 5 µl of 100 mM ATP was added and the plates were incubated at 25 °C for 30 min. Developing dye (section 2.7.7) was added and left at room temperature for 10 min before the absorbance was read at 620 nm. The test plate was normalised with a negative control plate containing no protein.

2.6.3 Small scale method for refolding

To test parameters for optimal refolding such as fine screening components, temperature, shaking trials and refolding time, a small scale refolding protocol was developed. For refolding 15 μl of IMAC prepared protein, speed for 5 sec and incubated at 18 °C unless stated otherwise. concentrated to $\sim 5 \text{ mg}\cdot\text{ml}^{-1}$, was placed on the lid of a 1.5 ml Eppendorf tube (Eppendorf, Germany) containing 1.5 ml of ice-cold refolding buffer. The lid was gently closed so the protein remained on the lid. The tube was vortexed at maximal

2.6.4 Large scale refolding using the dilution method.

Before the solubilised protein was added the refolding buffer cooled in an ice bath to reduce any rapid refolding. IMAC purified denatured protein was added using a peristaltic pump at $7 \text{ ml}\cdot\text{min}^{-1}$ to 1 L of ice cold, de-gassed refolding buffer in a conical flask. During the addition of solubilised protein the buffer was mixed as fast as possible using a flea and a mixing plate without the surface being disturbed. The IMAC purified 7WC and DbLnP was diluted in refolding buffer to a maximum of $50 \mu\text{g}\cdot\text{L}^{-1}$ to reduce aggregation and incubated at 18 °C for 7 days without being disturbed.

2.7 Protein analysis

2.7.1 Native-PAGE

The non-reducing PAGE was used to analyse native proteins. These were made in two parts using the recipe in Table 2-1; the resolving gel was typically made with 10 % acrylamide throughout most of the project although some 12 % gels were used, the stacking gel was made with 5 % acrylamide. Samples were prepared by mixing with 4 x native protein loading dye (125 mM Tris-HCl, pH 6.8, 50 % glycerol and 0.1 % bromophenol blue) before loading and the gel was electrophoresed in native gel running buffer (25 mM Tris, 192 mM glycine and 0.1 % SDS) at 80 V for 20 min or until the dye band went through the stacking gel and then 150 V for 75 min or until the dye front runs completely through the resolving gel.

A)		
Volume in ml	10%	12%
RO water	12	10
30 % acrylamide	10	12
Resolving buffer (1.5 M Tris, pH 8.8)	7.5	7.5
10 % APS	0.15	0.15
TEMED	0.015	0.015

B)	
Volume in ml	5%
RO water	8.5
30 % acrylamide	2.125
Stacking buffer (1.0 M Tris, pH 6.8)	1.6
10 % APS	0.063
TEMED	0.0063

Table 2-1 Components of the native PAGE gels.

A) The separating and the B) stacking gel.

The gel could be kept for one month at 4 °C when wrapped in a moist paper towel inside a plastic page. The running buffer contained 25 mM Tris, pH 8.5 and 192 mM glycine. The orientation of the electrodes was chosen based on the isoelectric point for each protein. The migration of the sample was compared with Bio-Rad precisions plus protein standards (BioRad, USA) as a standard. Bands were labelled I to VI.

2.7.2 SDS-PAGE

For analysis of denatured proteins the 10 % non reducing gels were used but the preparation of the samples and the running buffer used differed. The samples were prepared by adding reducing loading dye (125 mM Tris-HCl, pH 6.8, 50 % glycerol and 0.1 % bromophenol blue, 10 mM mercaptoethanol and 10 % SDS) and incubating for 5 min at 95 °C before loaded. The gel was run in SDS running buffer (25 mM Tris, pH 8.5 and 192 mM glycine and 1 % SDS) in the same way as

the native PAGE (section 2.7.1). The migration of the sample was compared with Bio-Rad precision plus protein standards (Bio-Rad, USA) to estimate the molecular weight.

2.7.3 Staining PAGE gels

2.7.3.1 *Coomassie stain*

Before staining, the gel was immersed in hot RO water and shaken for 2 min to remove any SDS from the gel. The gel was stained by microwaving the gel for 30 sec on high in coomassie stain (0.006 % coomassie blue-G, 30 mM HCl) and shaken gently for 5-10 min before discarding the stain. The gel is destained by microwaving the stained gel for 30 sec on high in RO water and gently shaken for 1-16 hours before visualisation.

2.7.3.2 *Silver stain*

Before staining, the gel was immersed the gel in hot RO water and shaken for 2 min to remove any SDS from the gel. The gel was immersed in fixing solution (50 % methanol, 12 % acetic acid, 0.05% formalin) and shaken overnight at room temperature. The gel was washed 3 x 10 min in 20 % methanol before sensitising solution (0.002 % sodium thiosulfate) was added and left shaking for 2 min. The sensitising solution was discarded and the gel washed with MilliQ for 1 min. Cold silver stain (2% silver nitrate, 0.002% sodium thiosulfate) was added and incubated for 20 min while being shaken. Gel was rinsed in MilliQ and developed with developing solution (6 % sodium carbonate, 0.004 % sodium thiosulfate, 0.05 % formalin). Developing typically took 3-10 min before being stopped using 12 % acetic acid.

2.7.4 Protein concentration measurement

2.7.4.1 *Nanodrop method*

The nanodrop method for determining protein concentrations was used if the expected protein concentrations was over 1 mg·ml⁻¹.

The protein concentration was measured using the Nanodrop ND-1000 (Thermo Scientific, USA). Absorbance at 280 nm was measured using 1-2 μl of sample. The concentration was determined from absorbance using the Beer's Law (section Equation 4) which describes the linear relationship between absorbance at 280 nm and protein concentration.

$$A=\epsilon\cdot c\cdot l$$

Equation 4 Beer's law.

Where A =Absorbance at 280 nm, ϵ =the theoretical extinction coefficient of the protein ($\text{M}^{-1}\cdot\text{cm}^{-1}$) and l =length of the path (cm).

The extinction coefficient is based on the number of aromatic residues and was estimated using ExPASy ProtParam (<http://ca.expasy.org/tools/protparam.html>).

2.7.4.2 *Bradford assay*

For the Bradford assay, an adapted method for use with a 96 well microplate was used using the Bio-Rad protein assay dye (Bio-Rad, USA). The following mixture was made: 40 μl of Bio-Rad protein assay dye, 1-25 μl protein and was made up to 200 μl with MilliQ. The reaction was mixed and incubated at room temperature for 5-10 min before measuring the absorbance at 595 nm using a Flourostar Optima microplate reader (BMG labtech, Germany). The test reading was compared against a standard curve using a range of BSA concentrations.

2.7.5 Concentrating protein

2.7.5.1 *Using protein spin concentrators*

For small scale concentrating of proteins; 0.5 ml, 2 ml and 20 ml Vivaspin ultracentrifugation devices with a molecular weight cut off of 30,000 were used (Sartorius group, Germany). The ultracentrifugation devices were centrifuged at 10,000 $\times g$ (0.5 ml) or 4,000 $\times g$ (2 ml and 20 ml) until the desired concentration was achieved.

2.7.5.2 *Using stir-cell concentrator*

After large scale refolding using the dilution method the refolded protein was concentrated using a 400 ml Millipore stirred ultrafiltration cell (Millipore, USA)

with an Amicon polyethersulfone (PES) 76 mm filter with a molecular cut-off at 30,000 Da (Millipore, USA). To pressurise the cell, nitrogen gas was used at 70 psi. The cell was kept cool using ice packs and typically took 2-3 days to concentrate to 50 ml. Once this volume was reached the protein was typically concentrated further using spin concentrators (section 2.7.5.1). Due to the long time to concentrate, the cell was kept at 4 °C during the night to reduce protein denaturation.

2.7.6 Buffer exchange via dialysis tubing

Sample was added into pre-wet Spectra/Por® tubing with a MWCO of 6000-8000 Da, ends closed and dialysed in dialysis buffer (~100-fold of the sample volume) overnight at 4 °C. The dialysis buffer was replaced at least once during the dialysis period.

2.7.7 Protein activity analysis

The hydrolysis of nucleoside di/triphosphate was measured by detection of free phosphate using a malachite green assay modified for the 96 well microtitre plate (Baykoz, et al., 1988). The assays were carried out at 25 °C in 100 µl and unless specified contained 100 mM Tris, pH 8.0, 750 µM CaCl₂, 0.02 µg·µl⁻¹ protein. The reaction was started with 0.2 mM substrate. After 30 seconds to 5 minutes the reaction was stopped with malachite solution (1.5 % ammonium molybdate, 0.18 % Tween 80, 0.1 % malachite green and 14 % sulfuric acid). After 10 min the absorbance of each assay was measured at 620 nm. The molar concentration of phosphate in the reaction assay was achieved by comparing against a NaPO₄ standard (0-100 nM). All assays were performed in triplicate and standardised against a blank assay containing no protein.

2.7.7.1 Protein activity analysis for calcium saturation curve.

To determine the half binding efficiency (EC₅₀) for the metal cation the NTPDase activity assay was performed with a range of divalent cation concentrations (0-800 µM). Data were analysed using Graphpad Prism software which calculated the EC₅₀ using a non-linear regression line of best fit.

2.7.7.2 Protein activity analysis for optimal pH

To determine the optimal pH for substrate hydrolysis the activity assay was conducted for each substrate with buffers at various pH. These included; 100 mM sodium acetate-acetic acid at pH 3.5-6.0, 100 mM HEPES at pH 6.5-8.0 and 100 mM glycine-NaOH at pH 8.5-10.5.

2.7.7.3 Determination of the initial rate.

To measure the initial rate of NTPDase activity the enzyme assay was performed for various periods of time (0-600 sec) and the amount of liberated phosphate ions were measured. The initial rate was determined by measuring the gradient of the linear region of the liberated phosphate vs. time plot and dividing this value with 10 % of the total concentration of phosphate ions in the assay (Equation 5). The total phosphate ion in the assay is 50 and 100 μ Mole for NDP and NTP, respectively.

$$t = \frac{\Delta y}{[\text{PO}_4^-]_{\text{total}}/10}$$

Equation 5 Determination of the time required for 10 % of the total substrate to be hydrolysed.

Symbols represent t =time (sec), PO_4^- total= 100 μ M for NTP, 50 μ M for NDP, Δy = gradient of the linear region of the (liberated PO_4^- /time) plot.

2.7.7.4 Michaelis-Menten kinetic analysis

To produce a Michaelis-Menten plot the NTPDase activity assay was conducted with various concentrations of substrates (0 - 1 mM). Data were analysed using Graphpad Prism software which calculated the K_M , V_{max} , and the using a non-linear regression line of best fit. It was assumed that there was only one active site so the k_{cat} was calculated by dividing the V_{max} value with the molar concentration of the enzyme. Since the product of the hydrolysis of NTP to NDP is also catalysed careful considerations were made to measure the initial rates. It was assumed that the measured NDP was negligible during these assays.

2.7.8 In-gel activity assay

After native-PAGE, the gel was left to incubate without shaking in the native running buffer containing 20 mM CaCl₂ and 10 mM ATP. After 1 hr the insoluble Ca(H₂PO₄)₂ was visualised. As a positive control the tested samples were compared with a commercially sourced apyrase from potato (Sigma-Aldrich, USA) which was run alongside samples.

2.7.9 Small scale thrombin cleavage trials of DbLnP

Small scale optimisation of thrombin cleavage was performed to determine the optimal ratio of thrombin to DbLnP. Thrombin was diluted with thrombin dilution buffer (50 mM sodium citrate, pH 6.5, 200 mM NaCl, 0.1 % PEG 8000, 50 % glycerol) to contain 0.04, 0.02, 0.01, and 0.005 U per assay. Each 50 µl assay contained 20 mM Tris-HCl, 150 mM NaCl, 2.5 mM CaCl₂ and 50 µg of size exclusion purified DbLnP. The reaction was incubated at room temperature (~25 °C) and 10 µl aliquots were removed and mixed with 4 x SDS running buffer after 2, 4 and 8 hrs. The extent of the cleavage was evaluated using SDS-PAGE comparing them with an uncut DbLnP.

2.7.10 Large scale thrombin cleavage

Large scale cleavage reaction was proportional to the small scale and was typically a total of 5-6 ml which was sufficient to cleave 1 mg of DbLnP. The reaction was incubated longer (4 hr) than determined in the small scale reaction to ensure that cleavage had occurred.

2.7.10.1 Purification of thrombin cleaved DbLnP

To remove the cleaved His-tag and any un-cleaved DbLnP, the reaction solution was mixed end over end with nickel charged sepharose resin (equilibrated in 50 mM Tris-HCl, pH 7.5 and 150 mM NaCl, 50 mM imidazole) for 20 min. To isolate cleaved DbLnP from the thrombin, the resin was centrifuged at 500 xg for 5 min and the supernatant was recovered.

2.7.10.2 *Site directed mutagenesis*

Site directed mutagenesis was performed using QuikChange™ kit (Agilent Technologies, Inc. (USA) as per manufacturer's protocol. Primers were developed using the Agilent primer design server (<http://www.genomics.agilent.com>).

2.8 Protein crystallography

2.8.1 General methods

After the final purification step, protein was concentrated as described in 2.7.5.1 as high as possible without the protein aggregating. This was typically ~10 mg · ml⁻¹ for proteins in this study. The protein was then prepared by centrifugation at 13,000 x-g for 5 min at 4 °C to remove any large particulates from the solution. The crystallisation trial was periodically observed using a Nikon stereomicroscope (model SM2800, Nikon Corporation, Japan). The drops were scored according to the following scoring system; 0. Clear drop, 1. Non-protein particles, 2-5. A gradient of precipitation where 5 is the greatest, 6. Phase separation, 7. Spherulites, 8. Needles, 9. Plates, 10. 3-dimensional crystals.

2.8.2 Initial crystallisation trials-robot

A Multiprobe® II HTE_x robot (PerkinElmer, USA) was used to dispense 75 µl of each condition into the wells of a 96-well Intelli-plate (Art Robbins Instruments, USA). The conditions were based on a high-throughput method of 480 conditions developed to minimise the protein used (Moreland, et al., 2005).

After the protein was prepared (section 2.8.1) initial screens were conducted using the Cartesian™ honey bee crystallisation system (Cartesian™ Dispensing system, USA) to dispense 100 nl of protein with 100 nl of precipitant solution in the sitting-drop 96-well Intelli-plates. The plates were sealed using ClearSeal film™ (Hampton Research, USA).

These trials were performed at the Maurice Wilkins Centre for Molecular Biodiscovery crystallisation facility at the University of Auckland.

2.8.3 Optimisation of crystallisation conditions

Potential conditions identified were further optimised either using the hanging drop or sitting drop method. All trials were carried out at 18 °C on shock-resistant shelves. Fine screens were produced by modifying the pH or the concentration of the components identified in the initial screen.

2.8.3.1 *Hanging drop method for crystallography*

The top of the wells of a VDX plate (Hampton Research, USA) was greased using glisseal®N (Borer chemie, Switzerland) and 500 µl of each condition was pipetted into each well. To the surface of a square siliconised cover slip 1-2 µl of protein was added to equal amount of mother liquor and the cover slip inverted to the top of the pre-greased well.

2.8.3.2 *Sitting drop method for crystallography*

The top of the wells of a VDX plate (Hampton Research, USA) were greased using glisseal®N (Borer chemie, Switzerland) and 500 µl of each condition was pipetted into each well. The sitting drop method was performed by pipetting 1-2 µl of protein with an equal amount of mother liquor onto a Micro-Bridge (Hampton Research, USA). To close the well, a glass cover-slip was placed on the top of the pre-greased well.

2.8.4 Seeding

Before seeding was performed, the drop was left 2-3 days to allow the drop to equilibrate with the reservoir solution. Seeds were transferred to the drop with either a cat whisker or a cryo-loop (Hampton Research, USA) (thoroughly washed in ethanol).

2.8.4.1 *Micro-seeding*

A seed stock was made if enough crystals were available and prepared fresh each time. The seed stock was made by transferring a drop with crystals into a 200 µl Eppendorf tube containing 100 µl of mother liquor. The crystals were aspirated with a pipette and crushed using a small pipette tip. The stock was centrifuged

(1,000 x·g, 2 min) and the supernatant containing the micro-seeds were recovered. The seed stock was diluted with the mother liquor if too many crystals formed after seeding.

2.8.4.2 *Streak-seeding*

The cat's whisker was gently streaked across a pre-formed crystal to pick up small protein crystal fragments and lightly run through the equilibrated drop. The cat's whisker was typically passed through 2-3 identical drops to dilute the crystal particles transferred.

2.8.4.3 *Macro-seeding*

Small crystals grown in the same conditions were transferred using cryo-loops into equilibrated drops.

2.8.5 Additive screening

The additive screen based on selected components from the Hampton additive screen HT (Hampton Research, USA). Crystallisation was performed using either the hanging or sitting method (section 2.8.3.1, 2.8.3.2).

2.8.6 IZIT dye

To determine if the crystal was made of salt or protein, 0.2 µl Hampton IZIT dye (USA) was added to a 2 µl drop. The drop was visualised after 1 hr. If the crystals turned dark blue the dye is able to pass through the large solvent channels not present in salt crystals.

2.8.7 Co-crystallisation with substrate

Co-crystallisation trials were conducted using either the hanging or the sitting drop methods (section 2.8.3). Because of the expensive reagents used in these trials, 10 µl mother liquor solutions including the substrate were made exclusively for the drop and were mixed with equal amounts of the mother liquor. The reservoir solution contained the mother liquor without substrate.

2.8.8 Substrate soaks

To soak substrate into the crystal, it was transferred using a Hampton cryo-loop into 2 μl of mother liquor containing desired substrate on a siliconised cover-slip. The cover slip was inverted onto a pre-greased VDX plate (Hampton Research, USA) containing 500 μl of mother liquor. The crystals were typically soaked overnight at 18 °C.

2.9 Preparation for data collection

2.9.1 Cryo-protectants

If required, cryo-protectant was added to the mother liquor. In this study PEG 3350, MPD and glycerol were routinely used. To test the cryo-protectants, mother liquor containing the cryo-protectant was collected in cryo-loops and flash frozen in liquid nitrogen. The cryo-loops were exposed to X-ray radiation and the resulting diffraction pattern was examined for ice rings at 1.93 Å, 2.07 Å, 2.26 Å, 2.68 Å, 3.45 Å, 3.67 Å, and 3.92 Å. The lowest concentration of cryo-protectant that did not result in ice rings was used.

2.9.2 Flash freezing

The pre-formed crystals were transferred to mother liquor solution containing cryo-protectant with a cryo-loop (Hampton Research, USA). Crystals were transferred to 3-4 drops for 30 sec each containing increments of cryo-protectant until the desired concentration was met. After the addition of cryo-protectant, the crystal was immediately immersed in liquid nitrogen and was either loaded into a SSRL automated mounting cassette (Crystal Positioning Systems, USA) or into CrystalCap cryo-vial (Hampton Research, USA).

2.10 Diffraction data collection

2.10.1 Home source data collection

Home source data collection was performed at the School of Biological Sciences at the University of Auckland. The X-rays were generated using $\text{CuK}\alpha$

radiation ($\lambda=1.5418 \text{ \AA}$) from a rotating copper anode Micro-MaxTM-007HF generator (Rigaku, Japan) and reflections were detected using a Mar345 detector (Mar Research GmbH, Germany). The loops were mounted on a MAR234dtb goniometer (Mar Research GmbH, Germany) and kept cold with a stream of liquid nitrogen from a Cobra cryosystem (Oxford Cryosystems, UK). Prior to data collection the MOSFLM's strategy function was used with 2 images 90° apart to aid in data collection.

2.10.2 Synchrotron data collection

Data were collected at the Australian Synchrotron, Melbourne, Australia using either MX1 or MX2 beam-line. Reflections were measured on an ADSC Quantum 315r detector (Area Detector Systems Corp., USA). Prior to data collection the MOSFLM's strategy function was used with 2 images 90° apart to aid in data collection.

2.11 X-ray data processing

2.11.1 Indexing with MOSFLM

The program MOSFLM was used to visualise, index images and integrate the reflections (Leslie, 1992). The program auto spot finder and auto-index functions were used to determine the cell parameters. The cell parameters were refined using 2 sets of 4 successive images 90° apart and the reflections prior to integration. MOSFLM was used to integrate the images, paying particular attention to the changes in RMS deviation and mosaicity during integration.

2.11.2 Beam centre estimation

The beam centre was estimated by submitting images to the Lawrence Berkley LABELIT server (<http://adder.lbl.gov/labelit>).

2.11.3 Scaling and merging of reflections

The images space group was confirmed using the POINTLESS program within CCP4 prior to merging the images using SCALA within the CCP4 program (Bailey, 1994). The output using all the reflections was examined for unfavoured

R_{merge} , I/σ_i , data completeness and typically re-scaled to optimise these parameters. For refinement confirmation in REFMAC 5.0, 5% of the reflections were marked as the R_{free} dataset.

2.11.4 Analysis of unit cell using Matthew's coefficient

To determine the number of protein molecules in the asymmetric unit, MATTEWS_COEF program in CCP4 was performed. The results were examined and selected based on a solvent content of 35-60%.

2.11.5 Phase determination by molecular replacement

Molecular replacement was achieved using either MOLREP or PHASER within CCP4 or PHENIX (McCoy, et al., 2007; Vagin, et al., 1997). The search model sequence was aligned and modified by pruning non-conserved residues back to the γ -atom using CHAINSAW in CCP4 (Stein, 2008). If the space group was unknown all symmetry related space groups were tested.

2.11.6 Model building and refinement

2.11.6.1 Automated building

Automated building was performed using the ARP/wARP program within CCP4 (Langer, et al., 2008). ARP/wARP was performed with automated model building starting from experimental phases.

2.11.7 Manual building

Manual building was achieved using COOT (Emsley, et al., 2004). The model was built into the $2|F_o| - |F_c|$ and $|F_o| - |F_c|$ maps contoured to 1σ and 3σ , respectively.

2.11.7.1 Model refinement

Model refinement was performed with either Refmac 5.0 within CCP4 (Murshudov, et al., 1997) or phenix.refine within PHENIX (Afonine, et al., 2005). Within both programs particular attention was paid to the R and R_{free} values to validate model building. Refmac 5.0 was performed with either restrained refinement or with input of TLS parameters (section 2.11.8.3).

The phenix.refine program was exclusively performed with the simulated annealing function using the Cartesian method with a starting temperature of 5000 K and reducing in 100 K steps to a final temperature of 300 K. Water molecules were added once refinement could not be improved further using the 'Add water' function in Refmac 5 and checked manually using COOT (Emsley, et al., 2004). Hydrogen bonds were assessed using the MolProbity server (<http://molprobity.biochem.duke.edu/>) (Davis, et al., 2007) and manual inspection using KiNG (Chen, et al., 2009).

2.11.8 Structural analysis

Images of structures in this thesis were produced using Pymol (DeLano, 2002) or UCSF chimera (Pettersen, et al., 2004). Topology maps were generated using the PDBsum server (<http://www.ebi.ac.uk/pdbsum/>).

2.11.8.1 *B-average analysis*

The B-average analysis was performed using Baverage program within CCP4 (Bailey, 1994).

2.11.8.2 *Ramachandran analysis*

Ramachandran analysis was performed using MolProbity either via the server (<http://molprobity.biochem.duke.edu/>) or through PHENIX (Adams, et al., 2010; Davis, et al., 2007).

2.11.8.3 *TLS analysis*

The TLS analysis was performed using TLSMD web server (<http://skuld.bmsc.washington.edu/~tlsmd>). The determination of rigid domains was performed using isotropic analysis. Rigid domains were identified by the observation of an acute angle in the residual number vs. domain number plot. The output from TLS analysis was used for further TLS refinement.

TLS parameters were visualised as ellipsoids by firstly processing the TLSMD results with the TLSANL program within CCP4 (Howlin, et al., 1993). Results from the TLSANL were used in UCSF chimera (Pettersen, et al., 2004) to produce images.

2.11.8.4 *Secondary structure comparison*

For comparing two structures for visualisation, Structural Similarity matching algorithm in the COOT or Pymol program was used (Bailey, 1994; DeLano, 2002). For structural similarity analysis between two structures, the DALILite Pairwise server (<http://www.ebi.ac.uk/Tools/dalilite/>) was used (Holm, et al., 2000). For structural comparison with the Protein Data Base (<http://www.pdb.org/pdb/home/home.do>), DALILite version 3 was used (http://ekhidna.biocenter.helsinki.fi/dali_server) (Holm, et al., 2000).

2.11.8.5 *Hinge region analysis*

Protein domain motion analysis was performed using the DynDom server (<http://fizz.cmp.uea.ac.uk/dyndom>) to identify hinge regions (Lee, et al., 2003). Domain size was restricted to 20 residues during analysis.

2.12 Mass spectrometry

Mass spectrometry was performed by Johnathan Puddick at the Brunner Daltonics Centre of Excellence, University of Waikato using an amaZon™ X mass spectrometer and data visualised using Compass DataAnalysis 4.0 (Brunner Daltonics, USA). Samples were prepared by spiking with 0.1 % formic acid or 5 mM NaOH (final concentrations shown).

3 Expression, refolding and purification of 7WC and DbLnP

3.1 Introduction

For biochemical and structural studies of the legume NTPDases it was necessary to produce large quantities of protein in an efficient manner. For this objective the open reading frames of the NTPDases were transformed into an *E. coli* based expression system. The expressed protein contained a His-tag which aided in the isolation of the expressed protein by immobilised metal affinity chromatography (IMAC). This system is a common method to produce large quantities of protein suitable for subsequent biochemical analysis.

3.2 Cloning and expression of 7WC

3.2.1 Cloning of mature 7WC peptide

The expression construct for 7WC was supplied by Chung Hong Chen (Chen, 2008). For cloning the mature 7WC, Chen selected the N-terminal residues based on the alignment of the mature native DbLnP and the mature native potato apyrase Figure 3-1. The N-terminal residues for both DbLnP and the potato apyrase are 5 residues upstream of the ACR1 (Chen, 2008).

Pot_apy1	-----MLNQNSHFIFILLAFLVLP-LSLLSKNVNAQIPLRRHLL---SHESEH	YAVIFDAGS	TG
7WC	MPTLYKKAGSAAAPFTMEFLIKLITFLI FLIPTITSQYLGNNLFTNRKIFLKQ	GTITSYAVIFDAGS	TG
DbLnP	---MNVVWPKTKSMSFLLLITFLIFSLPKLSSSQYV	GNSILLNHRKILPNQEI	ELLTSYAVIFDAGSSG

Figure 3-1 Protein sequence alignment of soluble plant NTPDases.

Residues that are within red box represent signal peptide. The blue box represents the first 5 residues in the mature peptide

The gene construct of the mature 7WC peptide was cloned into the pET28 expression vector and transformed into *E. coli* BL21 (DE3) Rosetta strain (section 8.1). Comprehensive expression trials were carried out with differing expression vectors and expression hosts (Chen, 2008). However 7WC was expressed as insoluble inclusion bodies.

3.3 Cloning and expression of DbLnP

3.3.1 Cloning of mature DbLnP peptide

The mature DbLnP peptide gene construct was supplied by GENEART (section 8.1). The sequence was optimised by GENEART for expression in *E. coli*. In this case, their software provides increased RNA stability by removing destabilising sequence elements. GENEART also inserted *NdeI* and *XhoI* restriction enzyme cleavage sites at the N- and C-terminus, respectively. The DbLnP sequence was cloned into the expression vector pET28 so that expression of this construct would have a His-tag followed by a thrombin cleavage site before the DbLnP translation start site. This was achieved by digesting the isolated GENEART plasmid DNA and the pET28 plasmid with the restriction enzymes *NdeI* and *XhoI*. The cut fragments were isolated by agarose gel electrophoresis. The insert fragment (DbLnP) 1,259 bp and the pET28 fragment, 5,296 bp, were excised from the gel Figure 3-2.

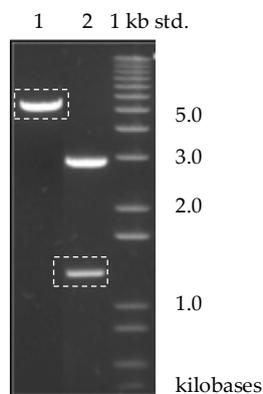


Figure 3-2 Agarose electrophoresis of the restriction digested DNA. The pET28 vector (lane 1) and GENEART DbLnP gene (lane 2). Dashed boxes indicate the excised fragments

Ligation of the isolated DNA fragments was performed using the T4 DNA ligase and subsequently transformed using heat shock into competent *E. coli* DH5 α TOP10 cells. Selections of correctly transformed cells were isolated by plating the cells on LB agar with 50 mg·L⁻¹ kanamycin. To ensure that the transformation was successful a selection of colonies were further cultured and plasmids were extracted, digested with restriction enzymes *NdeI* and *XhoI* and analysed using

agarose gel electrophoresis. Plasmids that showed correct fragment size were further assessed by sequencing the plasmids using the T7 forward and reverse primers. For protein expression the pET28:DbLnP gene construct was transformed as above into *E. coli* BL21 (DE3) Rosetta competent cells.

3.3.2 Initial expression trials of DbLnP

Small scale expression trials were conducted to determine if soluble expression of DbLnP was possible and the time post induction for DbLnP to express. Cultures were grown in LB media (section 2.4.2) and induced with IPTG. 1 ml samples were taken at 0, 1, 3 and 16 hr after induction and cells were isolated using centrifugation and were subsequently resuspended in lysis buffer. The lysed cells were centrifuged and the supernatant was collected (the soluble fraction) and the pellet was also collected (the insoluble fraction).

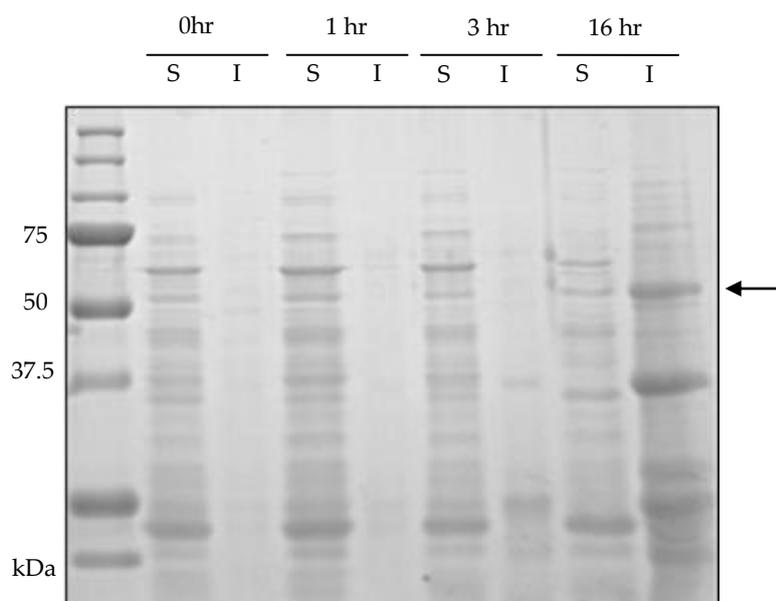


Figure 3-3 Small scale expression trial of pet28:DbLnP. Arrow indicates the expected molecular weight of DbLnP. S=soluble protein, I=insoluble protein.

The soluble and insoluble fractions were evaluated by SDS-PAGE analysis. It is difficult to determine if there is expression in the soluble fractions due to a similarly sized protein expressed without induction. Although this band remains at a similar intensity over the expression period and suggests that there is no or

very little expression of soluble DbLnP. At time zero the insoluble fraction had no observed expression suggesting that IPTG is required for expression. Expression of a protein with the predicted molecular weight was detected after 16 hr post induction in the insoluble fraction.

Due to the success in refolding NTPDases and the difficulty in expressing soluble NTPDases, expression trials to encourage soluble expression of DbLnP were not conducted.

3.4 Large scale expression and solubilisation of 7WC and DbLnP

3.4.1 Considerations for large scale expression and solubilisation

For large scale protein preparation expression cultures of 1 L were grown at 37°C over night to maximise the amount of insoluble NTPDase expression. The inclusion bodies were isolated and washed before solubilisation (section 2.6.4).

Inclusion bodies represent aggregated proteins. For effective refolding the protein is solubilised by adding a denaturant containing high concentration of a chaotropic salt. Typically solubilisation of inclusion bodies is achieved with 8 M urea or 6 M guanidine hydrochloride.

Inclusion bodies often contain di-sulfide bonds, which may interfere with the refolding process (Clark, et al., 1999; Fischer, et al., 1993). During solubilisation it was also important to ensure that the disulfide bonds are reduced as disulfides can persist while in 8 M urea (Schoemaker, et al., 1985). To ensure that disulfide bonds were reduced 2 mM of the reducing agent DTT was added in a slightly alkaline condition (Fischer, et al., 1993). The concentration of DTT was kept as high as possible without reducing the metal ions during IMAC purification.

3.4.2 IMAC chromatography in denaturing conditions of 7WC and DbLnP

Prior to refolding the solubilised legume NTPDases were purified using denatured IMAC purification. 7WC and DbLnP eluted from the nickel charged sepharose with approximately 700 mM and 300 mM imidazole, respectively (Figure 3-4, Figure 3-5).

Substantial 7WC and DbLnP eluted from the column as flow through which was recovered for future IMAC chromatography (Figure 3-4, Figure 3-5-peak 1). This could be repeated 3 times before most of the protein was recovered.

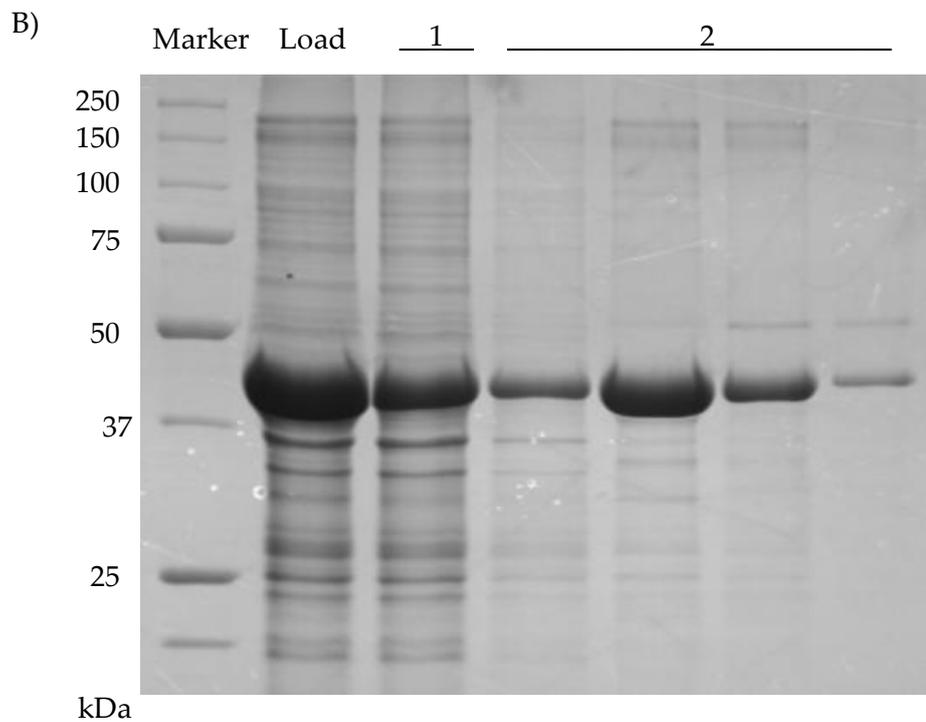
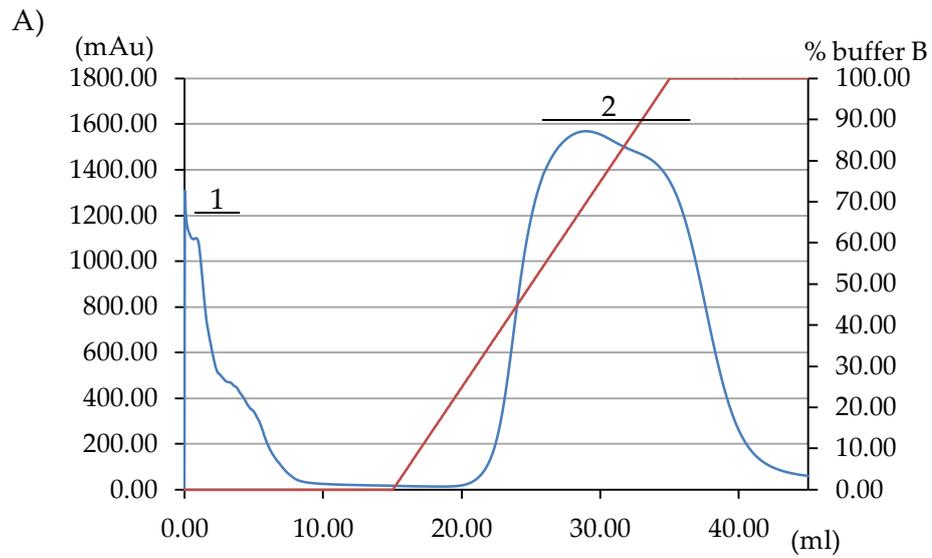


Figure 3-4 IMAC purification of 7WC in denaturing conditions.

A) The FPLC trace showing the absorbance in blue and percentage of buffer B containing 1 M imidazole in red. B) SDS-PAGE analysis of IMAC fractions.

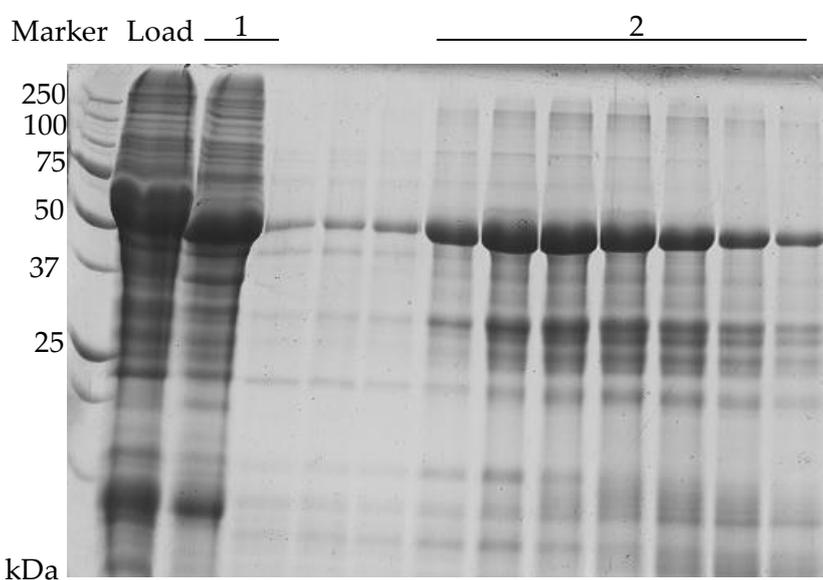
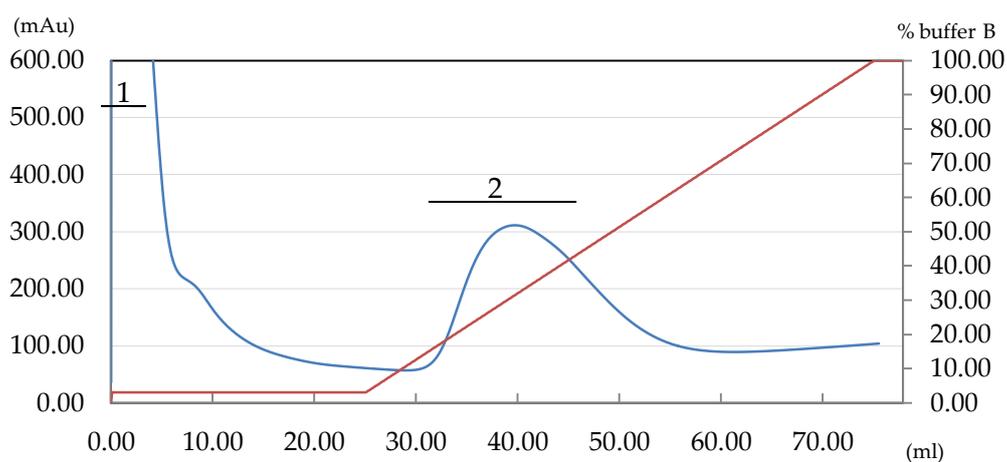


Figure 3-5 IMAC purification of DbLnP in denaturing conditions. A) The FPLC trace showing the absorbance in blue and percentage of buffer B containing 1 M imidazole in red. B) SDS-PAGE analysis of IMAC fractions.

3.5 Refolding

3.5.1 Introduction

A problem associated with over expressing non-host proteins in *E. coli*-based expression systems is the propensity for these to be incorrectly folded and sequestered as inactive insoluble aggregates or inclusion bodies. Active protein can be recovered from inclusion bodies by firstly solubilising the inclusion bodies in a high chaotropic salt solution followed by *in-vitro* refolding. The

renaturation of the protein requires conditions and techniques to correctly fold the protein to the native form.

Past research on the refolding of prokaryotically expressed NTPDases has shown that much of the protein where refolding was attempted formed structured soluble aggregates (Chen, 2008). Soluble aggregates can form when the unfolded protein (U) folds in to various intermediate forms (X_1 to X_n). Some of these intermediates are able to form soluble aggregates in a reversible manner or can subsequently form an insoluble aggregate which precipitates from solution. Those intermediates that do not aggregate can continue to correctly fold to an active protein molecule (N) (Figure 3-6). The rate of the aggregation is largely based on the protein concentration and the conditions of the refolding buffer.

In the past the dialysis method was employed, however this resulted in a very low yield of native protein (Chen, 2008). To refold the legume NTPDases the dilution method was chosen over the dialysis method to reduce protein aggregation (Clark, et al., 1999). The dilution method reduces the protein concentration quickly thus reducing the chance that refolding intermediates meet and aggregate. In contrast, the dialysis method reduces the denaturant slowly which reduces the rate of refolding therefore increasing the ratio of intermediates. Also, the dialysis method maintains a high protein concentration increasing the occurrence of intermediate aggregation.

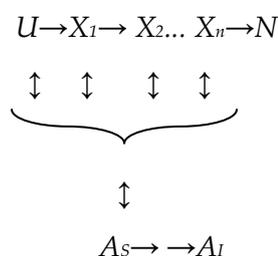


Figure 3-6 Schematic of how protein aggregates.
 Figure adapted from (Pain, 1995).

It can be assumed that the active protein is in a low energy state, however a local minimum may be found in the case of intermediates and hence additional free

energy may be required to overcome energy barriers for correct protein folding to occur. Hence considerations of temperature and time are essential parameters to optimise when developing a refolding protocol.

3.5.2 Refolding matrix

A large number of factors have an impact on correct refolding of proteins from their denatured forms. Defining correct conditions for each protein is empirical and difficult. Due to these reasons a high throughput protocol that used a 96-well microplate format based on the dilution method for refolding was developed (Vincentelli, et al., 2004). The components in the conditions are based on an extensive review of the REFOLD database (<http://refold.med.monash.edu.au>) (Chow, et al., 2006). Each well contained a unique combination of pH, salts, redox reagents and other additives proven to promote refolding. For the refolding of the NTPDases, refolding conditions were modified to specifically to refold NTPDases based on the refolding data previously obtained (Chen, 2008) and the physiological conditions at which NTPDases are thought to refold *in planta*. Extensive details of the buffers and additives were screened for 7WC and DbLnP are listed in appendix (section 8.2).

3.5.2.1 Selection of the refolding buffers for 7WC

The best condition previously reported for refolding 7WC was 50 mM acetic acid, pH 5.0, 10 % glycerol and 1 M arginine using the dialysis method (Chen, 2008). This protocol resulted in the successful refolding of 7WC with a very low yield as most of the protein precipitated. Also, native gel analysis and whole gel elution analysis of a native gel revealed that the refolded 7WC solution contained inactive soluble aggregates where a monomeric 7WC was the only active form. Based on the measured protein concentration and specific activity before and after refolding the yield of the active monomeric species is approximately 1.4 % of the initial protein before refolding. For crystallisation it is important to

produce high amounts of protein and so further steps were required to optimise this protocol.

Denatured IMAC purified 7WC in 8 M urea was concentrated to 5 mg·ml⁻¹ to minimise the amount of urea in the refolding condition and added to the refolding buffer described in section 2.6.2.

3.5.2.2 Determining the efficiency of refolding 7WC based on activity.

It can be assumed that if the refolded enzyme is active then correct refolding must have occurred. To determine the efficiency of refolding the catalytic activity was measured by adding 10 x activity buffer (500 mM Tris-HCl and 750 μM MgCl₂, pH 8.0). The activity buffer also contained the divalent metal co-factor, MgCl₂, required for efficient hydrolysis of substrate (Kettlun, et al., 1982; Tognoli, et al., 1981). The assay was started with ATP, which has been shown previously to be the optimal substrate for 7WC. The reaction was measured as described in section 2.7.7 and the measurements were subtracted from a blank microplate plate containing only buffer and no protein. A summary of the refolding results are shown in Table 3-1.

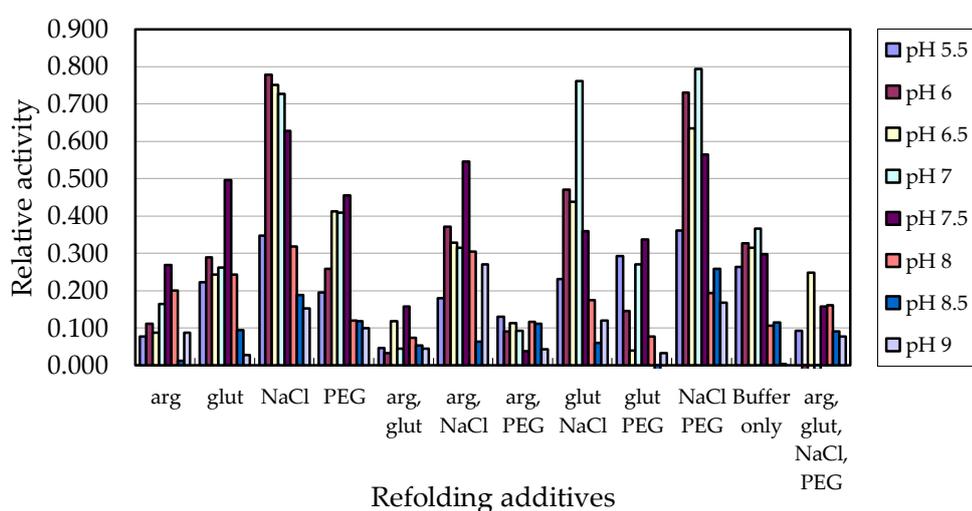


Figure 3-7 Activity assay of 7WC after refolding.

Components are as follows; arg. 50 mM arginine, glut. 5 mM:1 mM reduced:oxidise glutathione, NaCl. 150 mM NaCl, PEG. 0.2 % PEG 4000. Buffers are at 50 mM. pH 5.5-6.0. sodium acetate, pH 6.6-7.0. MES, pH 7.5-9.0. Tris-HCl

Highest peak #	pH	Components
1	6	150 mM NaCl
2	6	150 mM NaCl, 0.2 % PEG 3350
3	6.5	150 mM NaCl
4	6.5	150 mM NaCl, 0.2 % PEG 3350
5	7	150 mM NaCl
6	7	150 mM NaCl, 5 mM:1 mM red:ox glutatione
7	7	150 mM NaCl, 0.2 % PEG 3350

Table 3-1 Refolding conditions that showed the highest activity for 7WC.

It was clear that the refolding conditions that gave the best activity were all within pH 6.0 and 7.0 with the best at pH 6.0. The addition of 150 mM NaCl and 0.2 % PEG 3350 typically gave 2-fold increases in activity over the entire pH range when compared with buffer only.

Conditions containing arginine consistently showed low activity, although this may be due to the high amount of free phosphate present in the arginine detected in the negative control. It is possible that the amount of phosphate due to the background and the hydrolysis product was higher than the detection limit. If this is the case then this data would be an under-estimation of the protein activity. Further trials with arginine showed that this was not the case. These results contradict that of Chung Hong Chen's results although the buffers and the different refolding method used in this study could account for these differences (Chen, 2008).

3.5.2.3 *Determining the efficiency of refolding 7WC based on aggregation*

In general there is a negative correlation between aggregation and correct refolding. So the extent of aggregation after refolding was assessed by measuring the absorbance after refolding at 355 nm. Proteins in solution do not absorb light at this wavelength and only if the protein aggregates and increases the turbidity light is scattered resulting in an increase in absorbance (Leach, et al., 1960).

The results show, where a high absorbance indicates high aggregation, there is a relationship between the pH and protein aggregation (Figure 3-8). In this case, 7WC aggregates most at pH 6.5-7.0 which corresponded with the conditions that gave high activity (Table 3-1). There is a clear trend whereby all samples that contained arginine had no measurable turbidity despite the pH. This result is not unexpected as arginine is thought to increase the free energy of protein-protein interactions reducing aggregation (Arakawa, et al., 2003; Baynes, et al., 2005).

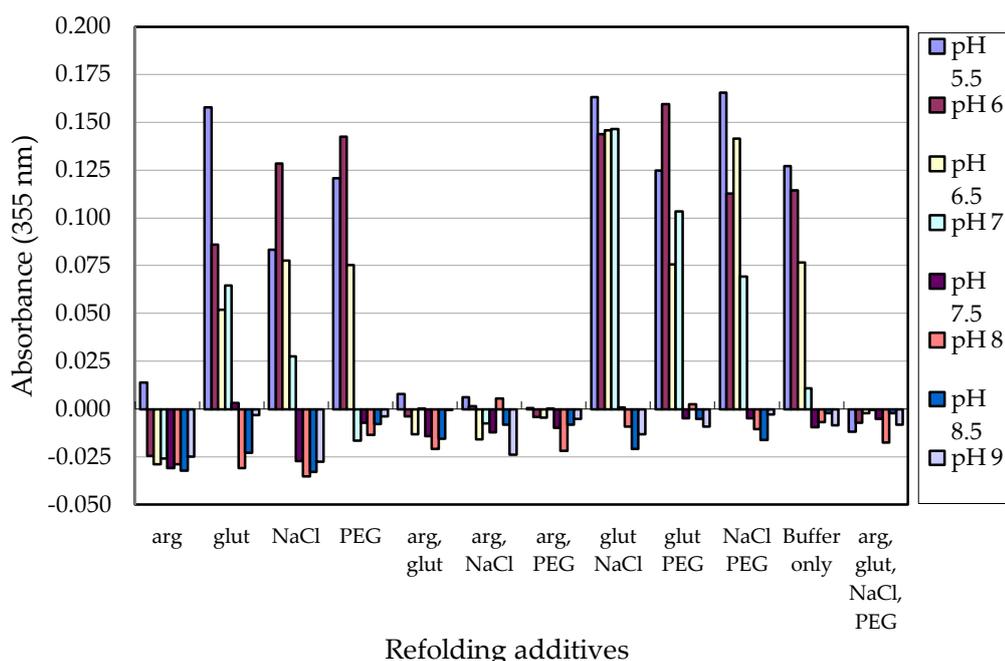


Figure 3-8 Aggregation assay of 7WC after refolding. Components are as follow; arg. 50 mM arginine, glut. 5 mM:1 mM reduced:oxidise glutathione, NaCl. 150 mM NaCl, PEG. 0.2 % PEG 4000. Buffers are at 50 mM. pH 5.5-6.0. sodium acetate, pH 6.6-7.0. MES, pH 7.5-9.0. Tris-HCl

3.5.2.4 Fine screening of the refolding conditions for 7WC

Results from the activity and aggregation assays did not correspond well (Figure 3-7 and Figure 3-8). Further iterations of the refolding conditions were based on positive results from the activity assays by optimising the concentration of the additives and evaluated by enzyme activity. Other additives were also tested such as a glutathione oxido-shuffling system which contains differing ratios of reduced and oxidised glutathione. This is typically used for appropriate disulfide bonds forming during refolding. The requirement for such a system is not surprising as 7WC contains 3 pairs of disulfide bonds (section 1.3.2).

The final refolding buffer for 7WC was 50 mM MES, pH 6.5, 700 mM NaCl, 0.6 % (w/v) PEG₃₃₅₀, 0.5 mM oxidised glutathione, 1 mM reduced glutathione.

3.5.3 Selection of the refolding buffers for DbLnP.

No previous research had been conducted on refolding DbLnP inclusion bodies thus, a wider range of additives were screened. This included β -cyclodextrin, which sequesters detergents, various NaCl concentrations, redox reagents such as DTT and glutathione oxido-shuffling system, PEG 4000 and urea. The total lists of components are shown in appendix (section 8.2).

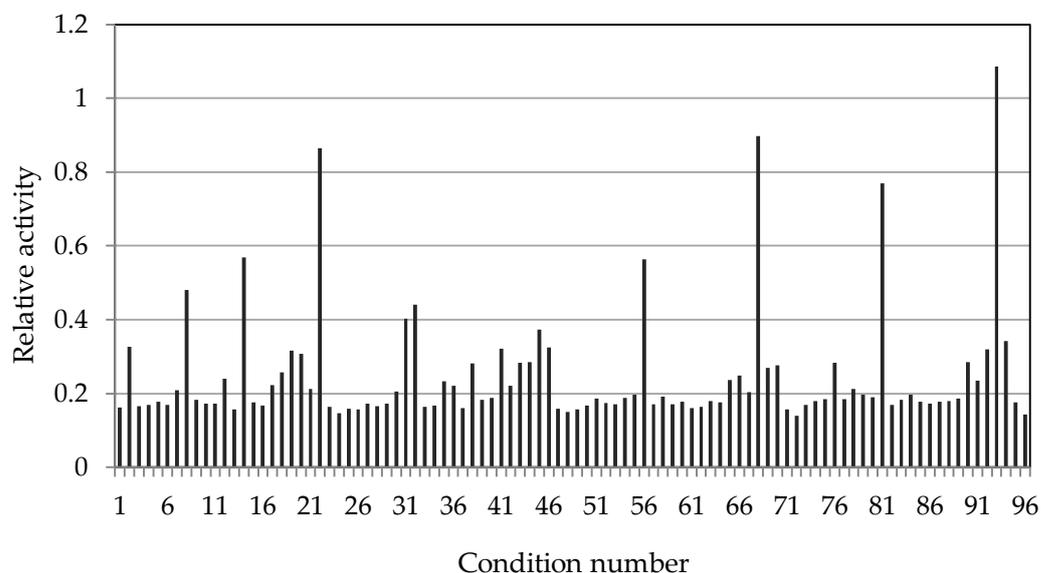


Figure 3-9 Evaluation of protein activity after refolding DbLnP. The conditions are shown in appendix (section 8.2)

3.5.3.1 Determining the efficiency of refolding DbLnP based on activity

DbLnP activity was measured in the same way (see section 3.5.2.2). The results showed a small number of conditions that were significantly more effective at refolding DbLnP (Table 3-2).

Condition number	pH	Components
93	7.0	250 mM NaCl, 500 mM arginine
68	6.5	250 mM NaCl, 500 mM arginine
22	5.5	150 mM NaCl, 500 mM arginine, cyclodextrin
81	7.0	500 mM Arginine, 1 mM EDTA
56	6.6	500 mM Arginine

Table 3-2 The top refolding conditions for DbLnP based on activity.

The top three conditions all contain 500 mM arginine and 250 mM NaCl and are in the range of pH 5.5-7.0. These results suggest that these components are important and were optimised further in subsequent refolding trials.

3.5.3.2 Determining the efficiency of refolding based on aggregation for DbLnP

The degree of aggregation after refolding DbLnP was evaluated (section 3.5.2.2.). The results are summarised in Table 3-3 which showed the conditions that led to the lowest absorbance at 340 nm.

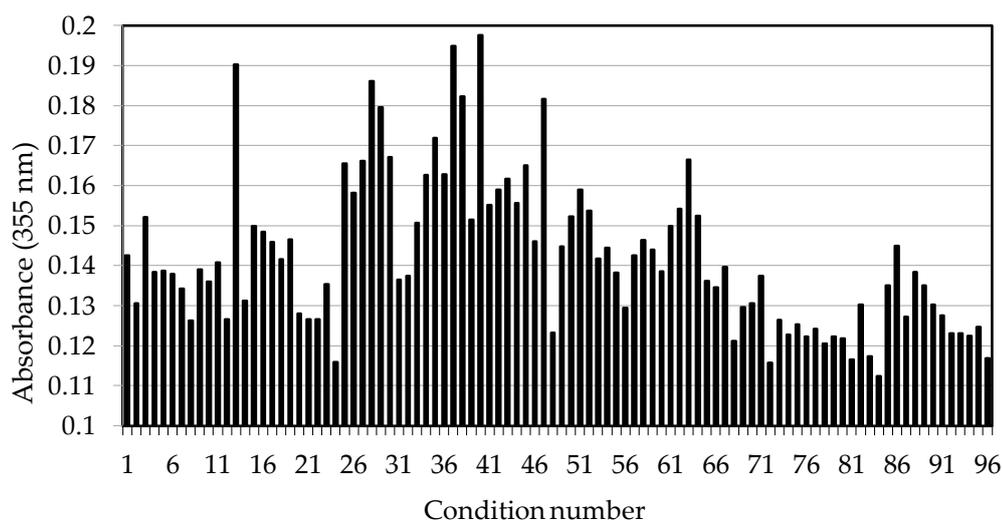


Figure 3-10 Aggregation of DbLnP after refolding

Condition number	pH	Components
84	7.0	100 mM NaCl, 600 mM urea, 12.5 mM cyclodextrin
81	7.0	500 mM arginine, 1 mM EDTA
83	7.0	100 mM NaCl
78	7.0	250 mM NaCl, 1 mM DTT

Table 3-3 Summary of the top condition by aggregation for refolding DbLnP

Refolding conditions that resulted in the lowest amount of aggregated protein are all at pH 7.0 and contain some quantity of NaCl.

Analysis of the aggregation test show a consistent trend with the best conditions are all at pH 7.0 and most contain NaCl. Low aggregation conditions correlate well with conditions that gave high activity.

3.5.3.3 *Fine screening of the refolding conditions for DbLnP*

The components identified that gave the highest activity and the lowest aggregation were subsequently optimised using small-refolding trials (section 2.6.3). The optimal buffer was found to be 50 mM Tris-HCl, pH 7.0, 500 mM NaCl and 250 mM arginine.

3.5.4 Additional parameters for optimal refolding for 7WC and DbLnP.

For optimal refolding the correct stable state that represents the correctly folded protein may have to go through energy barriers. Hence evaluating temperature that provides such energy is important. Midi refolding trials (1.5 ml) were carried out at various temperatures. Assessed by activity assay, the optimal temperature for refolding for both 7WC and DbLnP was ~15 °C (Figure 3-11). Large scale refolding was carried out at 18°C because of the availability of a temperature controlled room.

Other trials revealed that the refolding protein prefers to be still and not shaken while refolding was taking place. In the shaken samples visible precipitated protein was seen as a turbid solution after refolding.

The 'shuffling' of disulfides using the oxido-suffling system and overcoming energy barriers of structured intermediates may take a long time to occur and thus a time course experiment was conducted over 300 hr (Figure 3-12). Evaluation of the refolding efficiency was performed by measuring the specific activity and this showed that maximal activity was reached after 150 hr for both 7WC and DbLnP. These data lead to development of a standard protocol for refolding 7WC and DbLnP.

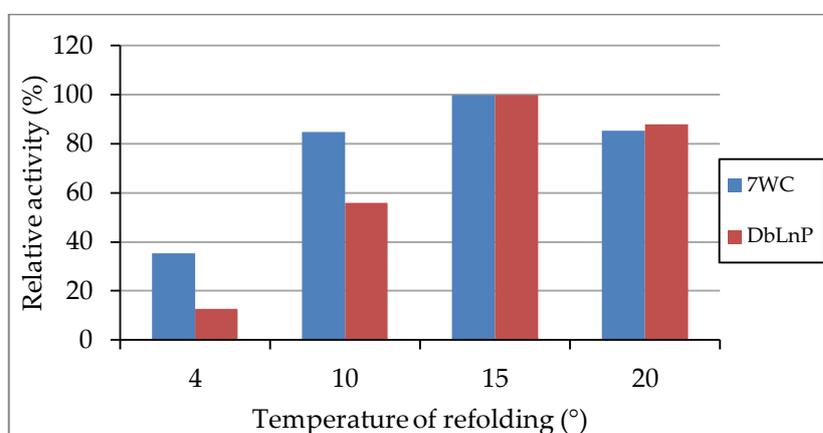


Figure 3-11 Activity analysis of refolding at various temperatures for 7WC and DbLnP.

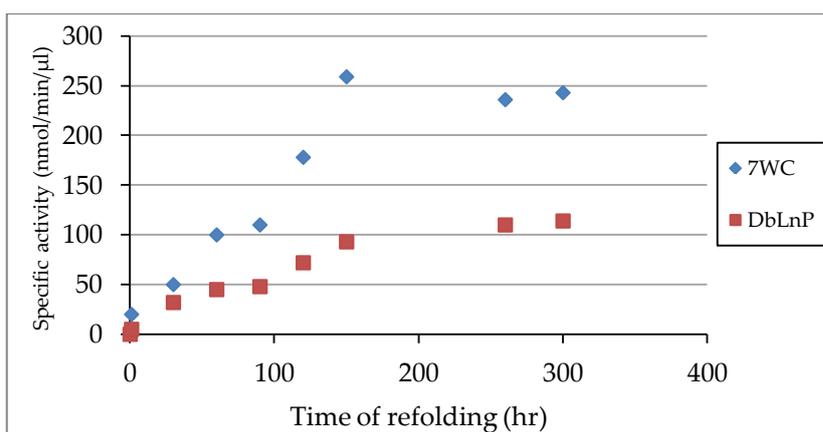


Figure 3-12 Activity analysis of refolding with increasing time for 7WC and DbLnP.

3.5.5 Large scale refolding preparation of 7WC and DbLnP

In addition to the components identified from small scale refolding experiments (section 3.5.2.4 and 3.5.3.3) further components and the preparation of the buffer were taken into consideration. The refolding buffer was thoroughly degassed and 1 mM EDTA was added. This was performed to inhibit the activity of metalloproteases that originate from the expression host and reduce the oxidation of thiols caused by molecular oxygen and metal ions (Misawa, et al., 1999). To inhibit bacterial and fungal growth 0.002 % sodium azide was added to the buffer.

3.5.6 Analysis and purification of refolded 7WC

Refolded 7WC was concentrated using the stir cell to ~ 50 ml (section 2.7.5.2). SDS-PAGE analysis of the refolded protein showed that only 7WC was present in solution however a native PAGE analysis revealed ~ 6 different sized proteins that formed discrete bands (termed laddering hereafter). These results suggest that after refolding 7WC is present in a number of multimeric forms.

3.5.6.1 *In-gel activity assay.*

An in-gel activity assay was developed to evaluate the activity of NTPDases after native PAGE (adapted from Zebisch, 2007). This method is based on the phenomenon that phosphate liberated from hydrolysed ATP can form insoluble calcium phosphate. 7WC was analysed by PAGE and visualised by either coomassie or silver stain and compared with the gel based activity assay. These data show that the fastest migrating band, that migrates similar to the (V) band of the marker, represents the active form of 7WC (Figure 3-8). These results are consistent with previous findings that used whole gel elution (Chen, 2008). Analysis of the potato apyrase was used as a positive control. The protein migrated through the native gel as a smear between I and III. The enzyme activity corresponded to the protein detected by silver stain analysis.

3.5.6.2 *Removing the soluble aggregate.*

For crystallisation and kinetic characterisation it is important to remove the active species from the inactive protein. This was achieved by dialysing the sample overnight at 4 °C in 50 mM acetate, pH 5.0. In this condition the active form was found to be stable whereas the inactive soluble aggregate precipitated from solution. The precipitated protein was removed by centrifugation at 40,000 x·g for 30 min at 4 °C and the pellet discarded. The sample analysed by native PAGE show that after the dialysis in 50 mM acetate, pH 5.0 the laddering is not observed by silver stain. The band representing the active form is only just visible in the silver stained gel (Figure 3-13-B) but is shown to be highly active in the in-gel activity assay (Figure 3-13-C). The recovered active form was then loaded on to a cation exchange column for further purification as a method to reduce the volume of the solution.

3.5.6.3 *Cation exchange of 7WC.*

The cation exchange chromatography showed a single peak after eluting with a 0 to 2 M NaCl gradient. The highest peak (at an absorbance of 280 nm) is at 700 mM NaCl which corresponded with the eluted monomeric 7WC and enzyme activity determined using the in-gel activity assay (Figure 3-13).

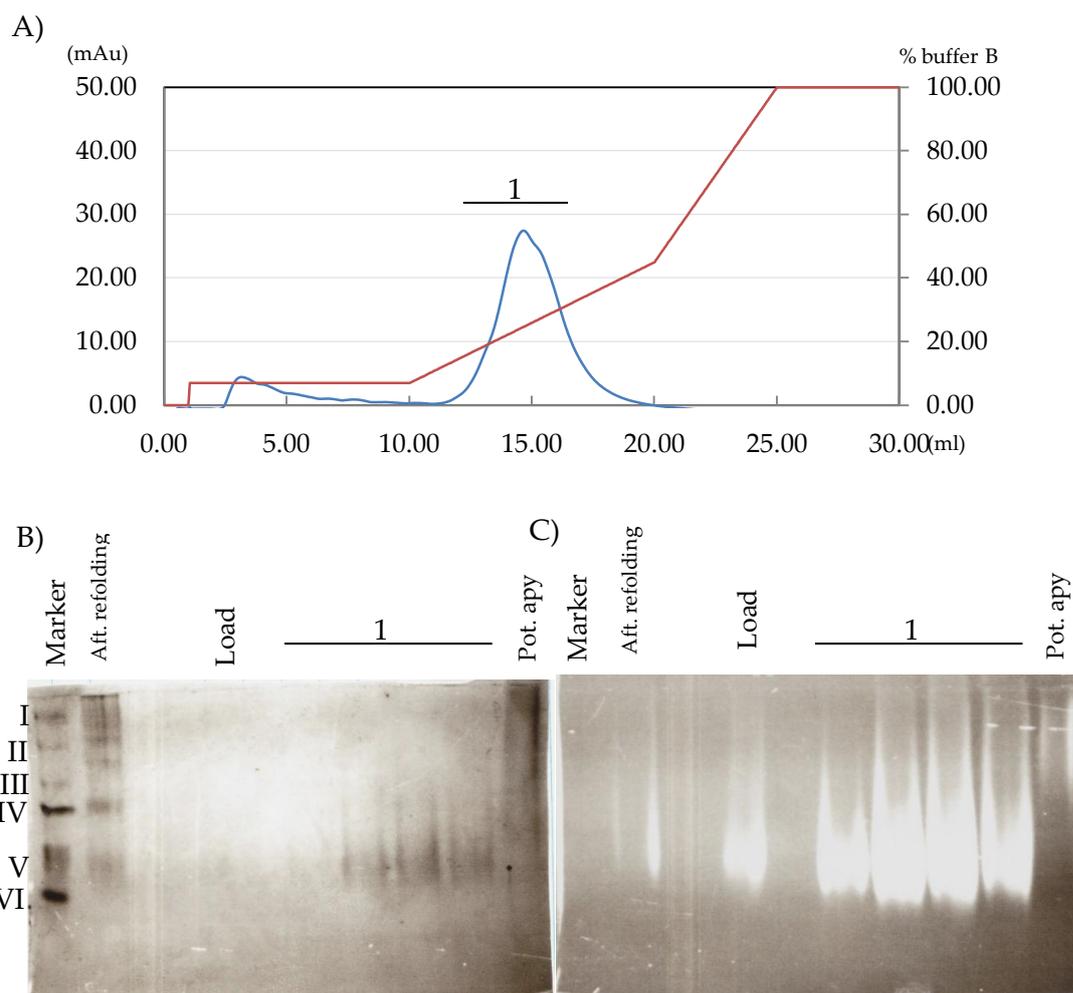


Figure 3-13 Native gel analysis of 7WC after cation exchange.
 A) FPLC trace. B) Silver stain of 10 % Native PAGE C) In-gel activity assay of 10 % Native PAGE. Marker labelled I-VI

3.5.6.4 Purification of 7WC by Size exclusion chromatography.

For enzyme kinetic analysis and crystallography a high purity of protein was essential. So size exclusion chromatography was performed after cation exchange chromatography. The run buffer initially contained 50 mM acetate, pH 5.0 which was later changed to 20 mM Tris, pH 8.0 and 150 mM NaCl because this was a successful buffer for crystallisation of human NTPDase1 (Zhong, et al., 2008). The change of buffer did not seem to have any effect on the stability of 7WC. 7WC eluted from the column at ~ 16 ml which when compared to a standard (section 2.5.2) suggested that 7WC exists as a monomer in solution.

The SDS-PAGE analysis confirms that only 7WC is present in pooled fractions and that observation of the native PAGE analysis confirmed that only the monomeric form, which migrated to ~ marker V, is present. Comparison with the in-gel activity assay (Figure 3-13-C) shows that this is the active form.

Typically, from a one litre culture 1 mg of 7WC could be isolated. Compared to the amount of 7WC isolated as inclusion bodies represented a 5 % yield. This protein preparation was used for all subsequent experiments.

A)

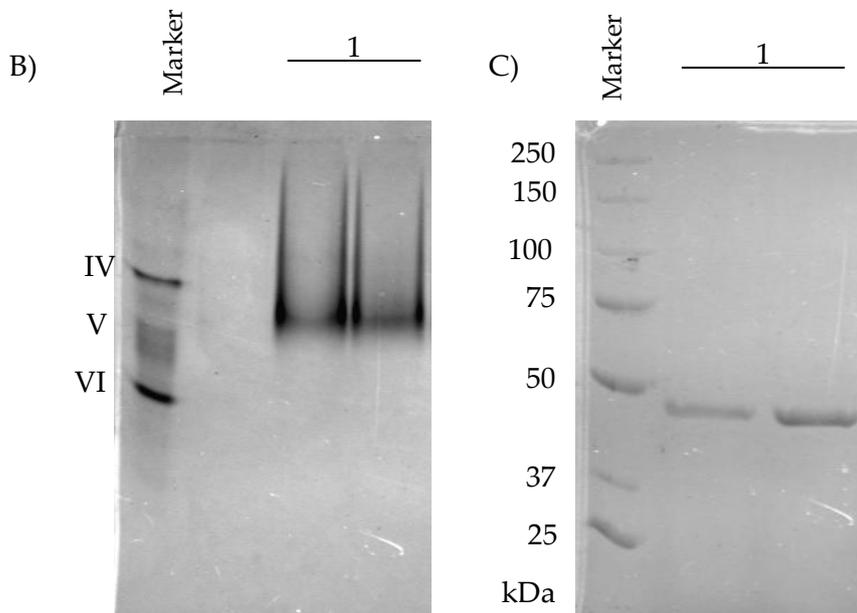
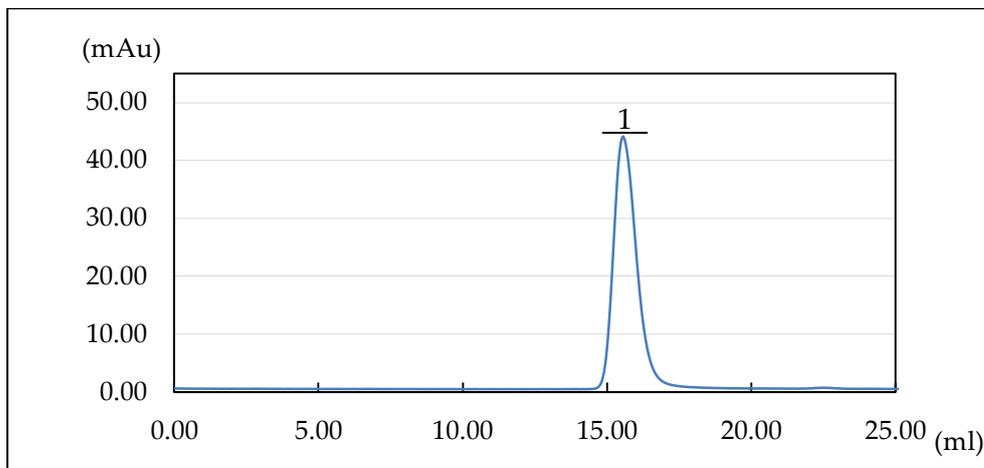


Figure 3-14 A) S200 chromatography of 7WC. B) 10 % native and C) 10 % SDS-PAGE. Both gels were stained with coomassie brilliant blue. Native marker labelled IV-VI

3.5.7 Analysis and purification of refolded DbLnP

Large scale refolding of DbLnP was concentrated using the stir cell to about 50 ml. Observation of the refolded DbLnP by native gel analysis showed that refolded DbLnP was largely held up in the stacking gel or at the interface of the stacking and separation gels with a protein running as a smear between the native marker I-VI Figure 3-15. A darker region was observed at approximately

the native marker V (Figure 3-15-arrow). DbLnP was purified further using size exclusion chromatography.

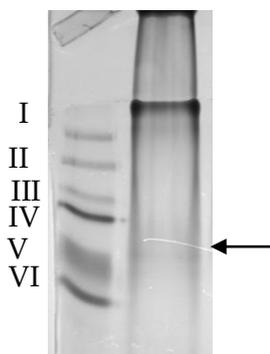


Figure 3-15 Native gel analysis of DbLnP after refolding. Arrow indicates a light band. Native marker labelled I-VI

3.5.7.1 Purification of active form of DbLnP by Size exclusion chromatography.

Protein preparation from the stir-cell was concentrated and loaded onto a Sephadex S200 16/60 or 10/300 and eluted with 20 mM Tris-HCl, pH 8.0 and 150 mM NaCl. The absorbance trace (280 nm) showed two pronounced peaks (Figure 3-17); peak one is ~ 5 ml immediately after the void volume indicating the eluted protein is soluble and of a high molecular weight. Peak two elutes at a volume consistent with the molecular weight of the monomeric DbLnP when compared to the protein standard curve for this column. This is in agreement with previous reports which showed DbLnP from root extract was a monomer by size exclusion chromatography (Quinn, et al., 1987).

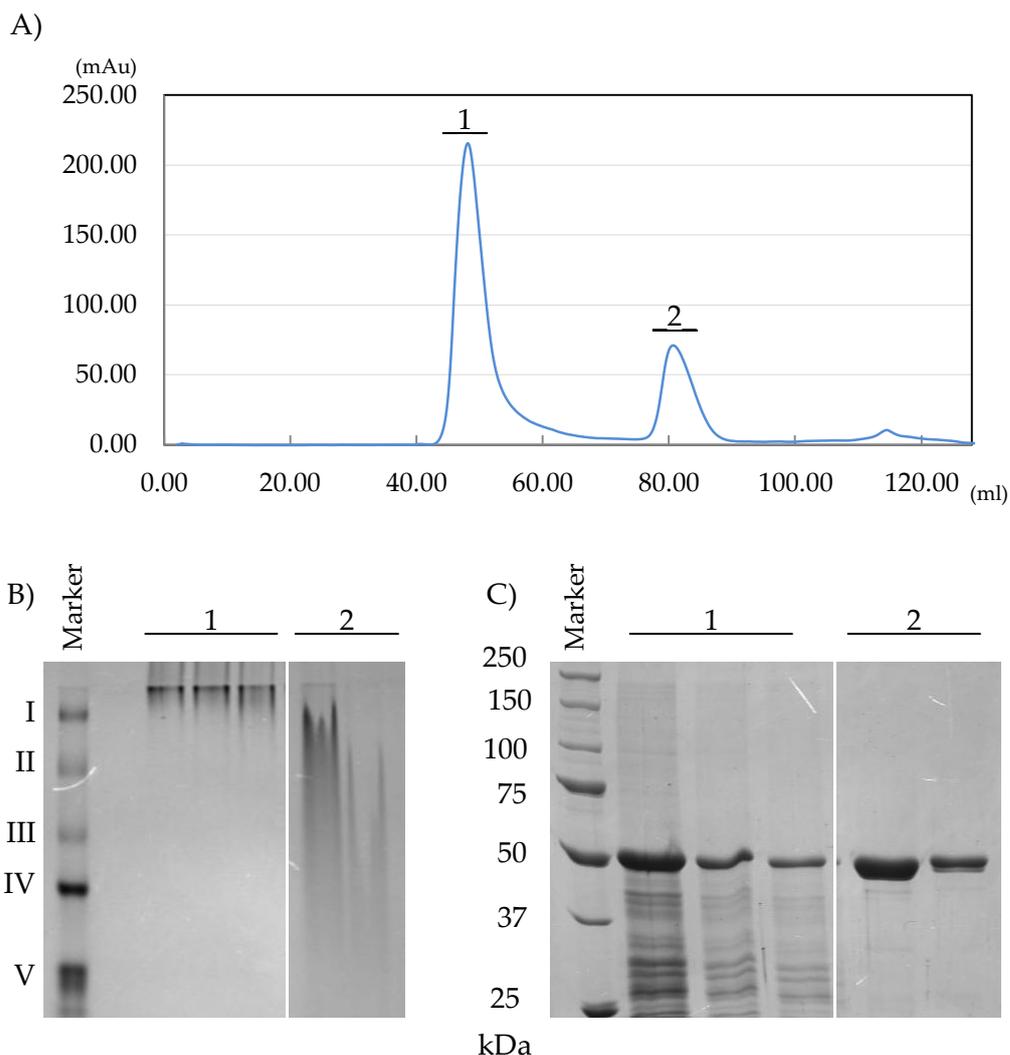


Figure 3-16 16/60 S200 Size exclusion chromatography of refolded DbLnP.
 A) The absorbance trace (280 nm). B) Native-PAGE analysis of the peak fractions. C) SDS-PAGE analysis of the peak fractions. Native marker labelled I-V

Comparison of the two UV trace peaks showed that peak 1 is held up as it enters the separating gel and also to a small degree migrates into the top of the stacking gel. Consistent with the UV trace, the protein is likely to be a large molecular weight soluble aggregate. Peak 2 showed that the monomer runs as a broad band between marker II and V. Such variation in the migration was commonly observed and maybe due to contamination of arginine or salt from the refolding buffer resulting in variation in the overall charge of the monomer. SDS-PAGE confirmed that both peaks contain DbLnP although peak 1 also contains a number of low molecular weight proteins (< 50 kDa) that are most likely to be degraded DbLnP encapsulated by the aggregated proteins. Low molecular

weight proteins were commonly observed, by SDS-PAGE analysis, after metal affinity chromatography purification of DbLnP.

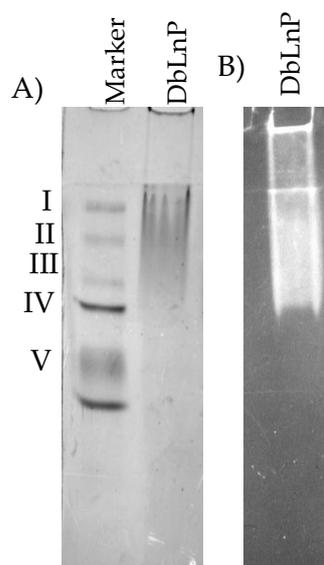


Figure 3-17 A) Native-PAGE and B) In-gel activity assay of the size exclusion purified DbLnP.

To confirm that the DbLnP eluted in peak 2 of the chromatogram was the active species, an in-gel activity assay was performed. This showed a comparable smear that observed in the native gel analysis. Protein activity was also observed in the stacking gel illustrating that not all the active DbLnP runs into the gel.

These data showed it was possible to separate the higher molecular weight species of DbLnP from the active monomeric form by size exclusion and this method was used for further purification of DbLnP. Compared to the 7WC purification method, this approach took less time and generally gave a higher yield of protein. Typically, from a one litre culture 5 mg of ultrapure DbLnP was isolated.

Longer refolding times increased the yield of the active form of DbLnP. Refolding DbLnP for 2 weeks (rather than the typical week) resulted in the ratio between first and the second peaks on the size exclusion trace changing from ~3:1 to ~1:1 (Figure 3-18). Despite this and due to time constraints, DbLnP was refolded for 1 week. The yields were ~5 % yield after a week and ~15 % after 2 weeks.

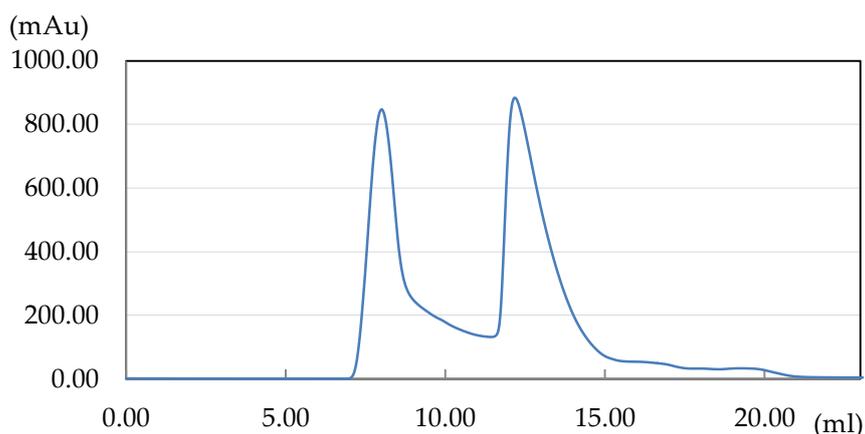


Figure 3-18 10/300 S200 Size exclusion chromatography of refolded DbLnP after two weeks.

Blue line represents the absorbance trace (280 nm).

3.6 Discussion and concluding remarks

The expression of non-host recombinant proteins in *E. coli* has become one of the most popular methods for protein production largely due to the fast growth rate, rapid accumulation of overexpressed protein and the capacity for high density and large scale fermentation. It is however, not uncommon that recombinant proteins incorrectly fold and form inactive, insoluble aggregates known as inclusion bodies. The tendency for mis-folding is increased with the use of strong promoters and if the recombinant protein contains disulfide bonds in the native state, these are impossible to form in the reducing condition of the *E. coli* cytoplasm. This chapter presents the expression of 7WC and DbLnP in *E. coli* as insoluble aggregates. To confront this issue, techniques to unfold or solubilise the NTPDases were utilised and correct refolding conditions were discovered. The use of high throughput methods for the refolding of NTPDases was essential to quickly and efficiently determine correct refolding conditions. Using this approach 7WC and DbLnP were successfully refolded from inclusion bodies.

Comparing the optimal refolding conditions for 7WC and DbLnP revealed that both enzymes optimally refold at neutral pH and require substantial NaCl but differ in the requirement for an oxido-shuffling mechanism and arginine.

Efficient refolding of 7WC was not as effective as DbLnP. As the optimal refolding condition based on enzyme activity also had high amounts of aggregation. This suggests that there are a number of side reactions leading to incorrect folding and aggregates. One possible reason for this is the fact that the soluble and the structured intermediates have a larger proportion of hydrophobic side chains exposed to the solution which leads to the intermediates associating with one another. One component that was effective against aggregation for both 7WC and DbLnP was the addition of arginine. Arginine in the refolding solution is thought to increase the free energy required for proteins to associate by restricting the exposure of the hydrophobic patches to the solvent (Baynes, et al., 2005). Another theory is that aggregation prone intermediates are somewhat protected through π -stacking interactions to aromatic residues, such as phenylalanine, tryptophan and tyrosine, before the molten globule transition state where the hydrophobic residues are buried in the hydrophobic core (Arakawa, et al., 2007; Tsumoto, et al., 2004). The addition of arginine to the refolding solution of DbLnP reduced the amount of aggregation and presumably allowed more protein to correctly fold to the native state. Despite arginine reducing aggregation with 7WC it did not increase the specific activity.

For both 7WC and DbLnP, 3 disulfide bonds need to be correctly formed. It is therefore common for an oxido-shuffling system to be present in refolding systems for proteins containing disulfide bonds. The aggregation observed with 7WC could be the formation of inter-molecular disulfide bonds. This is consistent with the results that showed improved refolding efficiency with the addition of the glutathione oxido-shuffling system.

The necessity to refold prokaryotically expressed NTPDases is a common issue. It was reported by a number of investigations including the rat NTPDase1-3 expressed in *E. coli* (Zebisch, et al., 2007) and human NTPDase5-6 in *E. coli* (Murphy-Piedmonte, et al., 2005) and also unpublished results on various NTPDases including an NTPDase from ryegrass (termed 6RG), white clover (termed 4WC) and the human NTPDase1 (CD39) (Chen, C., Cumming, M., Scott, R., Arcus, V., and Roberts, N. *unpublished results*). The components NaCl,

glutathione and arginine are a common feature in all of the refolding conditions for the NTPDases mentioned above illustrating the importance of these components for refolding NTPDases.

After refolding, native-analysis revealed that both 7WC and DbLnP contained soluble higher molecular weight forms of each. 7WC has different multimeric forms that were inactive and could be removed from solution by acid treatment. These multimers are likely to be incorrectly folded intermediates with various degrees of intra-molecular bonding. The treatment to an acidic environment destabilised the multimeric forms that precipitated from solution. The monomeric form was the active form and could be purified using cation exchange and size exclusion chromatography.

In contrast to 7WC, DbLnP did not form various multimeric forms after refolding rather existing as either a monomeric or a high molecular weight form. These could be isolated using size exclusion chromatography. Refolding for longer periods of time resulted in an increase in the monomeric form. This slow conversion from the high molecular weight form to the monomeric form illustrates equilibrium between intermediate forms and that once correctly folded the protein remains in this state and draws the equilibrium towards the active monomeric state.

4 Crystallisation and crystallography 7WC

4.1 Introduction

7WC is a root localised NTPDase that is believed to be involved in indeterminate nodule formation during the rhizobium symbiosis (Chen, 2008). 7WC hydrolyses nucleotide tri- and di-phosphates. There is no information about the structural or catalytic mechanism of 7WC. To investigate this, crystallisation trials were conducted and attempts to co-crystallise with substrates were attempted. These trials resulted in three structures including an apo- and two substrate-bound forms.

4.2 Initial crystallisation experiments.

Initial screening attempts for crystallisation used the sitting drop vapour diffusion method in a 96 well format at 18 °C by mixing 100 nl of protein solution with 100 nl mother liquor and equilibrating with 75 µl mother liquor. The initial screens were set up using the Cartesian honey bee crystallisation system at the Maurice Wilkins Centre for Molecular Biodiscovery crystallisation facility at the University of Auckland.

Two trials for 7WC were performed; 1. 7WC was purified in 50 mM acetate, pH 5.0 and concentrated to ~4.5 mg·ml⁻¹ or 2. 7WC purified in 20 mM Tris, pH 8.0, 150 mM NaCl and 5 mM AMPPNP, 20 mM CaCl₂ and concentrated to ~7 mg·ml⁻¹. The two domains of NTPDases are hypothesised to have flexibility allowing it to have an open and a substrate-bound, closed conformation as observed with several members of the actin super family (Kristensen, et al., 2008; Rangarajan, et al., 2006; Schuler, 2001). A non-hydrolysable substrate, AMPPNP and the co-factor, Ca²⁺ were added to the crystallisation trial in an attempt to 'lock' the enzyme in a conformation that was more rigid and more likely to crystallise.

Protein samples were centrifuged at 13,000 x-g for 5 min at 4 °C to remove any precipitation and particulates before being transferred to the crystallisation plates. Observation of both trials after two weeks of incubation showed eight promising conditions. However, these crystals showed no x-ray diffraction.

4.2.1 Further crystallisation trials.

Purified 7WC protein was dialysed against 50 mM Tris, pH 8.0, 150 mM NaCl before being concentrated to 7 mg·ml⁻¹ (section 2.7.4.2). Crystallisation experiments were conducted using the hanging-drop diffusion method using 24 well plates (Hampton) with a 500 µl reservoir and a 2 µl reservoir + 2 µl protein drops on siliconised coverslips (Hampton) and incubated at 18 °C. Initial screens of an appropriate reservoir solution were based on conditions determined for human soluble ecto-nucleoside triphosphate diphosphohydrolase 1 (20% PEG 3350 and 0.2 M ammonium dihydrogen phosphate) (Zhong, et al., 2008) and seeding from an unrelated protein crystal was required before thin rectangular plates (0.2 mm x 0.8 mm) formed in 1-2 weeks (Figure 4-1). The 7WC protein crystals were subsequently used to promote crystal growth in further trials. Fine screen trials were performed to produce large thicker plates. This was achieved by trialling different micro-seeding techniques either using a seed stock or using a cat's whisker brushed along a crystal (section 2.8.4) and by altering the concentration of the components in the precipitant solution. The best crystals were grown using reservoir solution that contained 0.2 M ammonium dihydrogen orthophosphate and 15 % PEG 3350 and carefully microseeded with a cat's whisker brushed over a 7WC crystal.

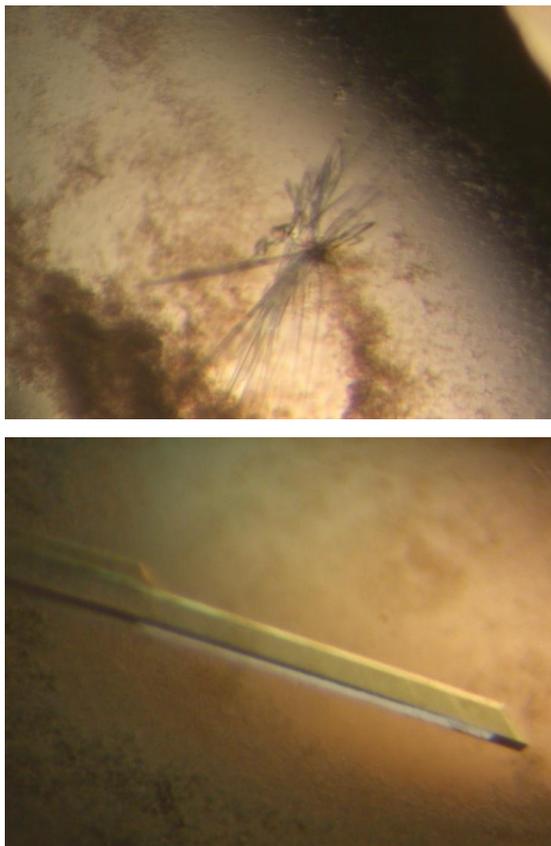


Figure 4-1 Crystallisation of 7WC.

A) Initial and B) final crystal forms of 7WC grown in 0.2 M ammonium dihydrogen orthophosphate and 15 % PEG 3350.

4.3 Data collection

7WC crystals were soaked in the reservoir solution with increasing amounts of PEG₃₃₅₀ to a final concentration of 40 % before being flash-cooled in liquid nitrogen. A data set was collected using a Rigaku Ru 300R rotating anode Cu-K α radiation source and using a MAR345 image plate detector. The images were collected over 360° at 0.5° increments for 3 min each and a detector distance of 140 mm (Figure 4-2). Images 90° apart were used to index the space group (MOSFLM) which showed a primitive monoclinic unit cell ($a = 54.0$, $b = 53.6$, $c = 72.8$, $\alpha = 90^\circ$, $\beta = 95^\circ$, $\gamma = 90^\circ$) with a space group P 1 2₁1 (Leslie, 1992). The full data set was integrated using MOSFLM using 1.89 Å as the resolution cut-off. The mosaicity was fixed at 1.0 because it rose to above 1.0 during integration to ensure that full reflections are recorded.

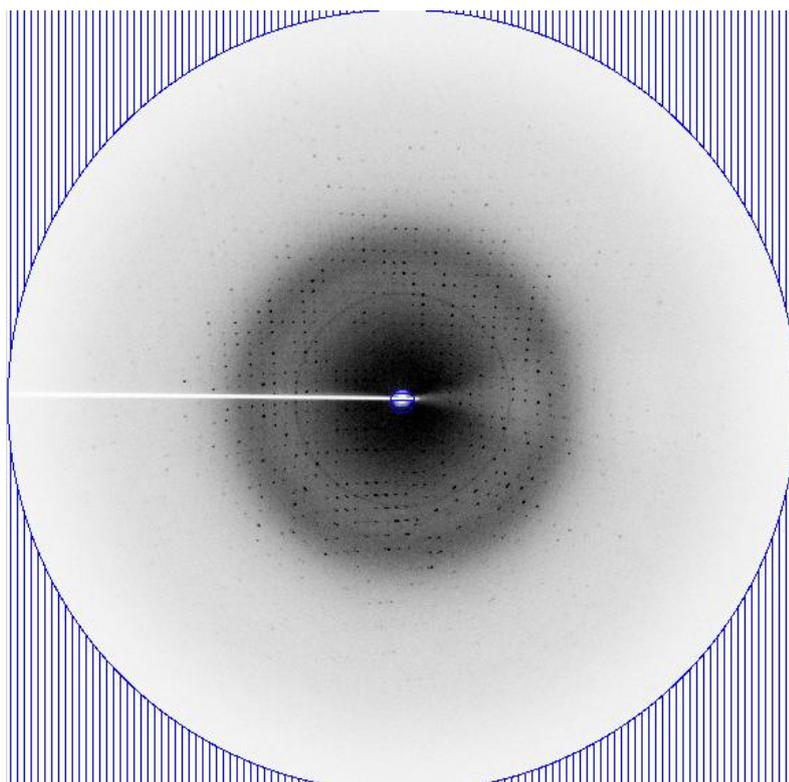


Figure 4-2 Diffraction pattern collected on the home source x-ray generator and MAR detector.

Detector is at 140 mm and the edge of the plate is 1.79 Å.

4.4 Data processing

All data were analysed using programs either within the CCP4 package (Collaborative Computational Project, Number 4, 1994) or in the PHENIX suite (Python-based Hierarchical Environment for Integrated Xtallography) (Adams, et al., 2010; Bailey, 1994).

The Matthew's coefficient analysis predicted that the unit cell contained one molecule in the asymmetric unit with 45.1 % solvent content and a Matthew's coefficient of 2.24 Å³/Da (Matthews, 1968). The reflections were scaled and merged using SCALA (section 2.11.3) using all the reflections. The R_{merge} in the outer shell of less than 0.5 was used to determine the resolution cut-off during SCALA. The indexing and merging statistics are shown in Table 4-1.

Data collection statistics for apo-7WC	
Wavelength (Å)	1.54
Space Group	P1 2 ₁ 1
Cell Dimensions	
<i>a b c</i> (Å)	53.9/ 53.6/ 72.8
<i>α β γ</i> (°)	90.0/ 94.9/ 90.0
Matthew's coefficient (Å ³ /Da)	2.24
Wilson B-factor (Å ²)	21.2
Molecules/asymmetric unit	1
Mosacity	1.0
Resolution range (Å)	72.6-1.89
No of measured reflections	366450 (37938)
Unique reflections	33604 (3718)
<i>R</i> _{merge} [†]	0.076 (0.47)
Mean <i>I</i> / <i>σI</i> (%)	25.8 (3.6)
Completeness	99.0 (96.5)
Multiplicity	10.9 (10.2)
[†] $R_{\text{merge}} = \frac{\sum_{hk\ell} \sum_j I_{hk\ell,j} - \langle I_{hk\ell} \rangle }{\sum_{hk\ell} \sum_j I_{hk\ell,j}}$	

Table 4-1 Data collection statistics for home source data for apo-7WC. Data in the parenthesis are for the outer-most shell.

4.5 Molecular replacement

The recently solved NTPDase 2 (PDB code 3CJI), an NTPDase from *Rattus norvegicus* with a sequence identity of 26.2% was used as a molecular replacement model because at the time it was the only NTPDase structure available (Zebisch, et al., 2008). To maximise potential success of molecular replacement CHAINSAW (ccp4i) was used to make a modified NTPDase2 model based on conserved residues between NTPDase2 and 7WC (Stein, 2008). The members of the actin super family typically have domains that have been shown to be flexible upon the binding of substrate and this has been suggested to occur for NTPDases (Kristensen, et al., 2008; Rangarajan, et al., 2006; Schuler, 2001). To accommodate any differences in the angle of the domains, N-terminal (Met1-Ile126) and C-terminal (Asp127-Glu420) models were made using CHAINSAW (CCP4). The domain cut-off was based on the NTPDase2.

Solution of the C-terminal domain model was achieved using PHASER (method 2.11.5) in the space group $P 1 2_1 1$ and the CHAINSAW model (McCoy, et al., 2007). Three solutions were determined and the best solution was based on the model that had the lowest number of clashes. Initially the phases of the N-terminal region were determined with PHASER using the rotation and translation function of the C-terminal domain solution. The resulting N-terminal domain model was aligned with the C-terminal domain using the most appropriate symmetry coordinates however this model did not agree with the electron density map. The correct orientation of the N-terminus was achieved by firstly performing PHENIX autobuild (Terwilliger, et al., 2008) with the C-terminal domain model and the full length sequence file and this resulted in the placement of the C-terminus along with a number of N-terminal residues including the portion of a β -strand (Val10-Ala13, Ser18-His24) and a portion of two α -helices (Leu62-Glu67 and Asp97-Ile102). Subsequently, the N-terminal domain model determined previously by PHASER was superimposed using least squares quadratic fitting (LSQ) with the N-terminal residues in the autobuild model in COOT (Emsley, et al., 2004; Krissinel, et al., 2004). The position of the N-terminus was manually added to the C-terminus model and was used in Arp/wARP to complete the model (Langer, et al., 2008). Using this method a starting model was built containing 231/426 residues in 17 chains which had an R value of 0.36 and an R_{free} value of 0.42.

The Arp/wARP model was improved and completed by manually building residues in COOT directed by the $2|F_o| - |F_c|$ and $|F_o| - |F_c|$ maps (contoured at 1.0σ and at 3.0σ , respectively) and refined using REFMAC 5 (Murshudov, et al., 1997). Finally, the model was improved by defining the two domains (Thr5-Ile125 and Ile126-Lys412) as rigid bodies and performing TLS (Translation Libration Screw-motion) refinement in REMAC 5 (Figure 4-3). The final refinement resulted in R and R_{free} values of 0.18 and 0.23 respectively. The final model contained 407 residues out of the 426 residues (this included the mature peptide and the His-tag), two free phosphates and 353 waters. The missing residues that could not be resolved in COOT (with the $|F_o| - |F_c|$ map contoured at

3.0 σ) were the first 3 residues at the N-terminus, 14 residues of the C-terminus (including the His-tag) and from one large loop between Thr174-Asn177 and Val 181-Glu185. Also, a number of side-chains were unable to be placed. The missing loops and side chains were modelled in with occupancies of 0.

The N- and C-terminal domains have small average B-factors of 15.2 \AA^2 and 12.1 \AA^2 , respectively with the overall average B-factor of 13.1 \AA^2 .

Protein geometry was validated using MolProbity which showed that 99.5 % of all residues had acceptable Ramachandran angles with 2 outliers, Ala89 and Gly96 (Lovell, et al., 2003)(Figure 4-4). The outliers had well defined density but are part of a large loop which might explain the unusual Ramachandran angles.

A full summary of the refinement statistics of the completed structure are displayed in Table 4-2.

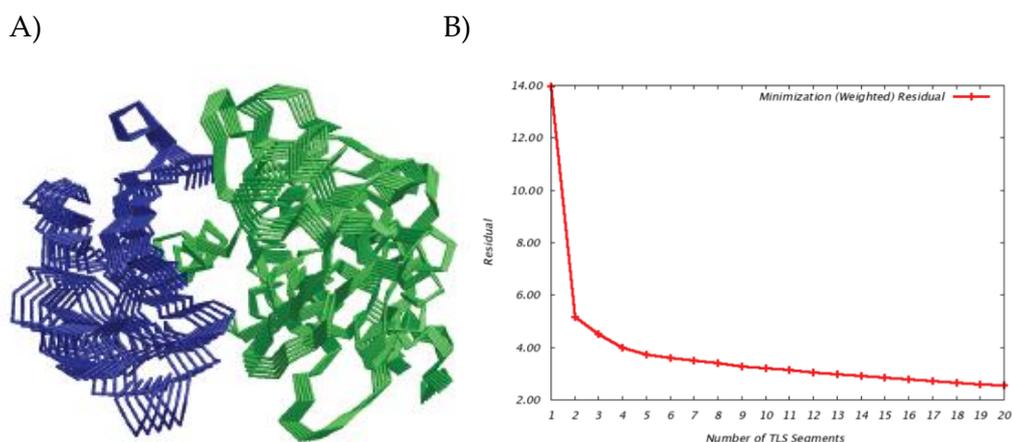


Figure 4-3 The result from the TLS analysis of apo-7WC.

A) A ribbon diagram showing the three domains used during refinement. B) Residual vs. Number of TLS segments.

Refinement and model statistics for Apo-7WC	
Resolution range (Å)	72.5-1.89
R-factor	0.18
R _{free} [†]	0.23
Average B-factors (Å ²)	
Overall (No. of atoms)	15.3 (3461)
Overall protein (No. of atoms)	13.1 (3098)
N-terminus (No. of atoms)	16.2 (930)
C-terminus (No. of atoms)	13.2 (2168)
Solvent (No. of atoms)	25.7 (353)
Ions (No. of atoms)	34.9 (10)
R.M.S deviation	
Bond lengths (Å)	0.02
Bond angles (°)	1.87
Ramachandren plot	
Percentage in favoured regions	96.1
Percentage in allowed regions	99.5
Percentage in disallowed regions	0.5
$R_{free}^{\dagger} = \frac{\sum_{testset} \left F_{obs} - F_{calc} \right }{\sum_{testset} F_{obs} }$	

Table 4-2 Refinement and validation statistics of apo-7WC.

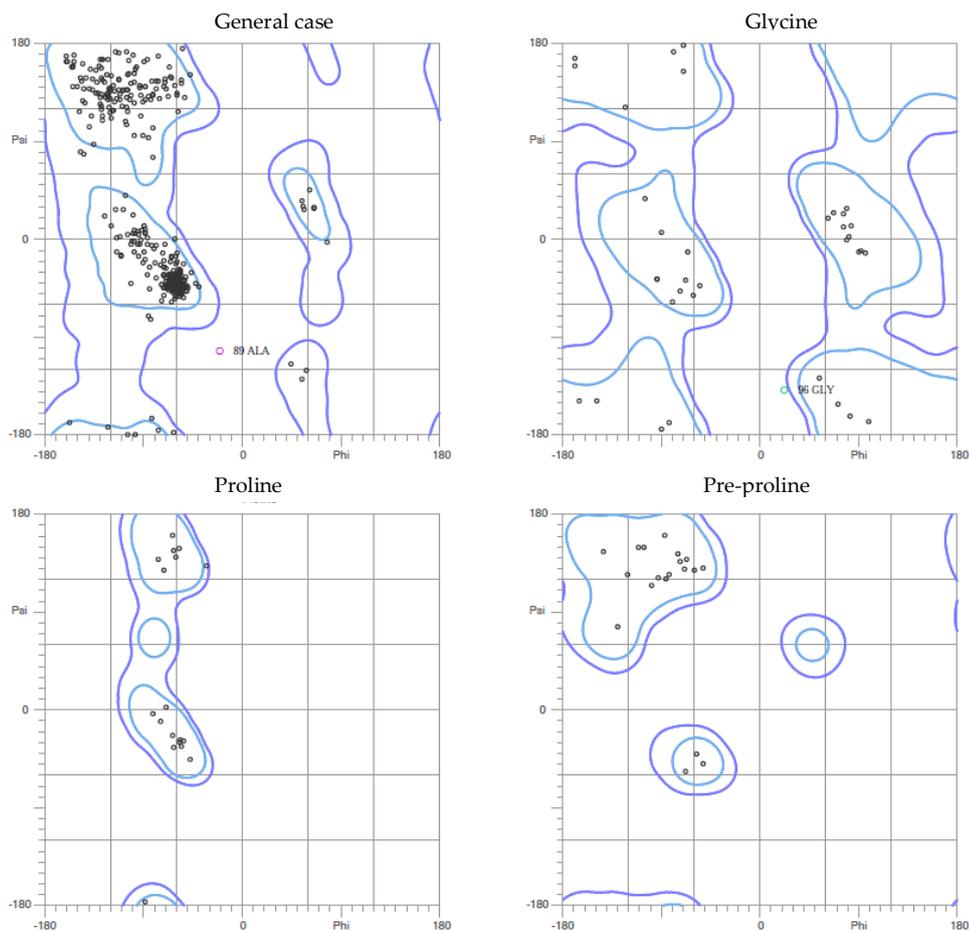


Figure 4-4 Ramachandran analysis of apo 7WC. Light blue regions indicate the preferred and the dark blue regions represent the acceptable ψ and ϕ angles.

4.6 Overall structure of the 7WC enzyme

The crystal structure of 7WC was determined at 1.89 Å resolution (section 4.5). The cartoon representation of the overall structure shows that 7WC consists of two domains either side of a putative hinge region (Smith, et al., 1999b; Vivian, et al., 2010; Zebisch, et al., 2008) which is orientated so that the ACRs make up the surface of the active site (discussed further in section 4.6). The N-terminal domain (Ile 4 - Ile 126) consists of a mixed, five stranded β -sheet where the second strand is anti-parallel to the rest (Figure 4-6). This is typical of the RNase-H fold represented by members of the actin-like ATPase superfamily (Smith, et al., 1999b). The RNase-H domain characteristically contains a β -sheet of 5 strands in the order 32145, with the second anti-parallel to the rest flanked by a single

α -helix one side and >2 α -helices on the other. The β -sheet of 7WC is surrounded by four α -helices on one side and the C-terminal α -helix that makes up ACR5 (Figure 4-9).

The C-terminal domain (Asp127-Phe411) is a larger domain that shares similar features to the N-terminal domain. The larger C-domain shows a modified RNase-H domain with inserted β -strands. The typical RNase-H domain would include (from N- to C-terminus) α_6 , α_7 , α_9 , α_{12} and α_{14} . However the C-terminal RNase-H domain contains a β -strand insertion of α_8 . This with an additional α_{15} increases the β -sheet by 2 β -strands. In contrast to the N-terminal domain, two β -hairpin turns are also present (α_{10} - α_{11} and α_{15} - α_{16}) (Figure 4-6). The mixed β -sheet is positioned so that most of the sheet curls around the conserved α -helix of ACR3 (Figure 4-9). Pairwise structural comparison between the N- and C-terminus demonstrated a rmsd of 4.2 Å using 282 residues. This similarity is thought to have arisen from gene duplication of the single domain that required dimerisation for function (Bork, et al., 1992).

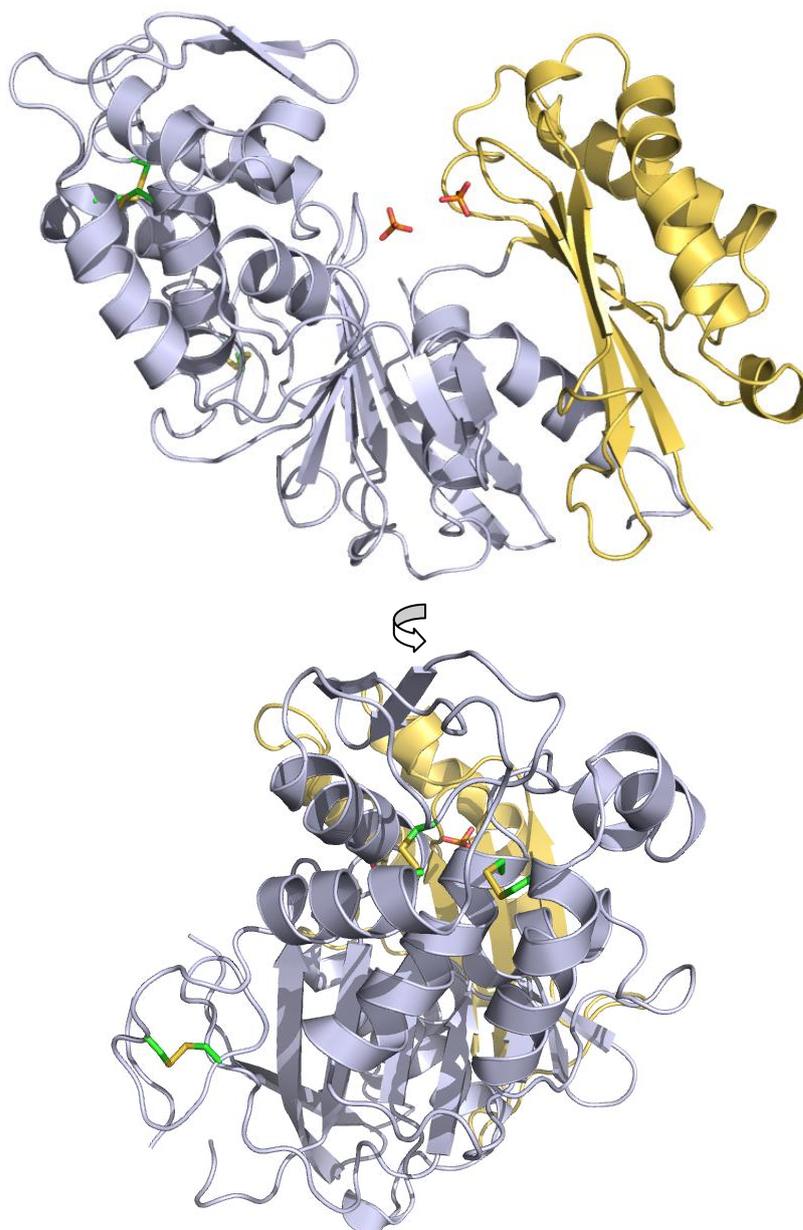


Figure 4-5 Overall structure of 7WC.

N- and C-terminal domains are coloured gold and grey, respectively. Disulfide bonds are shown in green and yellow. The two phosphates are shown in orange. Image was produced using Pymol.

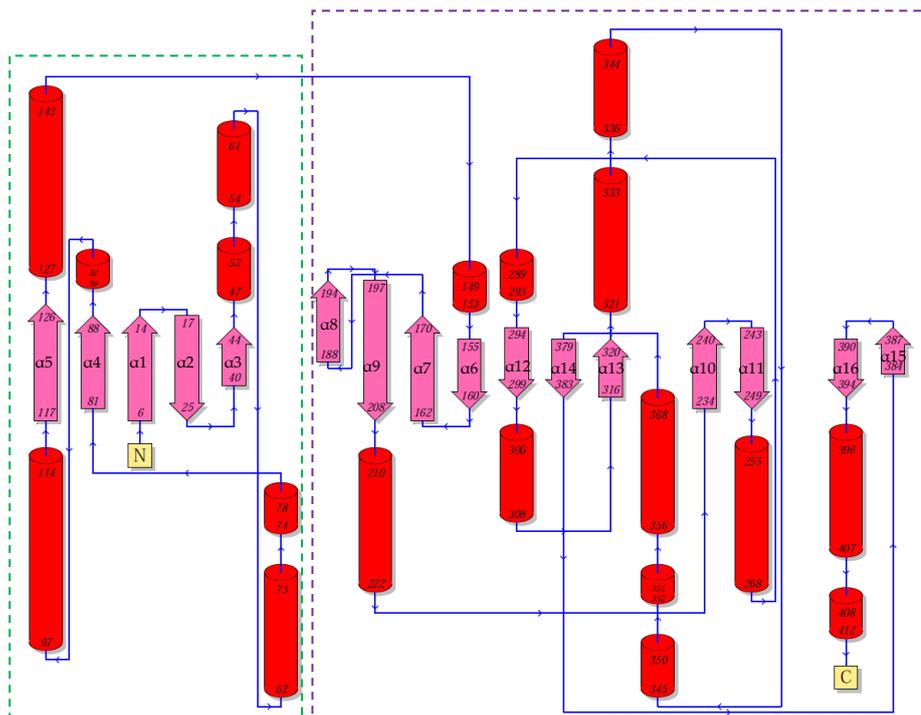


Figure 4-6 Topology model of 7WC.

Figure was produced using PDBsum. Green dashed line represents the N-terminal domain. Purple dashed line represents the C-terminal domain

Three disulfide bonds are formed within the C-terminal domain between cysteines conserved within all NTPDases. The disulfide bond between Cys229-Cys259 and Cys333-Cys357 results in a tight assembly of the C-terminal α -helices. Whereas the disulfide Cys278-Cys273 maintains a large loop from Asp270-Gly281 in the C-terminal domain (Figure 4-5).

The 7WC structure was superimposed with *R. novogicus* NTPDase2 (3CJA) with substrate analogue AMPPNP bound which revealed that the free phosphates (termed phosphate 1 and 2) present in the 7WC structure corresponded with that of the β - and γ -phosphate of AMPPNP bound in NTPDase2 (Figure 4-7). The residues in NTPDase2 involved with the binding the α - and β -phosphate of AMPPNP are Ser48, Ser49 and His50 which correspond with those that associate with phosphate 1, Ser15, Thr16 and Arg19. In 7WC, the analogous His50 (in NTPDase2) is not present but the side chain of Arg19 rotates to occupy the same space. The γ -phosphate of AMPPNP is associated with the C α chain of Gly204 and Ala205 in NTPDase2 and the phosphate 2 interacts with Gly161 and Gly162.

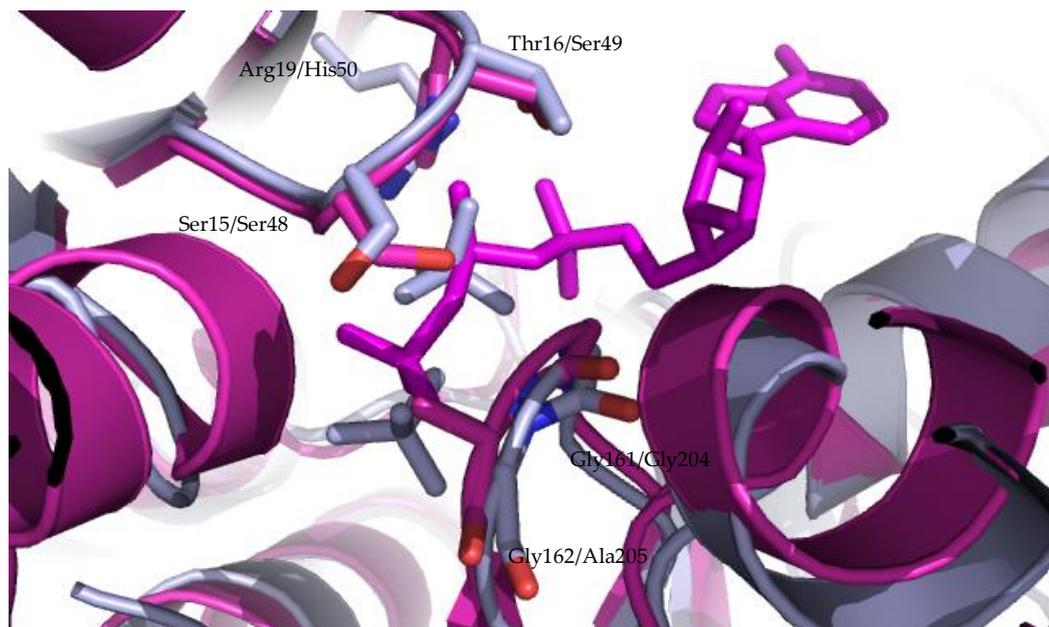


Figure 4-7 Superposition of apo-7WC (light blue) with rat NTPDase2 with AMPPNP bound (magenta). Each domain was overlaid individually to compensate for differences in the inter-domain angle. Residue labels are for 7WC and rat NTPDase2, respectively.

The position of the bound phosphates indicates where the phosphate tail of ATP could bind and the similarities between NTPDase2 and 7WC suggest that this is achieved in the same way. Using these results, the regions 'nucleotide phosphate binding site' and 'phosphate binding sites 1 and 2' within the active site can be assigned labels (Figure 4-8). These will be referred to throughout the thesis.

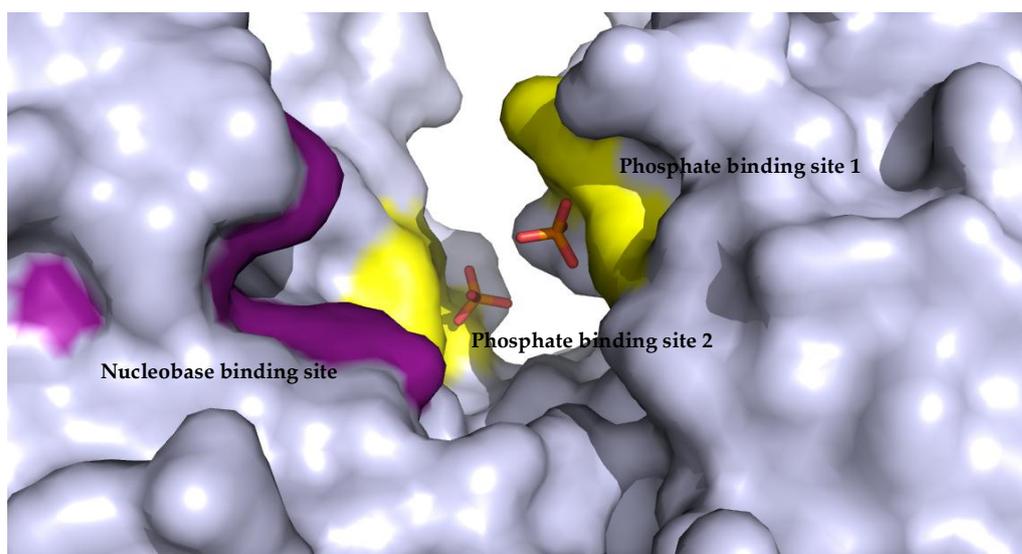


Figure 4-8 Surface model of apo-7WC demonstrating the three regions of 7WC that interact with substrate.

Yellow regions represent phosphate binding sites whereas purple represents the nucleobase binding site. Residues are shown in Table 4-3.

	Phosphate binding site 1	Phosphate binding site 2	Nucleobase binding site
7WC	Ser15, Thr16, Arg19	Gly161, Gly162, Ser163	Tyr303, Asp307, Phe350
DbLnP	Ser15, Ser16, Arg19	Gly161, Gly162 and Ser163	Tyr303, Arg304, Glu307, Glu346
LpNTPDase1	Ser52, Thr53, Arg56	<i>G189, Ala190, Ser191</i>	Gly299, Asn302, Tyr346, Tyr350
Rat NTPDase2	Ser48, Ser49, His50	Gly204, Ala205, Ser206	Arg394, Tyr398, Tyr350

Table 4-3 Residues involved in the substrate binding regions within the active site cleft. Residues are shown for known structures: 7WC, DbLnP, *L. pneumonia* LpNTPDase1 and *R. novogicus* NTPDase2. Italicised residues are based on sequence alignment rather than structural analysis.

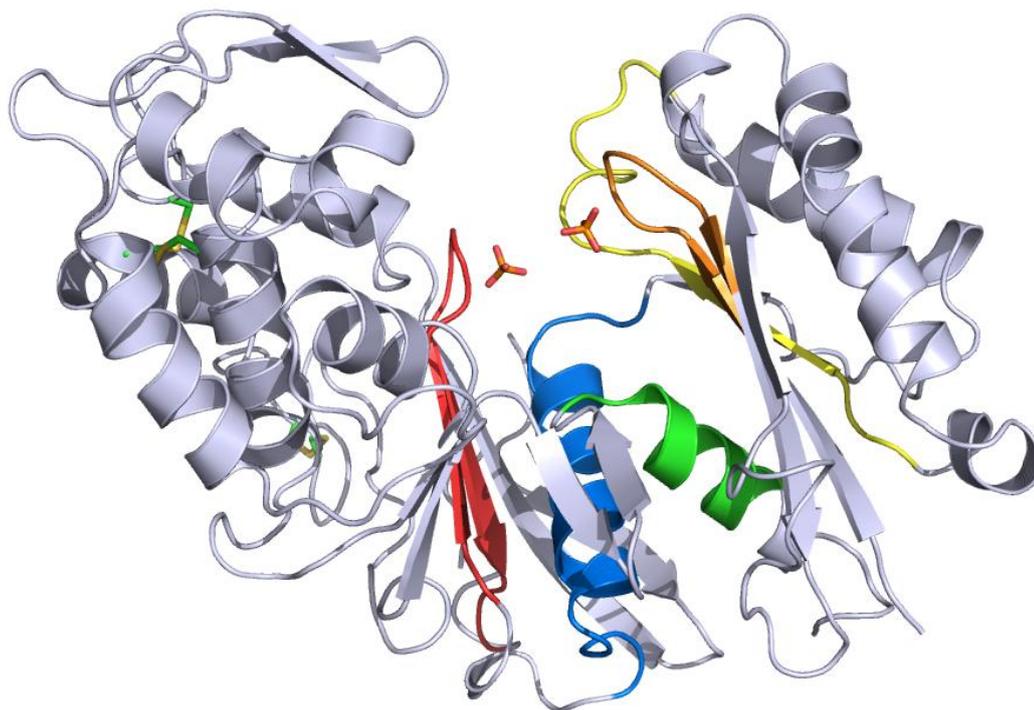


Figure 4-9 Cartoon representation of 7WC highlighting the ACRs. ACR regions are coloured Orange=ACR1, Yellow=ACR2, Blue= ACR3, Red=ACR4 and Green=ACR5. Disulfide bonds are represented as green and yellow sticks.

Highlighting the apyrase conserved regions (ACRs) (Handa, et al., 1996) on the 7WC structure illustrates that these regions are all located in the cleft between the two domains (Figure 4-9).

Using previous studies and data from this thesis it is possible to assign individual functions for each of the domains; ACR1 (Asp12-Arg19) and ACR4 (Thr153-171) are loops that function as phosphate binding sites that bind the phosphate tail of the substrate (discussed further in section 4.11 and 4.17). Residues within the ACR1 and ACR4 were also shown to coordinate metal binding in the *R. novogicus* NTPDase2 model (Zebisch, et al., 2008). The ACR2 (Thr81-Gly96) domain is a loop where residues within this region stabilize the catalytic residues as well as coordinate the metal associated waters. ACR3 (Ile125-Ser171) and ACR5 (Trp396-Ala401) are α -helices that are amid the two domains; the former contain residues that are involved in hydrolysis (discussed further in section 4.18.1) and also function as a hinge domain essential for trans-domain movement (discussed further in section 4.12). The ACR5 is a small

region that contains residues that coordinate waters that associate with the metal ion.

4.7 Comparisons with other NTPDases

A DALI search was performed using the entire 7WC model which identified the closest structural homologues (Holm, et al., 2000). Values determined by DALI neglect differences in the inter-domain angle which led towards the analysis of the structural similarities using the N- and C-terminal domains individually. DALI analysis of the two individual domains revealed closer structural homology.

DALI analysis identified NTPDase2 from *Rattus norvegicus* (3CJA) to have the closest structural homology with 7WC (Z-score of 35.7, rmsd of 3.4 over 419 residues). Analysis of the N-terminal domain is clearly closest to the *R. norvegicus* NTPDase2 with a rmsd of 1.8 Å, 0.8 Å lower than the next structural homologue, the LpNTPDase1 (3AAQ) from *Legionella pneumophila* (Z-score of 11.9 with a rmsd of 2.4 Å over N-terminal domain)(Vivian, et al., 2010; Zebisch, et al., 2008). The C-terminus has a lower homology to published PDBs with the closest being the *R. norvegicus* NTPDase2 with a rmsd of 2.6 Å. The higher homology of the N-terminal domain is not surprising since four of the five ACRs are within this domain.

The next closest structural homologue determined by DALI was the *L. pneumophila* NTPDase (3AAR) with an rmsd of 3.0 Å over 349 residues. Other structures that were revealed to be structurally similar are the NTPase from *Neospora caninum* and the exopolyphosphatase from *Aquifex aeolicus*.

Whole 7WC	PDB code	Z-score	rmsd	Residue #	Type	Species	Reference
1	3CJA	35.7	3.4	360	NTPDase	<i>R. novogicus</i>	(Zebisch, et al., 2008)
2	3AAQ	31.4	3.2	332	NTPDase	<i>L. pneumophila</i>	(Vivian, et al., 2010)
3	3AGR	31	3.6	364	NTPase	<i>N. caninum</i>	Unpublished
4	1T6C	21.3	3.5	285	PPX/GPPA	<i>A. aeolicus</i>	(Kristensen, et al., 2004)

N-terminal domain	PDB code	Z-score	rmsd	Residue #	Type	Species	Reference
1	3CJA	15.5	1.8	116	NTPDase	<i>R. novogicus</i>	(Zebisch, et al., 2008)
2	3AAQ	11.9	2.4	108	NTPDase	<i>L. pneumophila</i>	(Vivian, et al., 2010)
3	3MDQ	10	2.2	95	PPX/GPPA	<i>C. hutchinsonii</i>	Unpublished
4	2J4R	10	2.3	98	PPX/GPPA	<i>A. aeolicus</i>	(Kristensen, et al., 2008)

C-terminal domain	PDB code	Z-score	rmsd	Residue #	Type	Species	Reference
1	3CJA	23.6	2.6	249	NTPDase	<i>R. novogicus</i>	(Zebisch, et al., 2008)
2	3AAQ	21.3	2.7	231	NTPDase	<i>L. pneumophila</i>	(Vivian, et al., 2010)
3	3AGR	17.9	3.4	247	NTPase	<i>N. caninum</i>	Unpublished
4	1U6Z	17.8	3.6	185	PPX/GPPA	<i>E.coli</i>	(Rangarajan, et al., 2006)

Table 4-4 DaliLite structural alignment.

Displayed are the most structurally similar homologues for 7WC and the N- and C-terminal domains. Z-score is a DALI output that evaluates similarity where values above 3 are considered to have significant similarities.

While the overall structure between the 7WC and the rat NTPDase2 is very similar the latter contains a number of extensions that are absent in 7WC. This includes a hydrophobic loop; highlighted black in Figure 4-10 that contained a number of hydrophobic residues (Tyr183, Trp185, Val186, Trp189 and Ile190). This region is conserved in membrane bound mammalian NTPDases (NTPDase1-3 and 8) and is thought to act as a cell membrane anchor (Figure 3-1)(Knowles, 2011). The absence of this domain is consistent with 7WC being a

soluble enzyme (Etzler, et al., 1999; Quinn, et al., 1987). The other major difference is an extension of the β -strands highlighted in green in Figure 4-10, which forms a 13 residue extension forming an α -helix. In 7WC this is replaced by a shorter β -hairpin loop that resides on the surface of the C-terminal region. The NTPDase2 extension contains a disulfide bond (C241-C283), not seen in 7WC which is essential for activity in the related NTPDase3 where a point mutation of these cysteines to serines resulted in a 10 % loss in NTPase activity (Ivanenkov, et al., 2005).



Figure 4-10 Overlay of apo 7WC and NTPDase2.

The N- and C-terminal domains were overlaid independently to neglect differences in the domain angle. NTPDase2 is shown in pink, 7WC is shown in yellow. Highlighted regions black and green show major differences between 7WC and NTPDase2 (see text above).

Using the *R. novogicus* NTPDase2 as a reference the inter-domain angle was determined by aligning each domain with LSQKAB in the CCP4 suite and estimating the difference in angle between the aligned domains in COOT. This method gave a 2-dimensional difference in domain angle of approximately 5.5°. This difference in domain angle would explain why molecular replacement was not possible with the entire NTPDase2 model (section 4.5) and how the DALI searches with each domain have a higher homology compared with the entire model (section 4.7).

4.8 Substrate soaking.

To elucidate the binding mode and catalytic mechanism of 7WC for ATP the non-hydrolysable substrate adenosine 5' [β , γ -imido] triphosphate (AMPPNP) was used. AMPPNP has an imido group between the β - and γ -phosphate which is the site of hydrolysis. This substrate was chosen because compared with other non-hydrolysable substrates AMPPNP has no additional groups that may interfere with binding. Also, a recent study on rat NTPDase2, showed that AMPPNP was the most appropriate substrate for crystal soaking (Zebisch, et al., 2008). In addition to AMPPNP, trials were conducted to determine the binding mechanism of the product AMP.

In an attempt to bind AMPPNP and AMP into the active site of 7WC two methods, co-crystallisation and substrate soaking were trialled. Co-crystallisation was performed as in section 2.8.7 and 2.8.8 with 5 mM AMP or AMPPNP and either 5 mM CaCl₂ or MgCl₂. After 1-2 weeks small rough edged crystals grew which gave poor quality X-ray diffraction. Substrate soaks were carried out by transferring pre-grown crystals into mother liquor containing with 5 mM substrate and either 5 mM MgCl₂ or CaCl₂ and incubated overnight at 18 °C. The substrate soak method gave crystals with the highest quality diffraction for both substrates.

4.9 X-ray data collection of the AMP soaked crystals.

The AMP soaked crystals were prepared in a similar fashion to that described in section 4.3 with the addition of the appropriate substrate/cofactor in the cryo-protectant solution. Diffraction data were collected using the macromolecular crystallography beam line at the Australian synchrotron at a wavelength of 0.95 Å. Images were collected over 1° increments for 180°. Images were collected using an ADSC Quantum 315r detector at a distance of 300 mm. The images were integrated with MOSFLM using the same cell dimensions and space group

determined in section 4.3. Typically the quality of the images were much higher than those collected on the home source with the mosaicity between 0.3-0.7. The final collection statistics are shown in Table 4-5.

Data collection statistics for AMP bound 7WC	
Wavelength (Å)	0.9537
Space Group	P1 2 ₁ 1
Cell Dimensions	
<i>a b c</i> (Å)	53.8/ 51.9/ 69.5
<i>α β γ</i> (°)	90 / 108/ 90
Matthew's coefficient (Å ³ /Da)	1.96
Molecules/asymmetric unit	1
Mosacity (°)	0.5
Resolution range (Å)	65.9-2.15
No of measured reflections	72376
Unique reflections	19997
<i>R</i> _{merge} [†]	0.136 (0.462)
Mean <i>I</i> / <i>σI</i> (%)	6.2 (2.5)
Completeness (%)	100 (100)
Multiplicity	3.6 (3.7)

$$^{\dagger} R_{merge} = \frac{\sum_{hkl} \sum_j |I_{hkl,j} - \langle I_{hkl} \rangle|}{\sum_{hkl} \sum_j I_{hkl,j}}$$

Table 4-5 Data collection statistics for AMP bound 7WC
Data in the parenthesis are for the outer-most shell.

4.10 Molecular replacement for the AMP structure.

Initial attempts to solve the AMP bound structure were performed using the apo-7WC structure as a model using PHASER in PHENIX. Using this method most of the C-terminal domain (Glu131-Glu403) was correctly placed however the N-terminal domain was not. To overcome this difficulty Molrep was performed with the N-terminal domain and subsequently the C-terminal domain (Murshudov, et al., 1997).

This combined model was used in PHASER which gave an initial model that was 98 % complete with an *R* and *R*_{free} value of 0.44 and 0.53 respectively. The model

was further built and improved using Arp/wARP and manual building in COOT guided by the $2|F_o| - |F_c|$ and $|F_o| - |F_c|$ maps and refined using REFMAC 5 with restrained refinement. The final model contained Ile4-Glu412 residues with R and R_{free} values of 0.19 and 0.23, respectively. A small region of a loop (Thr183-Gly184) could not be accurately placed; these residues are not placed in the model.

The ϕ and ψ angles were verified using MolProbity which showed that 95.4 % of the residues were in favoured regions and 100 % were in acceptable regions of the Ramachandran plot.

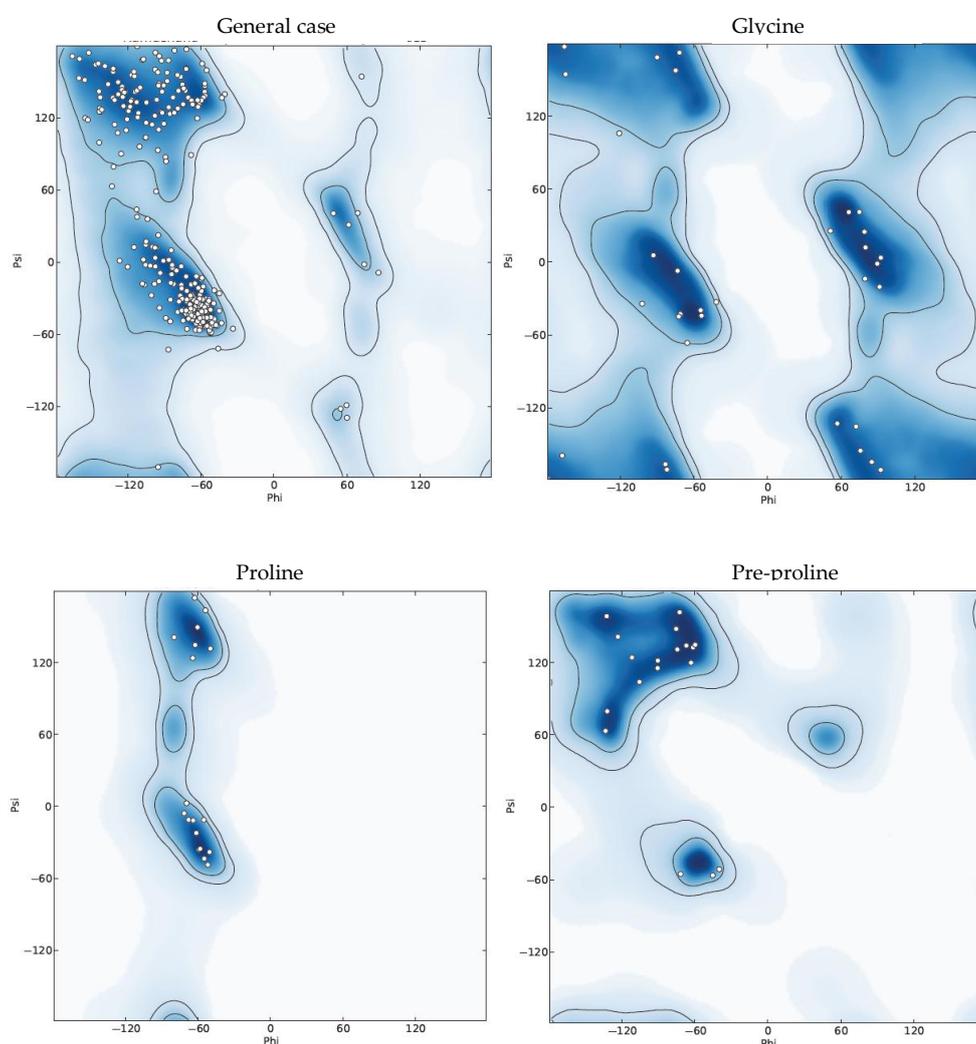


Figure 4-11 Ramachandran analysis of AMP bound 7WC.

Light blue regions indicate the preferred and the dark blue regions represent the acceptable ψ and ϕ angles.

The average B-factors for the entire structure were 34.2 Å, 44.1 Å for the N-terminal domain and 29.9 Å for the C-terminal domain. The full refinement and validation statistics are given in Table 4-6.

Refinement and model statistics for AMP bound 7WC	
Resolution range (Å)	65.94-2.15
R-factor (%)	0.191
R _{free} (%)	0.234
Wilson B-factor (Å ²)	30.8
Average B-factors (Å ²)	
Overall (No. of atoms)	34.5 (3135)
Overall protein (No. of atoms)	34.2 (2980)
N-terminus (No. of atoms)	44.1 (904)
C-terminus (No. of atoms)	29.9 (2076)
Solvent (No. of atoms)	35.9 (132)
AMP (No. of atoms)	67.0 (23)
R.M.S deviation	
Bond lengths (Å)	0.0157
Bond angles (°)	1.842
Ramachandran plot	
Percentage in favoured regions	95.4
Percentage in allowed regions	100
Percentage in disallowed regions	0

$$^{\dagger} R_{free} = \frac{\sum_{testset} \left| |F_{obs}| - |F_{calc}| \right|}{\sum_{testset} |F_{obs}|}$$

Table 4-6 Refinement and model statistics for AMP bound 7WC.
Values in parenthesis indicate the number of atoms

4.11 AMP binding mode in the active site

The AMP molecule was positioned in the structure based on $2|F_o| - |F_c|$ and $|F_o| - |F_c|$ maps contoured at 1.0σ and 3.0σ , respectively.

Examination of the modelled AMP shows the binding of the adenine ring is made possible through π -stacking with aromatic residues Tyr303 and Phe350. Such interaction would explain the range of nucleotides hydrolysed effectively by 7WC.

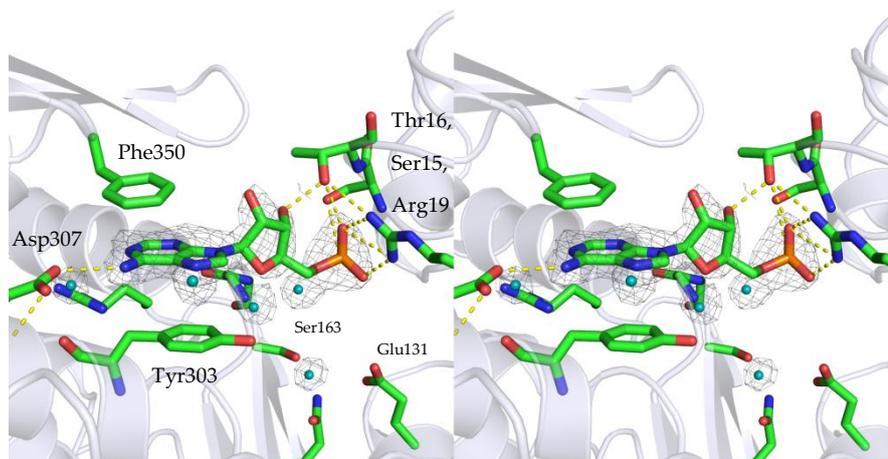


Figure 4-12 Stereo image of the active site of 7WC with AMP bound. Dashed lines indicate substrate interactions with residues. Waters are coloured in cyan.

Within the nucleotide binding pocket is Asp307 which interacts with the amino group on the adenine carbon 4. When combined with the kinetic data (which shows the preferred substrates are adenosine and cytosine - both contain amino groups this suggests that Asp307 contributes to the selection of nucleotides (discussed further in section 6.4). The 3' hydroxyl of the ribose interacts with the side-chain hydroxyl of Thr16 that would help position the phosphate into the phosphate binding site 1. The α -phosphate of AMP associates with the side chain of Arg19, with the backbone amides of Ser15 and Thr16 and possibly through a water mediated interaction with Gly161. The α -phosphate of AMP is 6.6 Å from the nucleophilic base, Glu131, so it is unlikely that hydrolysis could occur using the same catalytic mechanism of that proposed for *R. novogicus* NTPDase2 (Zebisch, et al., 2008). This is supported by the inability for 7WC to hydrolyse AMP (section 6.2.3).

Despite 5mM CaCl₂ being present in the soak solution, no metal ion was visible in the active site. Previous work on the members of the actin super-family would suggest that the metal ions would associate via 4 water molecules to Asp12, Thr38, Glu131, Asp158, Ser299 and Trp393 (Bork, et al., 1992; Drosopoulos, 2002; Smith, et al., 1999a; Zebisch, et al., 2008).

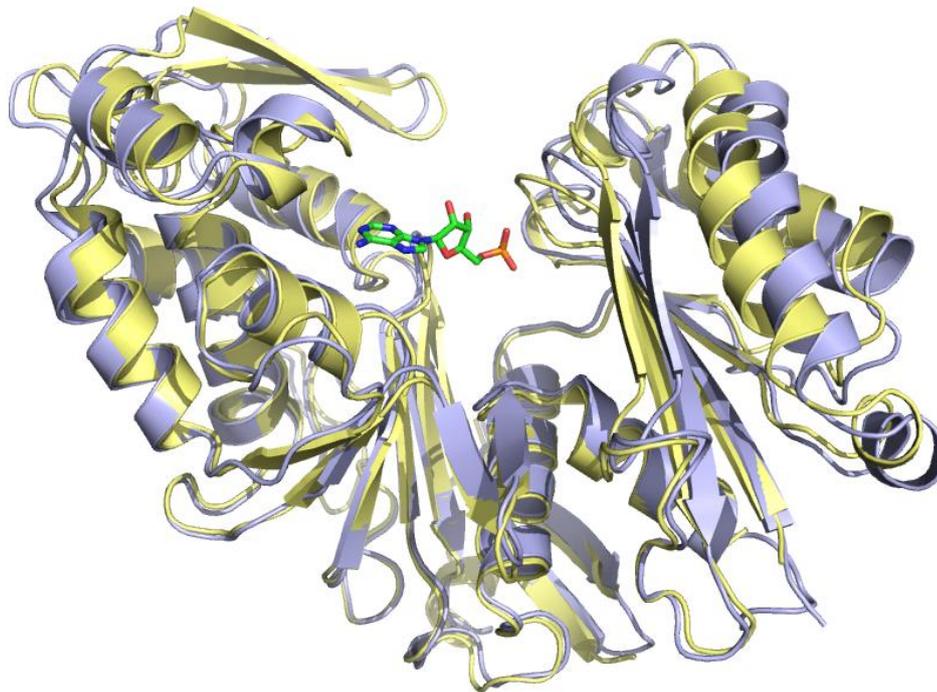


Figure 4-13 Superimposition of the apo-7WC and AMP bound 7WC using SSM. The apo-AMP is shown in purple whereas the AMP bound 7WC is in gold

4.12 Overall structure and comparison of the AMP bound with the apo 7WC structure.

Comparing the apo 7WC and the AMP bound structures by SSM reveals a domain shift (Figure 4-13). Due to the NTPDases homology with exopolyphosphatase/guanosine pentaphosphate phosphohydrolase and other members of the actin superfamily which have demonstrated conformational changes upon substrate binding (Kristensen, et al., 2008; Schuler, 2001) it has been hypothesised that this phenomenon also occurs for NTPDases but has not been shown. Further analysis of the substrate induced trans-domain shift will be discussed later in section 5.4.

The domain movement within the crystal results in higher B-factor values (Table 4-2 and Table 4-6). Visualisation of the B-factors by Pymol B-factor putty show high B-factors as thicker lines (Figure 4-14) and reveals that the difference is largely in the N-terminal domain.

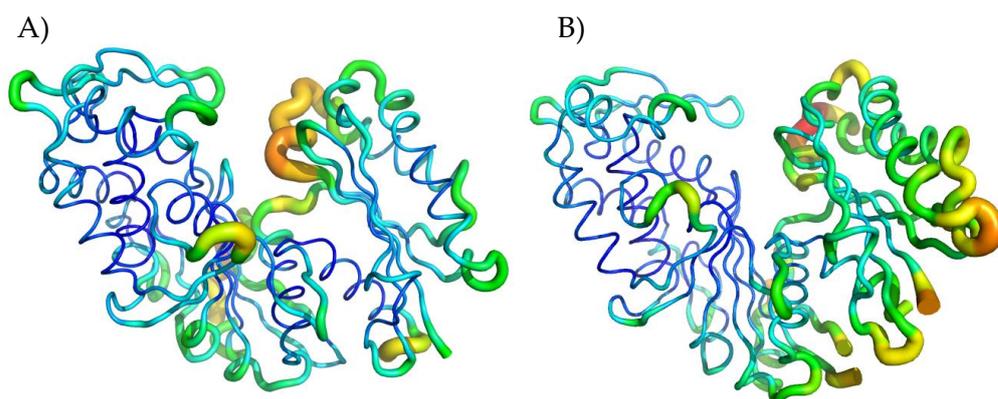


Figure 4-14 Pymol B-factor putty.
Decreasing B-factors are coloured; red, yellow, green, cyan, blue. A) apo 7WC, B) AMP bound 7WC.

The B-factors can be interpreted as the rms error in the atoms position. Differences in the crystal packing and the thermal motion can affect the B-factor value. This difference in B-factors illustrates the propensity of the atoms, particularly in the N-terminus to move which is in agreement with the DynDom analysis (section 5.4.1).

The higher B-factors could also be attributed to crystal packing constraints on the moving domains. If this is the case then the movement of the domains may be an under estimation. To overcome this problem co-crystallisation with AMP was performed. However this produced very poor quality crystals with bad quality diffraction data suggesting differences in the crystal packing.

4.13 Comparison of the AMP bound 7WC with other homologous structures.

Of the published NTPDase structures only the *R. novogicus* NTPDase2 has an AMP bound structure with (3CJ7) and without (3CJ9) a phosphate in the active site (Zebisch, et al., 2008). No phosphate was observed in the AMP bound structure of 7WC.

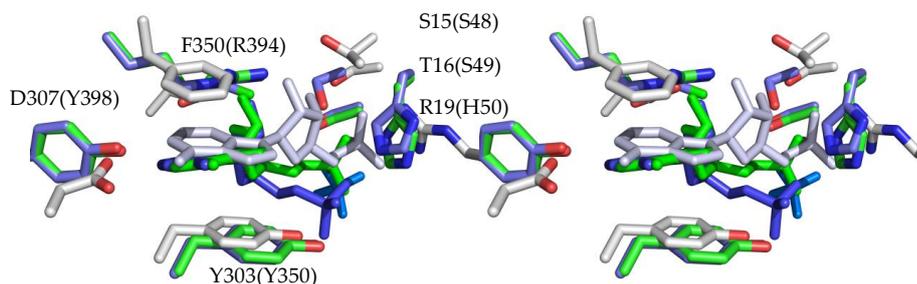


Figure 4-15 Stereo image of the active site comparison of *R. novogicus* NTPDase2 and 7WC.

Superimposition of 7WC (grey), NTPDase2 with AMP (green) and NTPDase with AMP and phosphate (blue). Labels in parenthesis are for NTPDase2.

The active site residues are very similar between the two structures with only a few substitutions; Arg394 of NTPDase2 is replaced with Phe350 in 7WC and this aromatic residue contributes to π -stacking of the adenine base. Ser49 corresponds with Thr16 and the most notable the His50 of NTPDase2 is in the same position as the Arg19 which rotates to fill the corresponding space. Alignment of NTPDase using protein sequences identify that an arginine is conserved in plant and prokaryotic NTPDases and in human NTPDase3, 4, 5, 6 and 7 whereas histidine has been observed for the mammalian NTPDase1, 2 and 8. Two structures of AMP binding have been published; one included a phosphate molecule deeper in the active site cleft and one with AMP only. Comparing the binding mechanism of AMP in 7WC shows that orientation has a higher similarity to the model without the phosphate in the active site. It was proposed that the phosphate deeper in the active site pocket of NTPDase2 is thought to distort AMP through charge repulsion (Zebisch, et al., 2008).

Similar to the 7WC structure, NTPDase2 with AMP bound has no visible metal ion in the structure. It could be speculated that the AMP, metal ion and phosphate leave the active site in a sequential manner with the AMP leaving the binding site last.

In contrast to 7WC, the NTPDase structure showed no domain movement to accommodate the binding of the AMP. Zebisch and Strater, 2008, are open to the idea that domain movement does occur in NTPDase2 and argue that this was either due to the trans-membrane domains being absent (shown previously for mammalian NTPDase1 (Grinthal, et al., 2007) or as a consequence of crystal packing constraints.

4.14 X-ray data collection of the AMPPNP soaked crystals

Data were collected and processed as in section 4.9 at the Australian synchrotron. Images were collected over 360 ° at 1 ° increments with the detector set 300 mm away.

4.15 Data processing of the AMPPNP soaked crystals.

The images covered the resolution range from 72 Å to 1.6 Å. But reflections were not visible in the outer shell. Despite this, the images were merged and integrated using the entire resolution range. This resulted in an R_{merge} in the outer shell of 0.725. The SCALA output was analysed using Xtriage contained within PHENIX to evaluate the $I/\sigma I$ for each resolution shell. An $I/\sigma I > 3.0$ was used to determine the resolution was cut off at 1.76 Å ($I/\sigma I$ of 3.13). Collection statistics are shown in Table 4-7.

Data collection statistics for AMPPNP bound 7WC	
Wavelength (Å)	0.9591
Space Group	P1 2 ₁ 1
Cell Dimensions	
<i>a b c</i> (Å)	53.9, 52.8, 72.1
α β γ (°)	90, 94.9, 90
Matthew's coefficient (Å ³ /Da)	2.18
Molecules/asymmetric unit	1
Wilson B-factor (Å ²)	17.27
Mosacity	0.7
Resolution range (Å)	71.9-1.50
No of measured reflections	195986
Unique reflections	44133
<i>R</i> _{merge} [†]	0.136 (0.725)
Mean I/σI (%)	8.7 (0.7)
Completeness (%)	98.5 (96.0)
Multiplicity	4.4 (1.4)

$$^{\dagger} R_{merge} = \frac{\sum_{hkl} \sum_j |I_{hkl,j} - \langle I_{hkl} \rangle|}{\sum_{hkl} \sum_j I_{hkl,j}}$$

Table 4-7 Data collection statistics for AMPPNP bound 7WC

4.16 Molecular replacement for the AMPPNP-bound structure.

Molecular replacement of the AMPPNP-bound structure was achieved by PHASER in the PHENIX suite of programs using the entire apo-7WC model. This resulted in a model containing 99 % of the residues (Ile4-Lys410) with an R value of 0.33. Further model improvement was undertaken as in section 4.5 which gave a final model that contains 406 (Thr5-Phe411) out of 426 residues with R and *R*_{free} values of 0.18 and 0.23, respectively. One region (Ala181-Ala184) could not be resolved with 2|F_o|-|F_c| and |F_o|-|F_c| maps contoured at 1.0 σ and 3.0 σ and were omitted from the model.

The average B-factors are slightly higher than the values for the apo-structure with an average B-factor of 20.23 Å² for the entire model.

Protein geometry was analysed by Molprobrity and revealed that 99.5 % of all residues had acceptable Ramachandran angles with 2 outliers; Ala181 and the terminal residue Lys410.

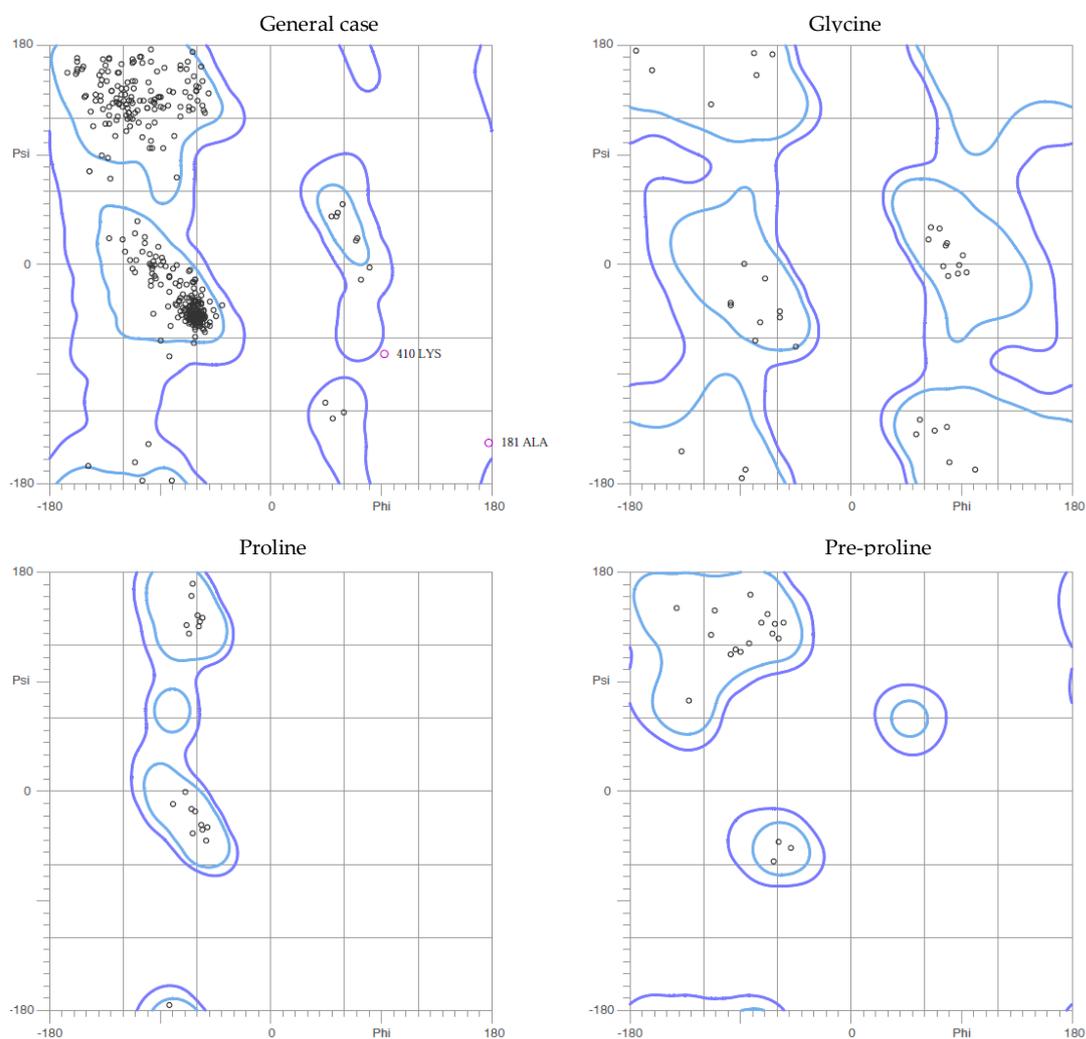


Figure 4-16 Ramachandran analysis of AMPPNP bound 7WC.

Light blue regions indicate the preferred and the dark blue regions represent the acceptable ψ and ϕ angles.

Refinement and model statistics for AMPPNP bound 7WC	
Resolution range (Å)	71.98-1.76
R-factor	0.18
R _{free}	0.23
Average B-factors (Å ²)	
Overall (No. of atoms)	22.2 (3532)
Overall protein (No. of atoms)	20.2 (3101)
N-terminus (No. of atoms)	22.5 (919)
C-terminus (No. of atoms)	19.2 (2182)
Solvent (No. of atoms)	31.3 (390)
Active site molecules	
AMPPNP	29.0 (31)
PO4	45.0 (5)
R.M.S deviation	
Bond lengths (Å)	0.007
Bond angles (°)	1.091
Ramachandren plot	
Percentage in favoured regions	96.4
Percentage in allowed regions	99.5
Percentage in disallowed regions	0.5
†	
$R_{free} = \frac{\sum_{testset} \left F_{obs} - F_{calc} \right }{\sum_{testset} F_{obs} }$	

Table 4-8 Refinement and model statistics for AMPPNP bound 7WC.
Values in parenthesis indicate the number of atoms

The overall AMPPNP structure more closely resembles the apo structure, with a rmsd of 0.4 Å in contrast to the AMP bound structure with an rmsd of 1.4 Å (Figure 4-17). The larger deviation with the AMP bound structure is due to domain movement rather than secondary structure differences since comparison of the AMP structure with the N- and C-terminus of the AMPPNP bound structure have an rmsd of 0.6 and 0.7, respectively.

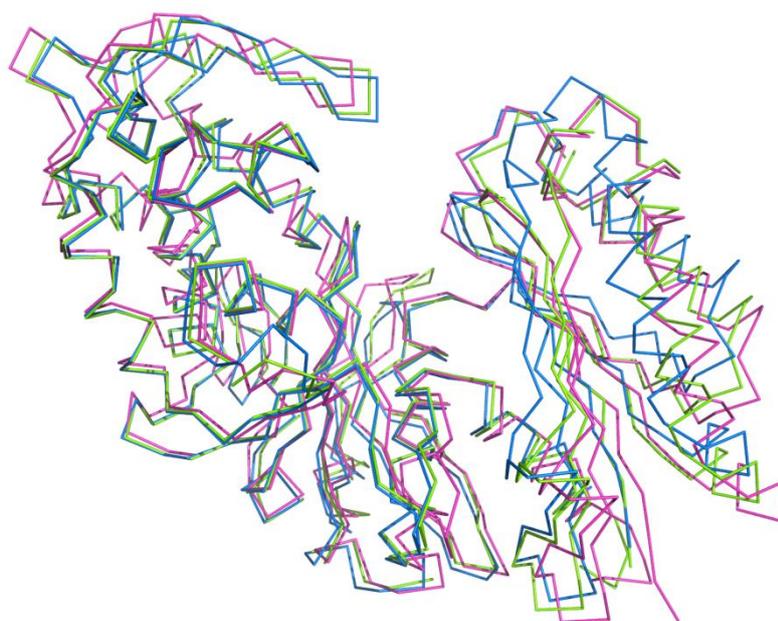


Figure 4-17 SSM superimposition of apo (pink), AMP bound (blue) and AMPPNP (green) 7WC structures.

4.17 The active site of the 7WC containing AMPPNP

To date, there are two crystal structures which have demonstrated AMPPNP binding; the *R. novogicus* NTPDase2 and *L. pneumonia* NTPDase1 (LpNTPDase1) (Vivian, et al., 2010; Zebisch, et al., 2008). These structures were overlaid using SSM and the active site was observed Figure 4-18.

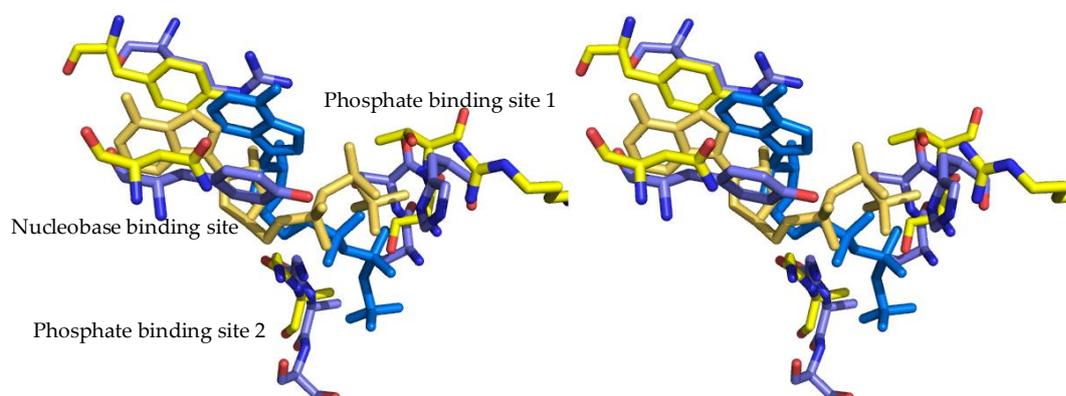


Figure 4-18 Stereo image of a comparison of the binding mechanism of AMPPNP in LpNTPDase1 and NTPDase2.

LpNTPDase1 is shown with yellow residues and gold substrate whereas NTPDase2 is shown with purple residues and, blue substrate. For residue names refer to (Table 4-3).

The binding mode for LpNTPDase1 is dissimilar to that of NTPDase2. In particular, the placement of the phosphate tail. In LpNTPDase1 the α - and β -phosphates bind to Asp186, Gly189 and A190 through a number of water mediated interactions whereas in NTPDase2 they are shown to be associated with the phosphate binding site 1. In the LpNTPDase1 the γ -phosphate interacts with the residues in the N-terminal domain phosphate binding site 1 (Ser52, Thr53 and Arg56) whereas in NTPDase2 the γ -phosphate binds to the N-terminal domain phosphate binding site 2 (Gly204 and Ala205) positioned to interact with the catalytic residues.

In the AMPPNP bound model of 7WC there was positive $|F_o|-|F_c|$ (contoured to 3σ) density to support both conformations. As the density was poor, the location of the AMPPNP was estimated using these models as guides.

4.18 Conformation of AMPPNP based on LpNTPDase1 structure

A recent study showed that the catalytically essential residues in *L. pneumophila* were found to be homologous to that of the mammalian NTPDase suggesting the same catalytic mechanism is shared by LpNTPDase1 and NTPDase2 (Sansom, et al., 2008). Despite this, the conformation of AMPPNP in LpNTPDase1 is orientated in a fashion where the γ -phosphate is $\sim 9\text{ \AA}$ from the site of hydrolysis and is unlikely to illustrate the correct binding mechanism of AMPPNP required for catalysis.

The AMPPNP in 7WC was positioned using the $|F_o|-|F_c|$ map (contoured to 3σ) based on the conformation as the LpNTPDase1 where the γ -phosphate resided in the phosphate binding site and was refined using REFMAC to confirm the position (Vivian, et al., 2010). Fine tuning of the binding position was guided by the positive $|F_o|-|F_c|$ map (contoured to 3σ).

The adenine, the phosphate tail and parts of the ribose fit well in the $2|F_o|-|F_c|$ map (contoured to 1σ) with no negative density in the $|F_o|-|F_c|$ map. A small

amount of positive $|F_o| - |F_c|$ density was observed around the adenine ring and oxygen of the γ -phosphate (Figure 4-19).

In this conformation the adenine has the same contacts as the AMP bound 7WC, where nitrogen in the adenine ring can form hydrogen bonds with Asp307 and the carboxylate of Gln306 (4.11). The α -phosphate is orientated to associate with the side chain of Ser300 and Leu159 via a water molecule, the β -phosphate forms an hydrogen bond with the Gly161 of the phosphate binding site 2 in the same manner as LpNTPDase1. The γ -phosphate has a myriad of associations in phosphate binding site 1, including the side chains of Ser15, Thr16, Arg19 and a water mediated interaction with Asp12. Ribose is present in a C2'-endo conformation in this structure, however, only partial density accounts for the position so it is conceivable that it is flexible.

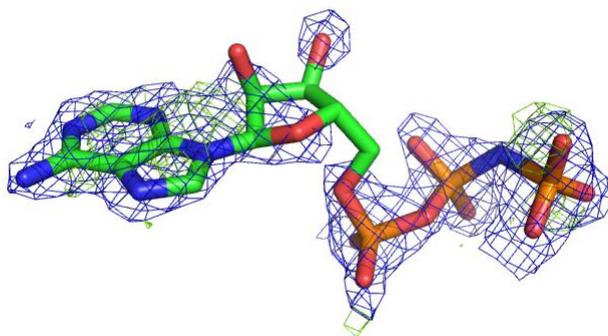


Figure 4-19 Close up of the AMPPNP with electron density map. $2|F_o| - |F_c|$ map (blue) contoured at 1σ and the $|F_o| - |F_c|$ map (green) contoured at 3σ .

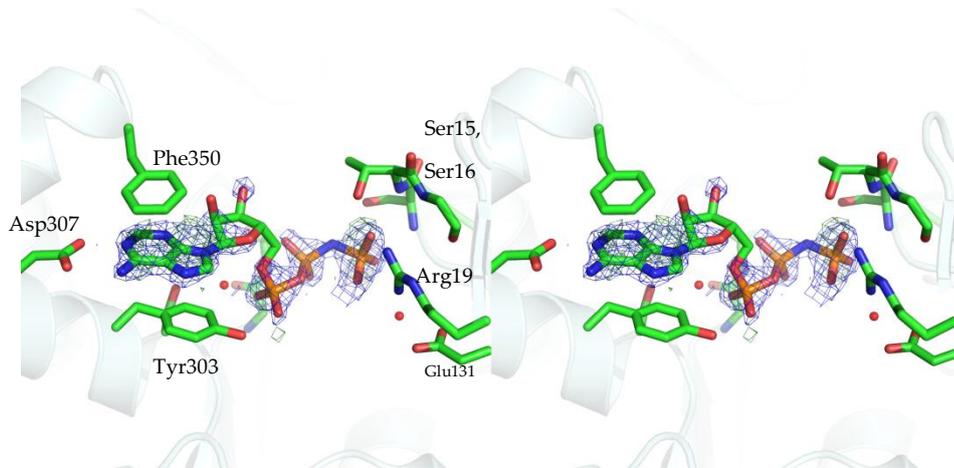
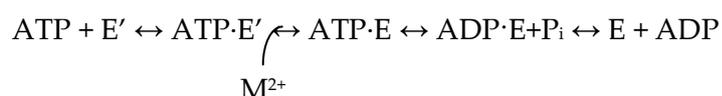


Figure 4-20 Stereo image of the binding mechanism of AMPPNP in 7WC based on the conformation of LpNTPDase1.

Residues that interact are shown as sticks. Waters are coloured red. Map represents the $2|F_o|-|F_c|$ (blue) contoured at 1σ and the $|F_o|-|F_c|$ map (green) contoured at 3σ .

B average analysis (CCP4) reveals that the AMPPNP had an average B-factor of 29.0 \AA^2 which indicates that this conformation is plausible. However the γ -phosphate of AMPPNP in this conformation is positioned in a way that γ -phosphate is 6.23 \AA away from the catalytic residues such as Glu131 and Ser163. Point mutation of Ser163 to Ala163 in 7WC resulted in no detectable activity for either ATP or ADP (data not shown) which with other studies show that it is unlikely that there is another catalytic site in NTPDases (Kirley, et al., 2006; Sansom, et al., 2008). For hydrolysis to occur the phosphate tail needs to enter further down the active site cleft, which could be mediated by the inter-domain movement. Domain movement was not observed in either LpNTPDase1 or 7WC. Another parallel between the two models is the absence of a metal ion bound in the crystal structure despite being present in the soak solutions. Previous experiments on G-actin have shown that the domain movement and the binding of the metal ion are intrinsically linked, where ATP alone was unable to induce a conformational change and only when a divalent cation was added does such a change occur (Frieden, et al., 1988). It has also been shown that the addition of the divalent cation switches the actins affinity for ATP from low to high (Kinosian, et al., 1993) and that this binding also induces the movement of domains (Frieden, et al., 1988; Schuler, 2001).

With this information it could be conjectured that since neither the metal binding nor domain movement was observed in either the LpNTPDase1 or 7WC that they correspond to an alternate binding state and that this conformation of AMPPNP represents the un-catalytic binding mode. The inclusion of the divalent cation would induce a different binding mode of AMPPNP, which may resemble that of the *R. novogicus* NTPDase2, which is then able to be catalysed.



Equation 6. Proposed method of hydrolysis.

E' and E represent the un-catalytic and catalytically active state of the enzyme.

4.18.1 Conformation of AMPPNP based on the NTPDase2 structure

The AMPPNP was placed into 7WC with the assumption that the binding mechanism was similar to that of the *R. novogicus* NTPDase2 where the γ -phosphate is positioned to be hydrolysed by the catalytically active Glu131 (Drosopoulos, et al., 2000). In this conformation electron density clearly supports adenosine base stacked between the Tyr303 and Phe350, a phosphate ion that associates with the Ser15, Thr16 and Arg19 (phosphate binding site 1) and the γ -phosphate that associates with the peptide bonds of Gly161, Gly162 and Ser163 (phosphate binding site 2). However, the position of the ribose, α -phosphate and β -phosphoramidate are not well supported by the $2|F_o| - |F_c|$ map (contoured to 1σ). Despite the electron density for a phosphate molecule binding to the phosphate binding site 1, the inter-domain cleft was too large to allow for the phosphate tail of AMPPNP to bind to both phosphate binding sites as was observed in rat NTPDase2.

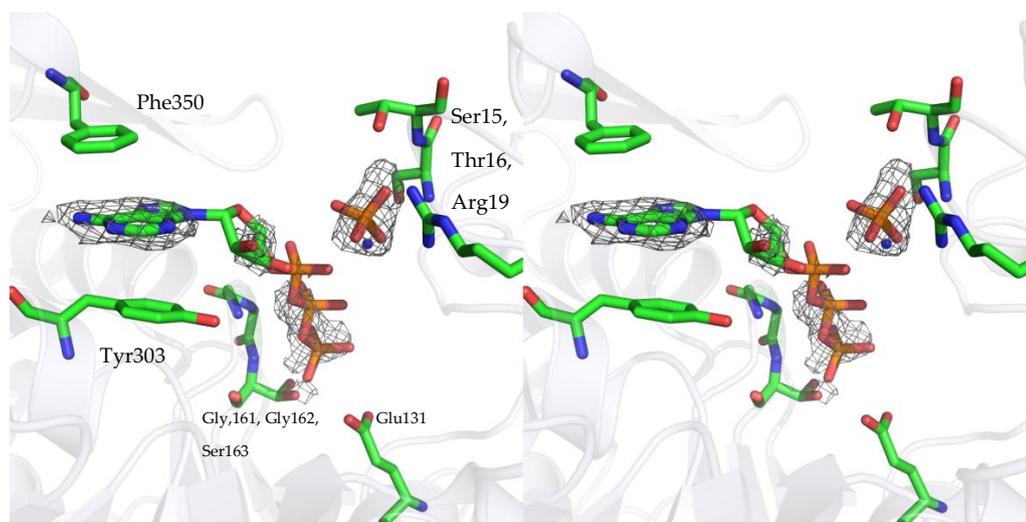


Figure 4-21 Stereo image of the active site of 7WC with AMPPNP bound in conformation 1.

The $2|F_o|-|F_c|$ map contoured to 1σ is shown in grey.

For these residues to be involved in the binding of the substrate the active site cleft width needs to be smaller either through the movement of the loop (Val10-Val20) containing the phosphate binding site 1 residues or movement of the C-terminal domain. It is therefore likely that a free phosphate binds to the first binding site observed in the apo-7WC structure (section 4.6). The AMPPNP bound NTPDase2 model showed a metal ion coordinated octahedrally with the β - and γ - phosphate and with four water molecules. However, there was no evidence of a metal binding in the 7WC structure.

The average B-factor analysis for AMPPNP and phosphate was 94.88 \AA^2 and 72.8 \AA^2 , respectively (Bailey, 1994). Taking into account all the data, there is no evidence that this represents the correct binding mechanism of AMPPNP.

4.19 Discussion

The apo-7WC structure was solved to 1.89 \AA resolution. This showed that 7WC exists as two domains that are bridged by a linker region. Consistent with other members of the actin superfamily, the two domains resemble an RNase-H domain where the C-terminal domain is modified with β -strand insertions within the domain making it larger than the N-terminal domain. Despite the low

homology in sequence 7WC had high structural homology with other NTPDase structures and other members of the actin superfamily. An important character of NTPDases is the ACR that contribute residues that make both sides of the active site cleft. The ACRs form two conserved phosphate binding sites and a nucleobase binding site. Using substrate crystal soaks and the binding mode of AMP and AMPPNP was found. The nucleobase of AMP associates with 7WC through π - π interactions with aromatic side chains, which shows how 7WC can accept a range of substrates. Analogous to the *R. novogicus* NTPDase2 the α -phosphate associates with the N-terminal phosphate binding site.

The binding mode of AMPPNP in 7WC was resolved. However, divalent cation binding (required as a co-factor of catalysis) was not observed as demonstrated in the *R. novogicus* NTPDase2. The most contrasting feature of the binding mode between the NTPDase2 and 7WC is the position of the phosphate tail. In 7WC, the AMPPNP is bound in a way that the γ -phosphate is bound to the phosphate binding site 1, where α - and β -phosphates of AMPPNP in NTPDase2 are. The orientation of AMPPNP in 7WC is unable to be hydrolysed as the terminal phosphate is too far from the catalytic residues for hydrolysis to occur, thus, represents an un-catalytically active binding site. From this it was conjectured a binding sequence, where the nucleotide first binds to the NTPDase in the alternative binding site. A divalent ion subsequently binds which brings about movement of the phosphate tail deeper into the active site. This model is supported by *Legionella pneumonia* LpNTPDase1 structure, which showed the same binding mechanism of AMPPNP (without divalent cation) and in-depth kinetic characterisation of other members of the actin superfamily.

It would be interesting to evaluate the binding kinetics of ATP or AMPPNP to this alternative binding site as this site may act as a high affinity binding site to recruit substrate. This would be beneficial if either the metal or the substrate concentration was limiting.

5 DbLnP crystallisation

5.1 Introduction

DbLnP has been known for many years to play an integral role in early stages of nodulation. Despite the importance that DbLnP plays, there is no understanding of the structural or catalytic mechanism. This chapter investigates these aspects to gain a better insight into the catalytic function of DbLnP.

No previous research has been conducted on the crystallisation of DbLnP, so extensive trials were conducted to find adequate crystals. To reveal the binding mechanism of substrate to DbLnP, attempts to crystallise the enzyme with a non-hydrolysable substrate were investigated and a catalytic mechanism is proposed. Further, analysis of the putative inter-domain movement that has been proposed in previous NTPDase studies was investigated.

5.2 Crystallisation of DbLnP

Using conditions identified to be successful for other NTPDases such as 7WC (section 4.2.1), rat NTPDases and human NTPDase1, crystallisation trials were conducted (Zebisch, et al., 2008; Zhong, et al., 2008). Despite DbLnP crystals forming, no conditions produced crystals that diffracted well. In an attempt to produce better crystals the His-tag was cleaved from the expression construct after the final size exclusion chromatography step using thrombin cleavage.

5.2.1 Thrombin cleavage

The full length expression construct contains an N-terminal His-tag followed by a thrombin cleavage site (see appendix, section 8.1). This made it possible to remove the first 20 residues from the N-terminus including the His-tag using thrombin.

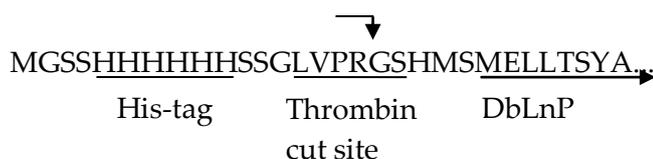


Figure 5-1 Thrombin cleavage site of the N-terminus of DbLnP

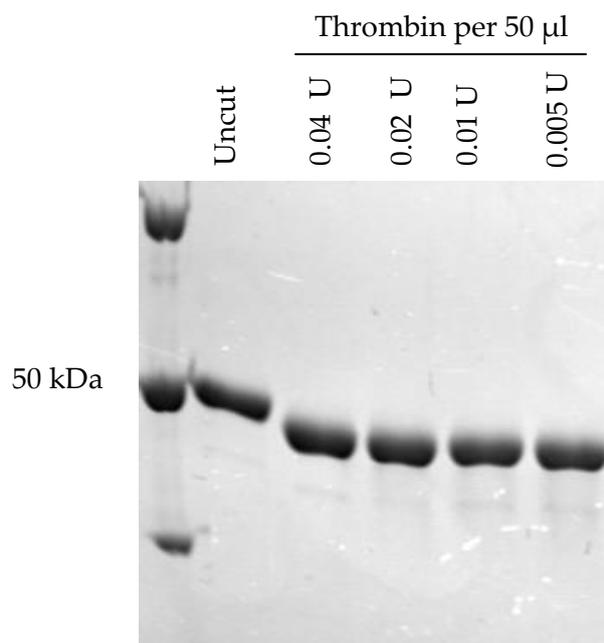


Figure 5-2 SDS-PAGE of the small scale optimisation of thrombin cleavage of DbLnP

The extent of the cleavage was evaluated using SDS-PAGE comparing cut with uncut DbLnP (Figure 5-2). From the small scale cleavage experiments the lowest dilution (0.005 U) of thrombin was able to effectively cleave the His-tag from DbLnP after 2 hr. A large scale cleavage reaction was proportional to the small scale as described in section 5.2.1 and evaluated for purity by S200 size exclusion chromatography. The thrombin cleaved DbLnP was separated from any DbLnP that was not cleaved by thrombin by performing IMAC in native conditions and eluting the thrombin cleaved DbLnP in buffer containing 50 mM imidazole.

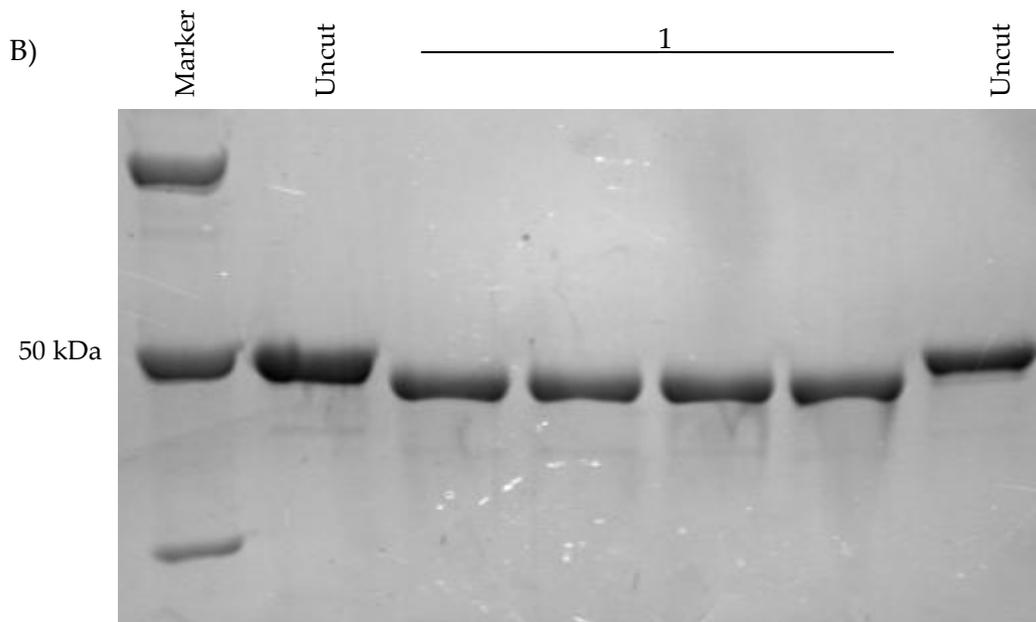
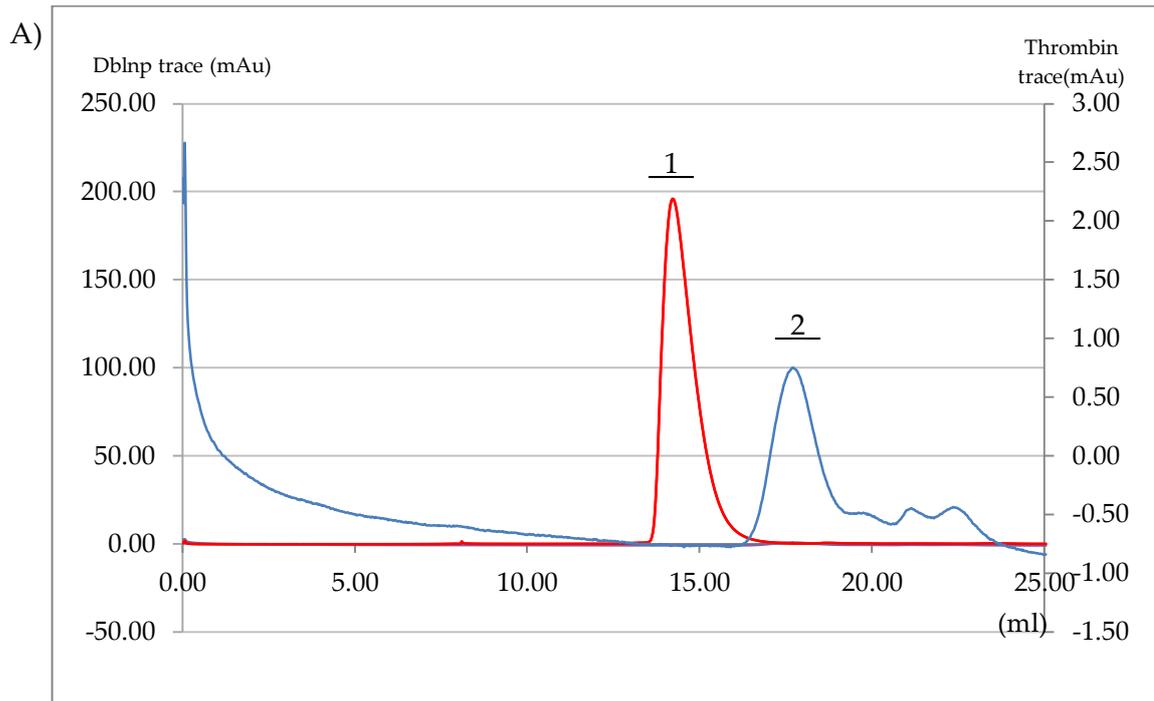


Figure 5-3 Size exclusion of thrombin-treated DbLnp.

A) Red trace shows thrombin treated DbLnp. The blue trace shows the elution of thrombin. B) SDS-PAGE of size exclusion chromatography showing the thrombin treated DbLnp. The concentration of thrombin (not shown) is too low to be detected by Coomassie stained SDS-PAGE.

To ensure that it was possible to separate thrombin from thrombin treated DbLnP by size exclusion chromatography 10 μ l of undiluted thrombin (\sim 1.0 U) was compared with thrombin treated DbLnP. The major peak of thrombin eluted at \sim 17 ml which was in agreement with the molecular weight of monomeric thrombin (37 kDa). The difference in elution volume is sufficient that the thrombin treated DbLnP can be isolated from thrombin using size exclusion chromatography. This protein was used for further crystallisation trials.

5.2.2 Crystallisation of thrombin treated DbLnP

The thrombin treated DbLnP was purified as described in section 5.2.1 and concentrated to 10 mg·ml⁻¹ before centrifugation at 13,000 x·g for 5 min at 4 °C to remove any particulates before being transferred to crystallisation plates. AMPPNP and MgCl₂ were added to the protein solution to a final concentration of 5 mM for both. The broad screens were performed using the Cartesian honeybee crystallisation robot as described in section 2.8.2. Observation of the plates after 4 weeks showed 5 conditions containing protein crystals that gave a positive izit dye test (section 2.8.6).

5.2.3 Further crystallisation trials

Crystallisation of thrombin-treated DbLnP was further investigated using the hanging-drop diffusion method as described in section 2.8.3.1.

Two conditions were investigated and will be reported here. These were based on successful crystallisation conditions identified for rat NTPDase2 (crystal form 1) or conditions identified by the robot screens (crystal form 2). Evaluation of the quality of the crystals by X-ray diffraction at the Australian synchrotron identified two crystal types that produced X-ray diffraction to \sim 3 Å resolution.

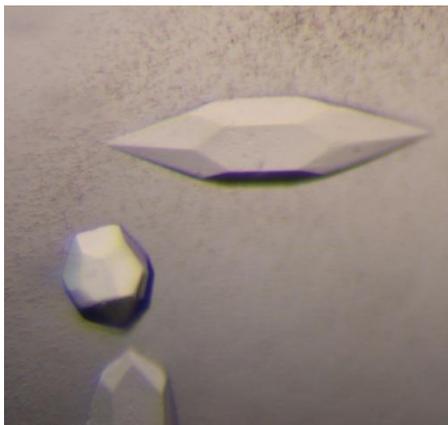


Figure 5-4 Crystal type 1 of DbLnP co-crystallised with AMPPNP

5.2.3.1 *Crystal form 1 of thrombin-treated DbLnP*

The best data were collected from crystals grown using a mother liquor that contained 200 mM ammonium phosphate, 25 % MPD, 100 mM HEPES, pH 8.0, 1-5 % PEG 3350, 5 mM MgCl₂, 5 mM AMPPNP and 5 mM spermidine. These formed after 2 months as 0.5 mm hexagonal prisms where the ends taper to a single point (Figure 5-4). It was possible to substitute AMP for AMPPNP and maintain the same crystal morphology, but in the absence of substrate, smooth ellipsoid crystals formed which did not diffract X-rays.

5.2.3.2 *Crystal form 2 of thrombin-treated DbLnP*

The second condition contained 200 mM MgCl₂, 50 mM acetate, pH 5.0, 9-12 % MPD, 10-20 % PEG 3350, and 5 mM AMPPNP. Crystals only grew after micro-seeding from other DbLnP crystals which after 2 weeks produced 0.2 x 0.4 mm plates and rods that would typically nucleate from the same point (Figure 5-5-A). Similar crystal morphology was observed if the AMPPNP was substituted with AMP. With no substrate in the precipitant small equilateral triangle plates with ~ 0.1 mm edges formed that showed no X-ray diffraction (Figure 5-5-B).



Figure 5-5 Crystals of DbLnP.

A) DbLnP with either AMP or AMPPNP, B) DbLnP with no substrate.

5.2.4 Data collection

5.2.4.1 *Crystal form 1 of thrombin-treated DbLnP*

The DbLnP crystals produced using the first method were soaked in the mother liquor with increasing amounts of PEG 3350 to a final concentration of 40 % before being flash frozen in liquid nitrogen. A data set was collected using the Australian synchrotron as described in section 2.10.2. The best data set was collected for AMPPNP co-crystallised crystals produced X-ray diffraction data up to 2.8-3.0 Å resolution.

Determination of the space group and unit cell parameters proved very difficult using MOSFLM which commonly gave inconsistent cell size and space group. Attempts to measure the reflections at thin regions of the crystal reduced the detected lattices and improved the results, however not enough for further analysis. Dehydration and annealing of the crystal were attempted but typically reduced the resolution. Dehydration and annealing of the crystal were attempted but typically reduced the resolution. Due to the difficulty in processing these data this crystal form was abandoned.

5.2.4.2 *Crystal form 2 of thrombin-treated DbLnP*

DbLnP crystals were soaked in mother liquor with increments of 5 % glycerol to a final concentration of 20 % before being flash cooled in liquid nitrogen (100 K).

A data set was collected for DbLnP co-crystallised with either AMPPNP and AMP as described in section 2.10.2. The AMPPNP and AMP crystals produced X-ray diffraction data to 2.6-3.0 Å using radiation from the Australian synchrotron (Figure 5-6). The images were collected over 360° in 1° increments with 1 second exposures and a detector distance at 270 mm. Images 90° apart were used to determine the space group with MOSFLM for both AMP and AMPPNP co-crystallised DbLnP. The unit cell was found to be monoclinic ($a = 119$, $b = 71$, $c = 78$, $\alpha = 90$, $\beta = 129$, $\gamma = 90$) with a C2 space group. The data set was integrated with MOSFLM using all reflections.

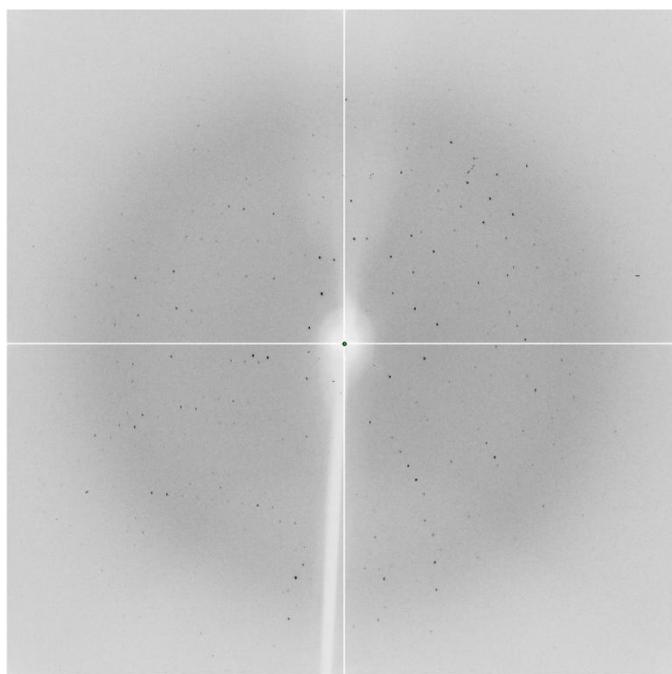


Figure 5-6 X-ray diffraction of crystal form 2 of DbLnP co-crystallised with AMP. The edge of the plate represents 2.5 Å.

5.2.5 Data processing of the AMP and AMPPNP co-crystallised DbLnP

The Matthew's coefficient analysis for both the AMP and AMPPNP crystals showed that there was 1 molecule in the asymmetric unit with 57.2 % solvent content and a Matthew's coefficient of $2.88 \text{ \AA}^3/\text{Da}$ (Matthews, 1968). The reflections were scaled and merged using SCALA using 2.6 Å and 2.9 Å resolution as the cut-off for the AMP and AMPPNP co-crystallised crystals,

respectively. A summary of the data statistics are shown for the AMP and the AMPPNP crystals in Table 5-1.

Data collection statistics.	AMP + DbLnP	AMPPNP + DbLnP
Wavelength (Å)	0.9536	0.9537
Space Group	C2	C2
Cell Dimensions		
<i>a b c</i> (Å)	120.1/71.1/78.1	119.9/71.2/77.8
<i>α β γ</i> (°)	90.0/ 129.0/ 90.0	90.0/128.8/90.0
Matthew's coefficient (Å ³ /Da)	2.88	2.88
Molecules/asymmetric unit	1	1
Mosacity	0.54	0.17
Resolution range (Å)	46.7-2.60	46.7-2.9
No of measured reflections	122050	75859
Unique reflections	16426	11389
$R_{\text{merge}}^{\dagger}$	0.10 (0.38)	0.33 (0.478)
Mean $I/\sigma I$	15.4 (5.0)	5.4 (3.5)
Wilson B-factor (Å ²)	29.86	31.47
Completeness	99.9 (99.8)	99.5 (99.9)
Multiplicity	7.4 (7.2)	6.7 (7.5)

†

$$R_{\text{merge}} = \frac{\sum_{hkl} \sum_j |I_{hkl,j} - \langle I_{hkl} \rangle|}{\sum_{hkl} \sum_j I_{hkl,j}}$$

Table 5-1 Data processing statistics for AMP and AMPPNP co-crystallised DbLnP
Data in the parenthesis are for the outer-most shell.

5.2.6 Molecular replacement of the DbLnP co-crystallised with AMP

Structural determination was achieved via molecular replacement using 7WC as a model which has 63.3 % sequence identity with DbLnP. 7WC is slightly larger with two insertions in the amino acid sequence (Asn314 - Lys315 and Phe350). The numbering of the DbLnP structure was aligned with 7WC for ease of analysis, resulting in the full length construct of the thrombin-digested DbLnP containing five residues before the Leu1. These residues do not feature in the DbLnP structures. Due to the high sequence identity it was not necessary to truncate the model with CHAINSAW. To accommodate differences in the inter-

domain angles, two models were made; the N- and C- terminal domains which were selected based on those used for 7WC structural determination (Leu4-Ile126 and Asp127-Glu412). These models were used independently in PHENIX AutoMR which found one solution using each model. Both models gave statistically significant molecular replacement solutions with log likelihood gain of 175.3 and 98.0 for the N- and C-terminal, respectively. The C- α chain of the N-terminal domain was resolved well. However the C-terminal domain had a region that was poorly resolved which included a solvent exposed loop turn (Thr174-Pro179) and a region of the C-terminal α -helix (Glu402-Phe411) that made up part of the ACR5. The N- and C- terminal domains were merged manually to form the complete model containing 390 of the 420 well resolved residues with an initial R value of 0.46 and an R_{free} of 0.47.

5.2.7 Model refinement of the DbLnP co-crystallised with AMP model

The merged model was improved and completed by manually building residues in COOT directed by the $2|F_o| - |F_c|$ and $|F_o| - |F_c|$ maps contoured to 1σ and 3σ respectively and verified using either Refmac 5 (CCP4) or phenix.refine. Subsequent iterations of the model were produced by simulated annealing as part of phenix.refine program and by defining two domains as rigid domains (identified by TLSMD) and performing TLS refinement in Refmac 5 (section 2.11.6). The TLS rigid domains were determined using the TLSMD server (<http://skuld.bmsc.washington.edu/~tlsmd>) which showed there were possibly 3 rigid domains, identified by a small change in angle in the residual vs. number of TLS segments. The rigid domains were identified as Ile3-Lys172, Lys173-Leu364, and Val365-Phe414. Further discussion of TLS can be found in section 5.3.2.

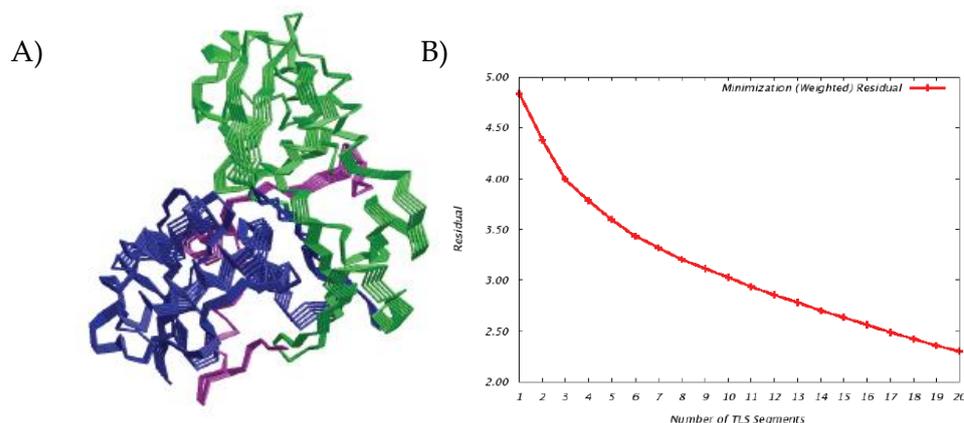


Figure 5-7 The result from the TLS analysis of DbLnP+PO₄.

A) A ribbon diagram showing the three domains used during refinement. B) Residual vs. Number of TLS segments.

The final refinement model had an R value and R_{free} of 0.20 and 0.28 respectively. The final model contains 408 of the 420 residues of thrombin-digested DbLnP. Residues that could not be resolved in COOT included five N-terminal residues, a loop turn Ala175-Asn177, and the last residue at the C- terminus, Ile414. The final model contains a number of side chains for which there was insufficient electron density to define their rotomers and the side chains were modelled in with occupancy of 0.

In the final model 3 phosphates and a manganese ion were modelled in the active site cleft (see section 5.2.10) and therefore this model hereon will be referred to as DbLnP+PO₄. There was no evidence in the electron density map to confidently support the placement of an AMP molecule in the active site.

The average B-factors for the N- and C-terminal domains were determined to be 17.9 Å² and 17.3 Å², respectively with the entire structure having an average B-factor of 17.7 Å².

Protein geometry was validated using Molprobity which showed that 100 % of the residues have allowable Ramachandran angles with 97.3 % of residues were found in the preferred region.

Refinement and model statistics for DbLnP+PO ₄	
Resolution range (Å)	39.02-2.57
R-factor (%)	17.7
R _{free} (%) [†]	24.8
Average B-factors (Å ²)	
Overall (No. of atoms)	17.7 (3292)
Overall protein (No. of atoms)	17.0 (3156)
N-terminus (No. of atoms)	17.9 (930)
C-terminus (No. of atoms)	17.3 (2214)
Solvent (No. of atoms)	15.5 (108)
PO ₄	26.9 (15)
Mn	17.9 (1)
R.M.S deviation	
Bond lengths (Å)	0.0178
Bond angles (°)	1.9
Ramachandren plot	
Percentage in favoured regions	97.3
Percentage in allowed regions	100
Percentage in disallowed regions	0

$$^{\dagger} R_{free} = \frac{\sum_{testset} \left| |F_{obs}| - |F_{calc}| \right|}{\sum_{testset} |F_{obs}|}$$

Table 5-2 Model and refinement statistics of the final model of DbLnP+PO₄.

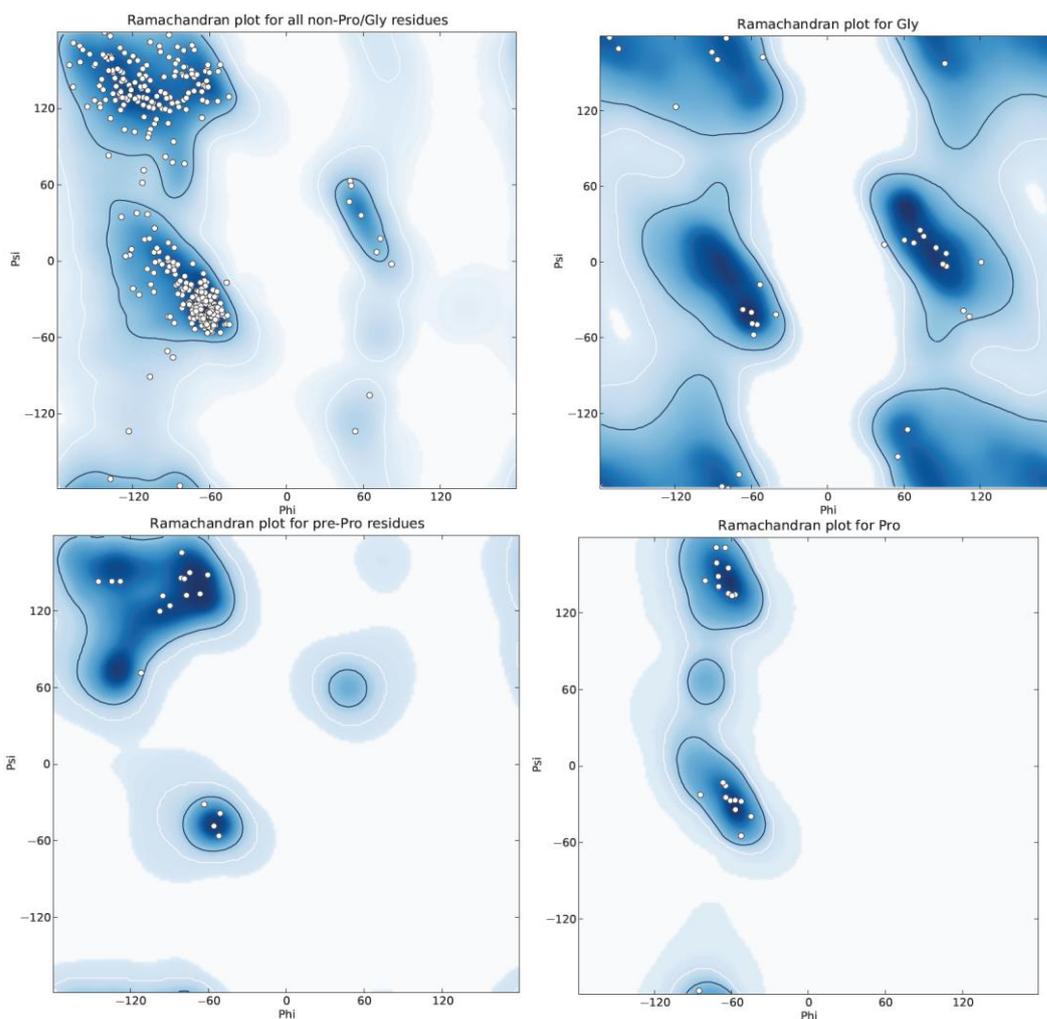


Figure 5-8 Ramachandran analysis of DbLnP + PO₄. Dark blue regions within the dark blue lines indicate preferred regions and the light blue regions within the white lines represent the acceptable ψ and φ angles.

5.2.8 The overall structure of DbLnP+PO₄

The crystal structure of DbLnP that contains 3 phosphates and a metal ion in the active site was determined at a resolution of 2.6 Å. The cartoon representation of DbLnP reveals two domains that are orientated to form the active site cleft. Structurally DbLnP is very similar to that of 7WC containing many similar features including; the N-terminal domain (Thr3 – Thr127) made up of a 5 mixed β -sheet to form the typical RNase-H domain and the C-terminal domain that resembles a modified RNase-H domain with the insertion of the α 8 helix and two β -strand hairpin loops (Figure 5-10). However, the C-terminal domain differs to that of 7WC because the C-terminal tail forms a small β -strand that associates

with the C-terminal domain β -sheet that extends the number of β -strands to 8 (Smith, et al., 1999b). This may be a consequence of DbLnP being in the closed conformation (section 5.4).

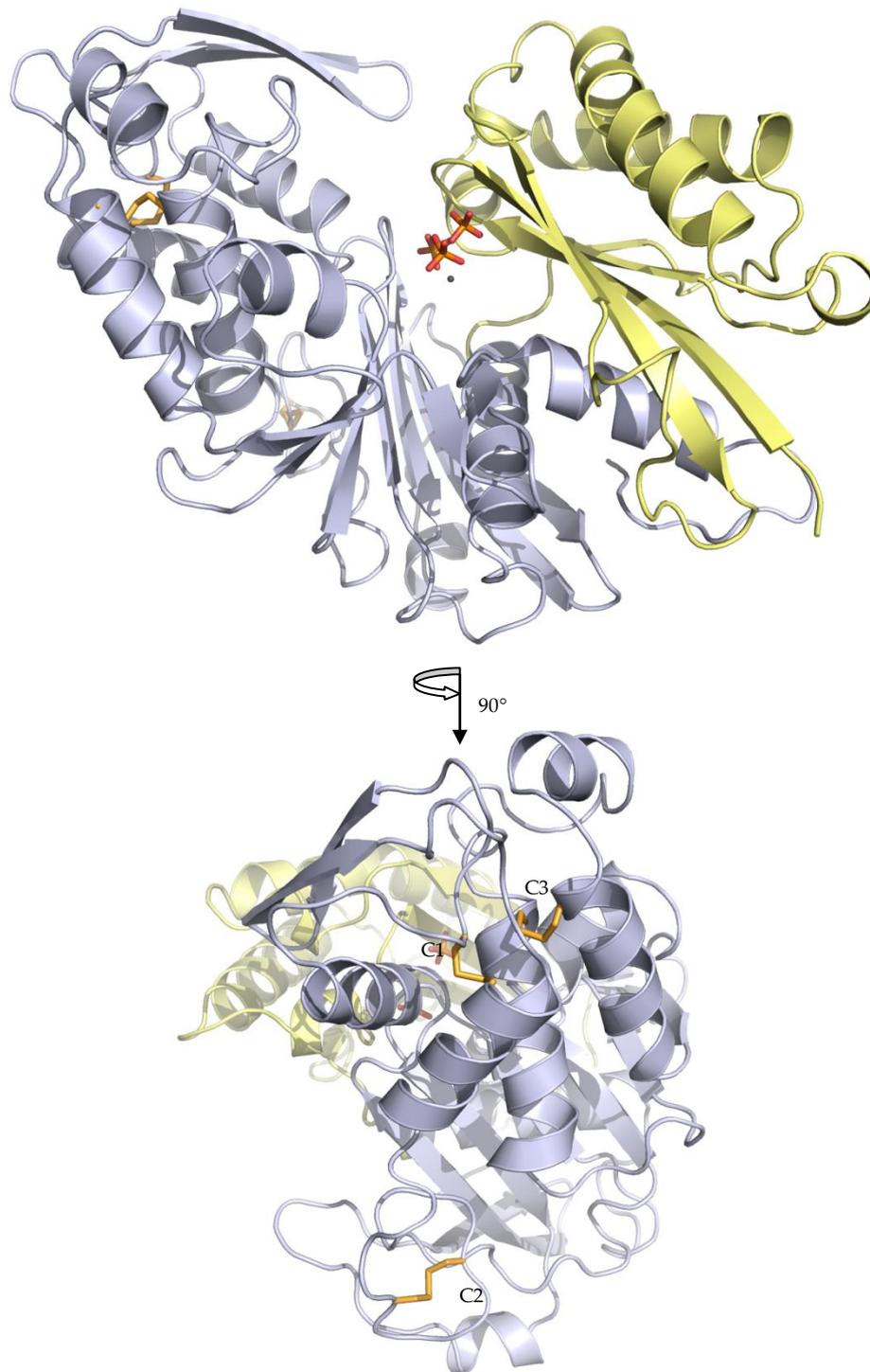


Figure 5-9 Cartoon representation of DbLnP+PO₄.

The C- and N-terminal domains are coloured grey and yellow, respectively. B) Disulfide bonds are shown in orange and labelled C1-3. Three phosphates (orange) and a metal ion (grey) are in the active site.

Similar to 7WC three disulfide bonds form within the C-terminal domain (Cys229-Cys259 (Figure 5-9-C1) and Cys331-Cys354 (Figure 5-9-C3)).

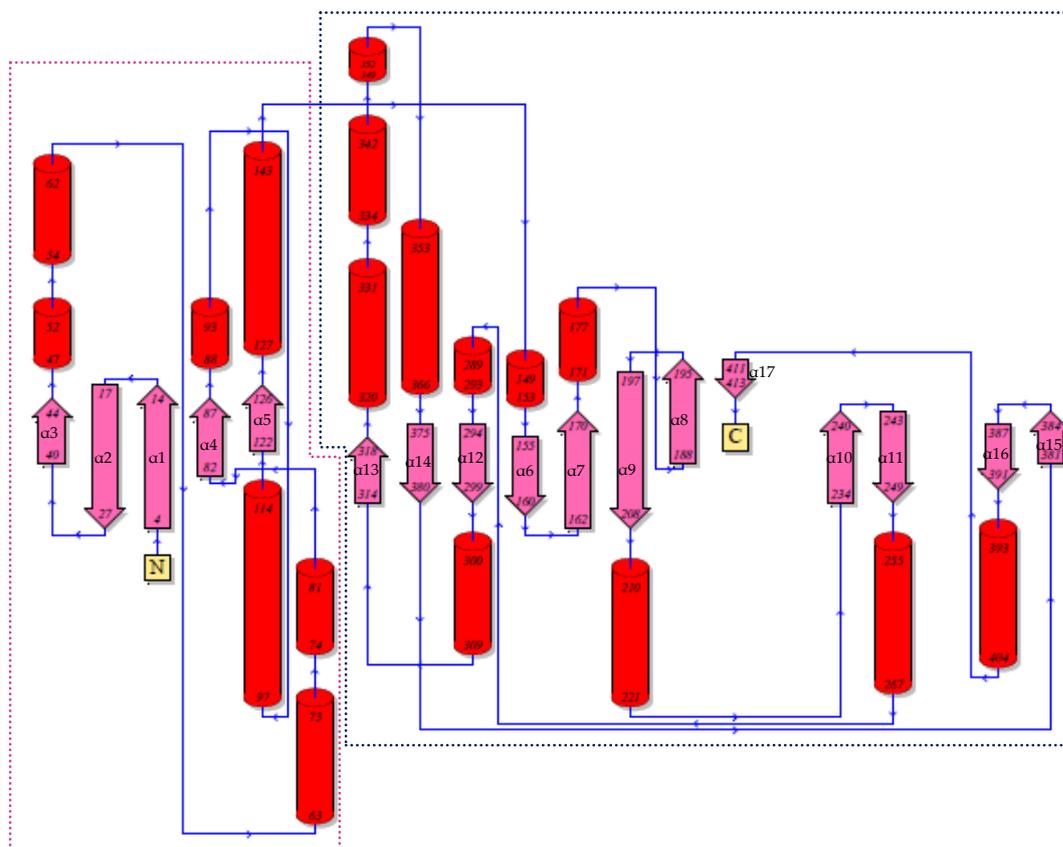


Figure 5-10 Topology model of DbLnP+PO₄. Pink dash lines represent the N-terminal domain. Dark blue dashed lines represent the C-terminal domain.

Despite the presence of AMP and MgCl₂ in the crystallisation solution there was no evidence to support the presence of AMP in the active site. However three phosphates were modelled in the active site (Figure 5-11). The NTPDases contain 2 phosphate binding sites (Vivian, et al., 2010; Zebisch, et al., 2008). The first consists of the N-terminal domain residues Ser15-Arg19, the second consists of the C-terminal domain residues Gly162-Ser163.

With reference to Figure 5-11, the phosphate 1 is bound only to the Arg19 of the first phosphate binding site and the Lys43. The phosphate 2 is located so that it interacts with residues in the phosphate binding site 1 including Ser15, Ser16 and the amide bond of Gly17 but not Arg19. In addition Lys43, part of the second β -strand in the N-terminal β -sheet, interacts directly with phosphate 2. This lysine

is highly conserved in the NTPDase family, including human NTPDase1, NTPDase3, *Glycine soja* GS52 and 7WC. Reports of the specific activities of NTPDases that contain lysine in this position all have an equal preference for NDP and NTP, whereas others, such as human NTPDase2, have non-conservative substitutions that exhibit significant differences between the specific activity of NDP and NTP hydrolysis (Kukulski, et al., 2005; Tanaka, et al., 2011b). This residue has yet to be a target for mutagenesis and may provide an insight into the discrimination between the hydrolysis of NDP and NTP. The third phosphate associates with the phosphate binding site 2 (Gly161, Gly162) and the Ser15 of phosphate binding site 1. This phosphate also contributes to the coordination of the metal ion (discussed later).

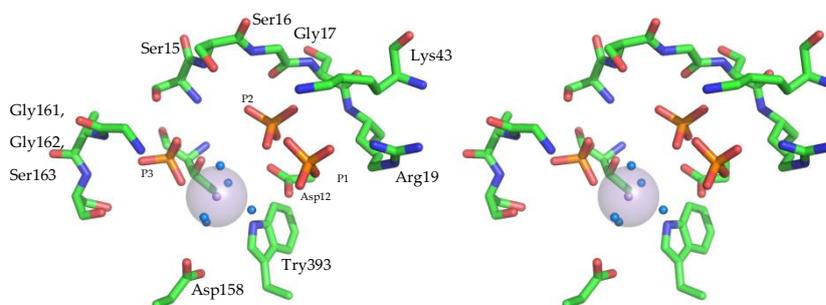


Figure 5-11 Stereo diagram of the active site of DbLnP+PO₄. Phosphates 1-3 are labelled P1-3. Waters and the metal ion are blue and grey, respectively.

5.2.9 Comparison of the DbLnP model with apo-7WC

The apo-7WC structure showed phosphates bound in the active site that corresponded with the α - and γ - phosphate of AMPPNP in the rat NTPDase2 structure. LSQ matching of the N-and C-terminal domains of 7WC with DbLnP was performed to compare the bound phosphates.

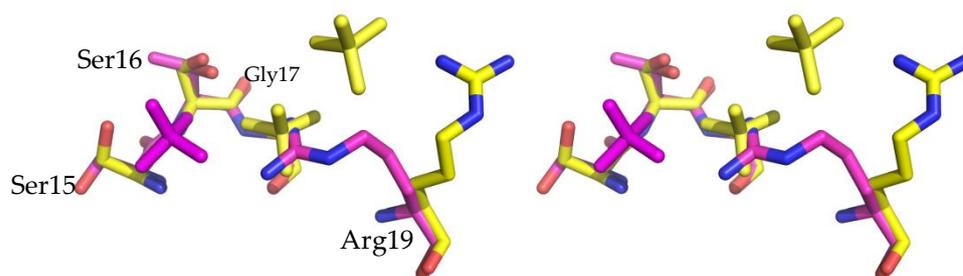


Figure 5-12 Stereo diagram of phosphate binding site 1 of apo-7WC (magenta) and DbLnP+PO₄ (yellow).

The phosphate binding site 1 of the apo-7WC structure (Figure 5-11-pink) shows the phosphate surrounded by the side chains of residues Ser15, Thr16, the backbone nitrogen of Gly17 and the guanidinium group on the side-chain of Arg19. The phosphates that are bound near the phosphate binding site 1 in the DbLnP structure do not correspond very well with the position of the phosphate in the 7WC structure. The main difference is that the side chain of Arg19 which in the 7WC structure curls towards the phosphate binding site to form a binding pocket is in a straighter conformation in DbLnP. As a consequence one phosphate has moved closer to the Gly17 while a second phosphate, not seen in the 7WC structure, associates with the Arg19. Overlay of the C-terminal phosphate binding site of apo-7WC and DbLnP reveals that both contain one phosphate that associates with the residues Gly161, Gly162 and Ser163. The orientation of the phosphates is slightly different due to the presence of a metal ion in DbLnP that is not present in the 7WC structure that would coordinate the position of the phosphate's oxygen.

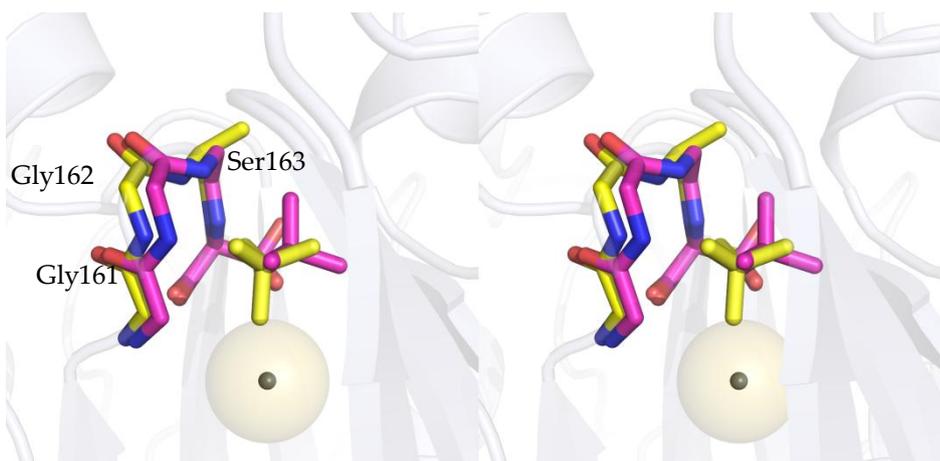


Figure 5-13 Stereo diagram of an overlay of the C-terminal phosphate binding site. The apo-7WC is shown in magenta, DbLnP+PO₄ in yellow and the DbLnP metal ion in grey.

5.2.10 Metal ion is present in the active site of DbLnP

There is evidence that a metal ion is present in the active site of DbLnP. X-ray fluorescence scanning was performed on regions of the crystal in and outside of the loop with a range of photon energies (from 1000-15000 eV). The results showed a fluorescence peak at 5900 eV and 6492 eV which is the expected L- and K-edge for manganese (Figure 5-14). This result was unexpected as manganese ions were not present in the crystallisation buffer (magnesium ions were present at a concentration of 200 mM). If magnesium was present one would have expected a fluorescence peak at 1253 eV and 1305 eV. It is possible that the manganese was a contaminant present in buffers used during refolding and purification which bound to DbLnP and did not dissociate once bound. A small peak at 8012 eV was detected which is the minimal energy required for copper to ionise. The detection of manganese or copper ions in the active site is plausible as other NTPDases have shown to utilise a range of divalent cations during catalysis (Chen, 2008; Kukulski, et al., 2003; Zebisch, et al., 2007).

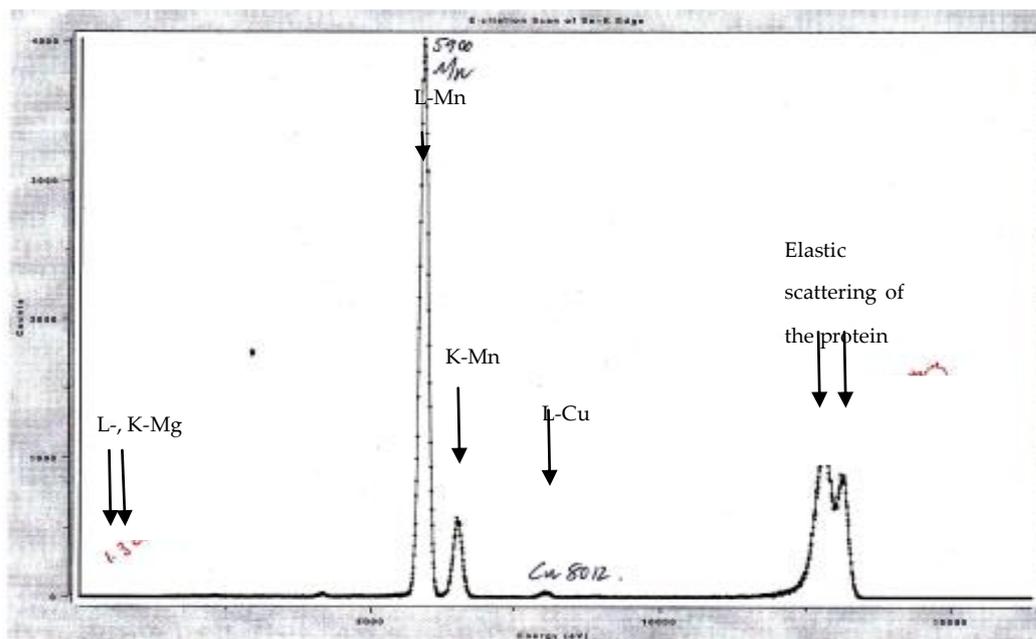


Figure 5-14 X-ray fluorescence scan of DbLnP+ PO₄ crystal outside of the loop. The L- and K-edge for various metal ions are indicated with arrows.

Single wavelength anomalous dispersion at the manganese absorption edge (6.5 keV) was attempted however due to technical difficulties these experiments did not advance.

The manganese ion is octahedrally coordinated with 5 waters and an oxygen from the phosphate that is bound in the C-terminal phosphate binding site. The distance of the metal coordinated oxygens for the manganese ranges between 1.8-2.3 Å, with a mean of 2.1 Å. These distances are within the range of O-Mn distances seen in the protein database which typically are between 2.1-2.2 Å (Harding, 2004). The six-coordinate manganese in the DbLnP structure agrees with the coordination number observed in the majority of manganese bound structures in the protein database (Harding, 2001).

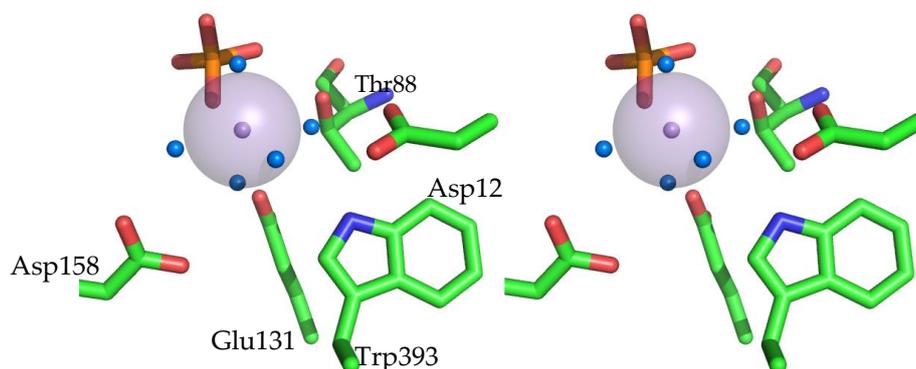


Figure 5-15 Stereo diagram of manganese coordination in the DbLnP+PO₄ structure. Manganese is represented as a solid grey sphere. The transparent grey sphere indicates 2.0 Å. Water is represented as blue spheres.

The waters that surround the manganese are coordinated by; Asp12, Asp158, Thr88, Trp393 and Glu131. These residues are highly conserved throughout members of the NTPDase family and point mutation of these residues leaves the enzyme inactive or with a lower affinity for metal ions (Drosopoulos, 2002; Drosopoulos, et al., 2000; Smith, et al., 1999a). One water is not coordinated by any residues but is flanked on either side by two phosphates.

The only other NTPDase structure published that contains a metal ion is the rat NTPDase2 with a calcium ion and AMPPNP (Zebisch, et al., 2008). The position of the metal ion in DbLnP is consistent with metal binding in the rat NTPDase2 where the divalent calcium is coordinated octahedrally with 4 water molecules and two oxygens from the AMPPNP. Like manganese, calcium is most likely to be octahedrally coordinated, but has larger contact distances with oxygens, typically between 2.35 and 2.45 Å (Harding, 2004; Hayward, 1999). Superposition of the rat NTPDase2 and DbLnP reveals that the same residues coordinate the waters surrounding the metal ion (Figure 5-16). The water that has no residue contacts is in the equivalent position as oxygen of the AMPPNP β-phosphoramidate.

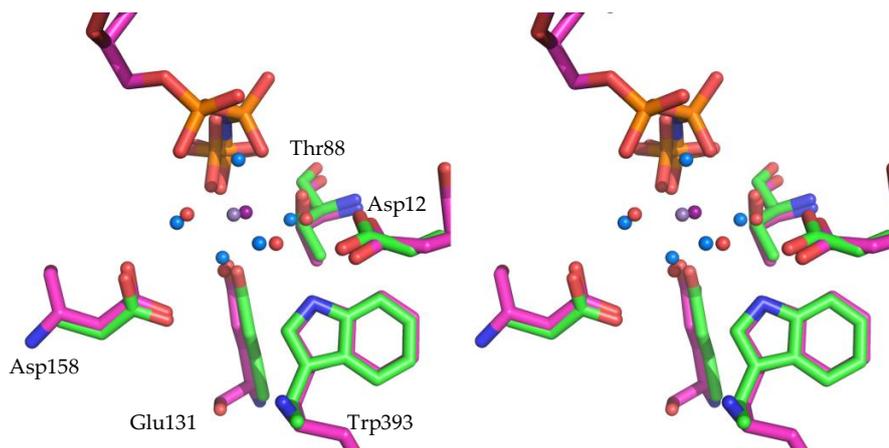


Figure 5-16 Stereo diagram of an overlay of DbLnP+PO₄ (green) and rat NTPDase2 with AMPPNP bound (pink).

DbLnP+PO₄ manganese and waters are shown in grey and blue, respectively. NTPDase calcium and waters are shown in purple and red, respectively.

5.2.11 Water is coordinated for nucleophilic attack

Close inspection of the catalytic site reveals the presence of a water that is coordinated by Ser163. The position of the water molecule is homologous to the water proposed to act as the nucleophile in the rat NTPDase2 model (Zebisch, et al., 2008).

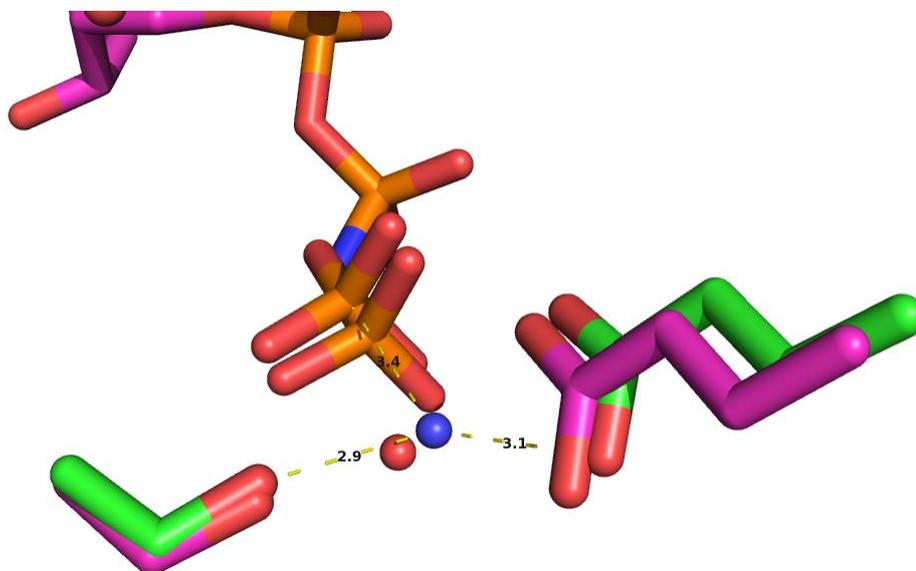


Figure 5-17 Overlay of DbLnP+PO₄ and rat NTPDase2 with AMPPNP bound that illustrates the position of the nucleophilic water.

Green residues and blue water represent DbLnP. Pink residues/AMPPNP and red water represent the rat NTPDase2 bound with AMPPNP. Distances (in Å) are shown for DbLnP.

The water is directly in-line with the phosphate molecule and is the best candidate for nucleophilic attack on the phosphate. The position of the water is 2.9 Å from the coordinating residue Ser163, 3.1 Å from the glutamate, which is thought to act as a general base and activates the water and 3.4 Å from the phosphorus atom. The angle between the posterior oxygen, the phosphorus and the water molecule is 173.5° which is comparable to the NTPDase2 model of 168.7°. The different metal ion was demonstrated to have a direct impact on the angle and distance of the water molecule from the substrate in *Dicystostelium* actin which may be a reason to explain these differences in the angles and distances between DbLnP and NTPDase2 (Vorobiev, et al., 2003).

5.2.12 Structural analysis of the DbLnP+PO₄ structure

Structural searches with DbLnP using DALI identified the published structures that had the highest structural homology (Holm, et al., 2000). In addition, analysis was carried out on the N- and C-terminal domains individually (Table 5-3).

DALI analysis of the entire structure identified the *Rattus norvegicus* NTPDase2 containing the highest structural similarity with DbLnP (Z-score of 40.9 and rmsd of 2.1 Å over 365 residues). The N-terminus has the highest structural homology with the rat NTPDase2 with a Z-score of 16.2 and rmsd of 1.9 Å over 118 residues. Analysis of the C-terminal domain exhibited a lower structural homology than the N-terminal domain with the most structurally similar protein, the rat NTPDase2 with a Z-score of 24.5 and rmsd of 2.2 Å over 249 residues. The lower structural homology in the C-terminal domain is not surprising as the C-terminal domain have a number of residue extensions that differ between structures.

The next closest structural homologue is the *Legionella pneumophila* NTPDase1 (LpNTPDase1) with the entire protein having a Z score of 34.6 and an rmsd of 2.9 over 335 residues, the N-terminal domain having a Z score of 12.6 and an rmsd of 2.3 over 110 residues and the C-terminal domain having a Z score of 21.8 and an rmsd of 2.6 over 232 residues.

Whole DbLnP	PDB code	Z-score	rmsd	Residue #	Type	Species	Reference
1	3cja	40.9	2.1	365	NTPDase	<i>R. novogicus</i>	(Zebisch, et al., 2008)
2	3aaq	36.6	2.9	335	NTPDase	<i>L. pneumophila</i>	(Vivian, et al., 2010)
3	3cer	23	3.6	295	PPX/GPP A	<i>B. longum</i>	Unpublished
4	3agr	22.2	3.4	360	NTPase	<i>N. caninum</i>	Unpublished

N-terminal domain	PDB code	Z-score	rmsd	Residue #	Type	Species	Reference
1	3cja	16.5	1.9	118	NTPDase	<i>R. novogicus</i>	(Zebisch, et al., 2008)
2	3aaq	12.6	2.3	110	NTPDase	<i>L. pneumophila</i>	(Vivian, et al., 2010)
3	3agr	10.2	2.3	118	NTPase	<i>N. caninum</i>	Unpublished
4	2j4r	9.4	2.4	97	PPX/GPP A	<i>A. aeolicus</i>	(Kristensen, et al., 2008)

C-terminal domain	PDB code	Z-score	rmsd	Residue #	Type	Species	Reference
1	3cja	24.5	2.2	249	NTPDase	<i>R. novogicus</i>	(Zebisch, et al., 2008)
2	3aaq	21.8	2.6	232	NTPDase	<i>L. pneumophila</i>	(Vivian, et al., 2010)
3	3agr	17.4	3.1	245	NTPase	<i>N. caninum</i>	Unpublished
4	2j4r	13.2	3.2	180	PPX/GPP A	<i>A. aeolicus</i>	(Kristensen, et al., 2008)

Table 5-3 DALI-Lite structural alignment.

Displayed are the most structurally similar homologs for entire and the N- and C-terminal domains of DbLnP+PO₄. Z-score is a DALI output that evaluates similarity where values above 2 is considered to have significant similarities.

Others structures that emerged as being structurally similar were the NTPase from *Neospora caninum* (3AGR) and the exopolyphosphatase from *Bifidobacterium longum* (3CER) (both structures are unpublished).

All identified structural homologous are members of the Actin/Hsp70 superfamily and the actin family. The two closest structural homologues, 3CJ7 and 3AAR contain two PFAM GDA_CD39 domains (PF01150). The lesser

homologues 3GAR and 2J4R contain formyl_trans_N domain (PF00551) and the PPx-GPPA domain (PF02541), respectively.

Pair-wise structural alignment of the apo-7WC structure (section 4.6) and the DbLnP model using DALI-pairwise alignment revealed an rmsd of 3.0 using 406 residues for the entire structure. The structural similarity increased significantly when each domain was compared individually. The N-terminal domain had an rmsd of 0.9 using 124 residues and the C-terminal domain exhibiting an rmsd of 1.5 using 283 residues. Such high structural similarity between 7WC and DbLnP was expected as the sequence identity is very high. Also, the length of the protein is very similar with 7WC slightly longer with two insertions (a proline and an asparagine between Thr312 and Pro313 of DbLnP and valine between Ile352 and Pro353 of DbLnP) in the C-terminal domain.

5.3 DbLnP co-crystallised with AMPPNP

5.3.1 Molecular replacement of DbLnP co-crystallised with AMPPNP

Molecular replacement was undertaken using the DbLnP+PO₄ as a molecular replacement model in PHASER from the CCP4 suite of programs (McCoy, et al., 2007). It was possible to solve the completed structure using the entire model. The results from the molecular replacement resulted in an initial model with R and R_{free} values of 29.5 and 34.2 respectively, containing 386 of the 420 amino acids. Refinement statistics are shown in Table 5-4.

Refinement and model statistics for AMPPNP+DbLnP	
Resolution range (Å)	46.7-2.9
R-factor (%)	24.3
R _{free} (%) [†]	30.1
Average B-factors (Å ²)	
Overall (No. of atoms)	30.3 (3292)
Overall protein (No. of atoms)	30.4 (3148)
N-terminus (No. of atoms)	31.2 (942)
C-terminus (No. of atoms)	23.9 (2206)
Solvent (No. of atoms)	33.2 (108)
PO ₄	11.2 (5)
ANPPN*	20.0 (27)
Mn	14.3 (1)
SO ₄	53.4 (5)
R.M.S deviation	
Bond lengths (Å)	0.0132
Bond angles (°)	1.6
Ramachandran plot	
Percentage in favoured regions	94.1
Percentage in allowed regions	99.3
Percentage in disallowed regions	0.7

$$^{\dagger} R_{free} = \frac{\sum_{testset} ||F_{obs}| - |F_{calc}||}{\sum_{testset} |F_{obs}|}$$

Table 5-4 Model and refinement statistics of the final model of DbLnP co-crystallised with AMPPNP.

* See section 5.4.3 for details.

5.3.2 Model refinement of the DbLnP co-crystallised with AMPPNP

The initial model was improved by manual building in COOT directed by the $2|F_o| - |F_c|$ and $|F_o| - |F_c|$ maps contoured to 1σ and 3σ respectively and verified using either REFMAC 5 (CCP4) or phenix.refine. Further iterations were performed using TLS refinement. The rigid domains were determined using TLSMD server (<http://skuld.bmsc.washington.edu/~tlsmd>) which showed no clear rigid domains by identifying a sharp angle in the residual vs. number of TLS segments (Figure 5-18). Despite this, fixing the two domains, Leu3-Lys173 and Ala174-Phe414, results in a small improvement during refinement using TLS

refinement in Refmac 5.0. These results, with the low overall B-factor, indicate that the protein has little propensity for movement.

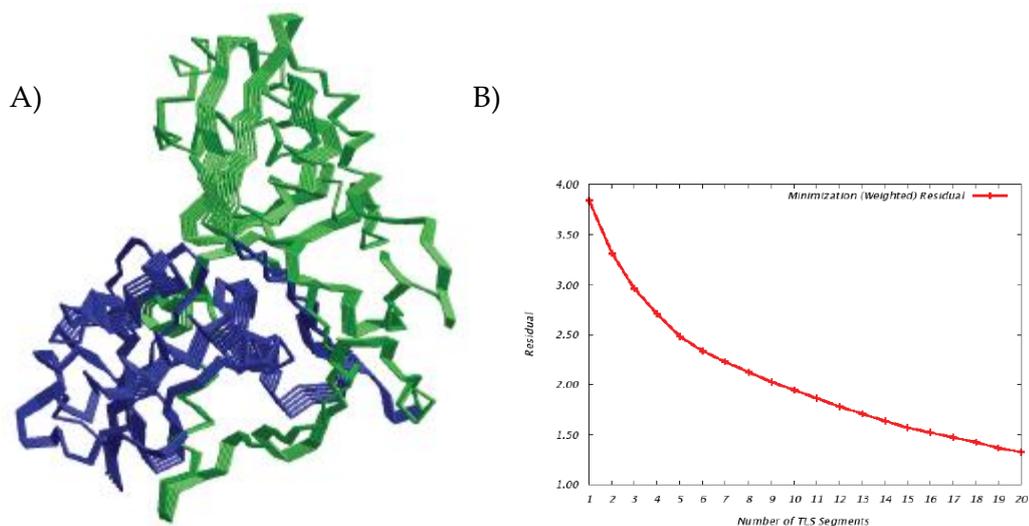


Figure 5-18 The results from the TLS analysis of DbLnP+AMPPNP. A) A ribbon diagram showing the two domains used during refinement. B) Residual vs. number of TLS segments.

The atomic displacement can be visualised by representing each atom as a thermal ellipsoid. These are derived from the anisotropic TLS using the most likely rigid domains. Observation of the thermal ellipsoid figures of the phosphate and AMPPNP bound DbLnP reveals that the average displacement is very small, shown as small spherical ellipsoids for both (Figure 5-19). In contrast, the apo-7WC structure which showed inherent flexibility (determined by two defined anisotropic domains and high B-factors) has large elongated ellipsoids that indicate more potential for a larger atom displacement. These results suggest that this protein structure is in a locked conformation.

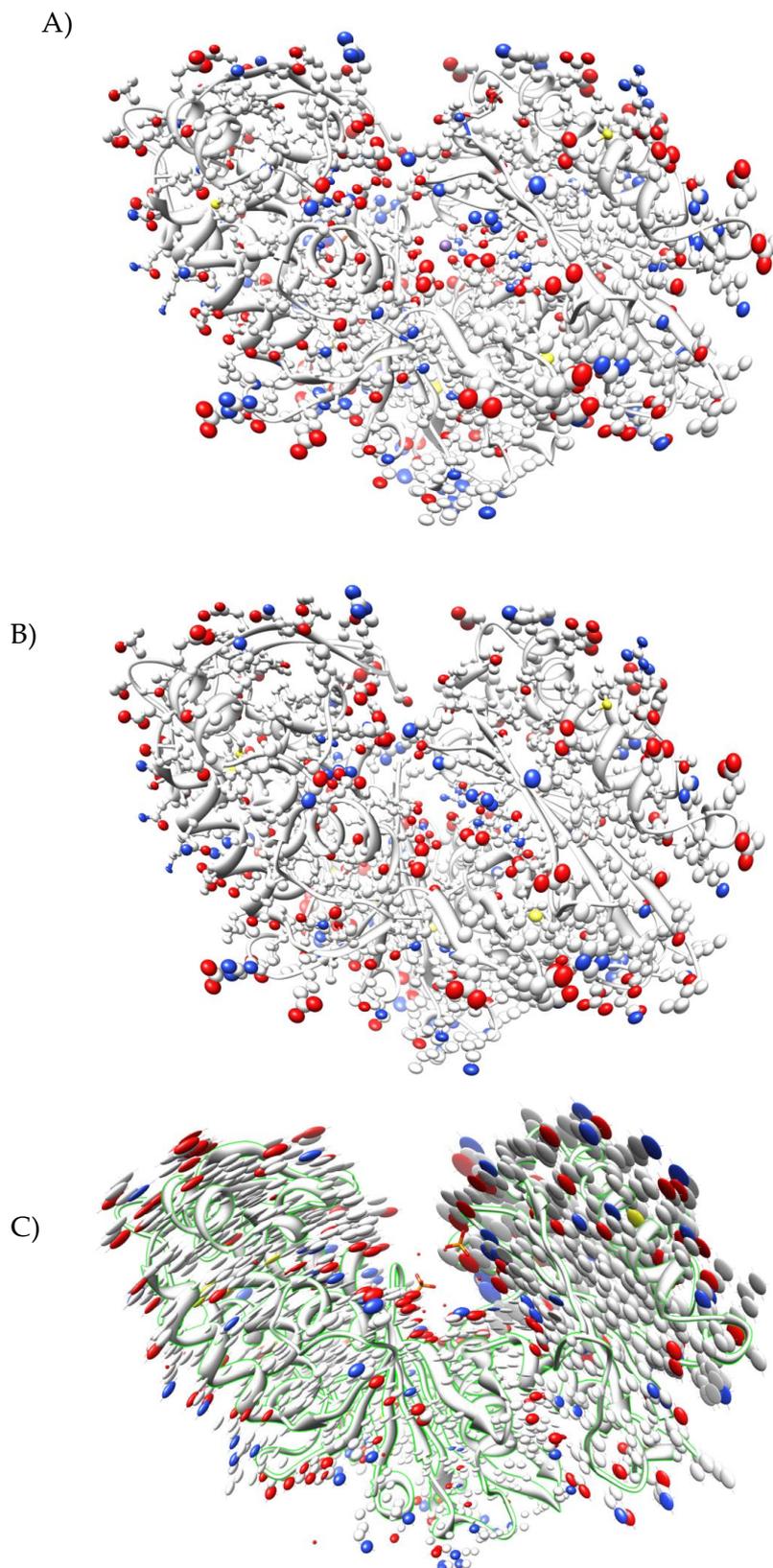


Figure 5-19 Thermal ellipsoid diagrams of NTPDases. Thermal ellipsoid diagrams of NTPDases of the A) DbLnP+PO₄, of B) DbLnP co-crystallised with AMPPNP and of C) apo-7WC.

Comparison of these results with the phosphate bound DbLnP which similarly showed very little displacement of atoms suggests and either a substrate or a metal ion is essential for forming this 'locked' conformation. This hypothesis is further supported by comparing the apo-7WC and the AMP bound 7WC results with the DbLnP structure. The TLS analysis of the 7WC structures indicated two rigid domains with substantial displacement of atoms (section 4.5). Comparing the thermal ellipsoids of the AMPPNP co-crystallised DbLnP and the apo-7WC shows a marked difference in the thermal displacement particularly in the lobes of each domain. Unlike the DbLnP structures, 7WC structures do not contain a metal ion, a phosphate ion or an AMP bound in the active site. This implies that both the substrate and/or phosphate ions and a metal ion need to be present for the enzyme to adopt a rigid form.

5.3.3 Structural analysis of the two DbLnP structures

Performing structural similarity matching and DALI pair-wise alignment of the DbLnP+PO₄ and AMPPNP co-crystallised DbLnP structures show no significant difference in the overall structure with an rmsd of 0.3 Å using 411 residues.



Figure 5-20 Overlay of DbLnP structures co-crystallised in AMP (green) and AMPPNP (cyan).

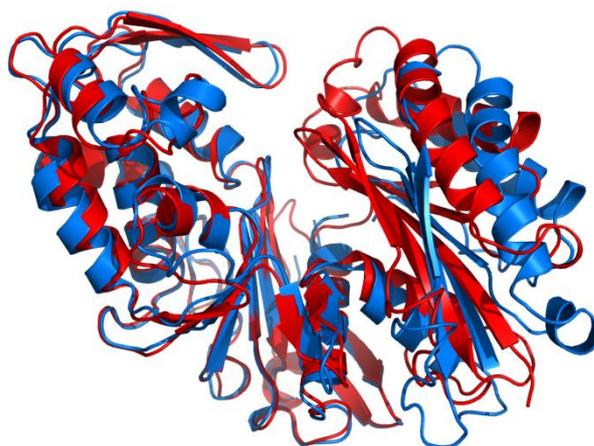


Figure 5-21 SSM overlay of apo-7WC (blue) and AMPPNP co-crystallised DbLnP (red). The C-terminal of each structure has been fixed in the same position.

5.4 Substrate induced trans-domain movement of 7WC and DbLnP

Upon the binding of substrate to members of the actin superfamily a conformational change has been observed which brings the two domains closer to one another reducing the width of the active site cleft described as a butterfly-motion. Such a conformational change is essential for catalysis (Kristensen, et al., 2004; Kristensen, et al., 2008; Schuler, 2001). Due to the homology of NTPDases

with members of the actin superfamily (Bork, et al., 1992; Smith, et al., 1999a) it has been proposed that a conformational change occurs upon the binding of the substrate (Vivian, et al., 2010; Zebisch, et al., 2008). However, conformational changes in NTPDases have yet to be demonstrated and thus the full detail of the catalytic mechanism has yet to be shown. This section shows comparisons of the substrate bound and apo-form of 7WC and DbLnP that shows the movement of domains, identifies hinge regions and gives a better picture of the catalytic mechanism.

5.4.1 Hinge domain analysis using apo-7WC and 7WC with AMP bound.

Differences were observed between the orientation of the domains between the apo- and AMP bound 7WC (section 4.12). To investigate the substrate binding domain movement, the apo- and-AMP bound 7WC were analysed using the DynDom web server (<http://ekhidna.biocenter.helsinki.fi/frontpage>)(Hayward, et al., 1998) (Figure 5-22). This revealed that the conformational change was due to the movement of two rigid domains, the N-terminal domain (Ile4-Ser124) and the C-terminal domain (Ile125-Glu412). Dyndom analysis revealed a rotation angle between the two domains of 11.8° centred on the putative hinge region, Ser124-Ile126 located between the two domains (Figure 5-22-HR1). The N-terminal domain contains three other regions that have minor hinge contributions including one that is located in the β -strand (Arg19-Val22) which positions residues that associate with the α -phosphate and place AMP closer to the C-terminal domain (Figure 5-22-HR2). If this movement did not occur AMP would be too small to bridge the two domains and binding would be poorer as a result.

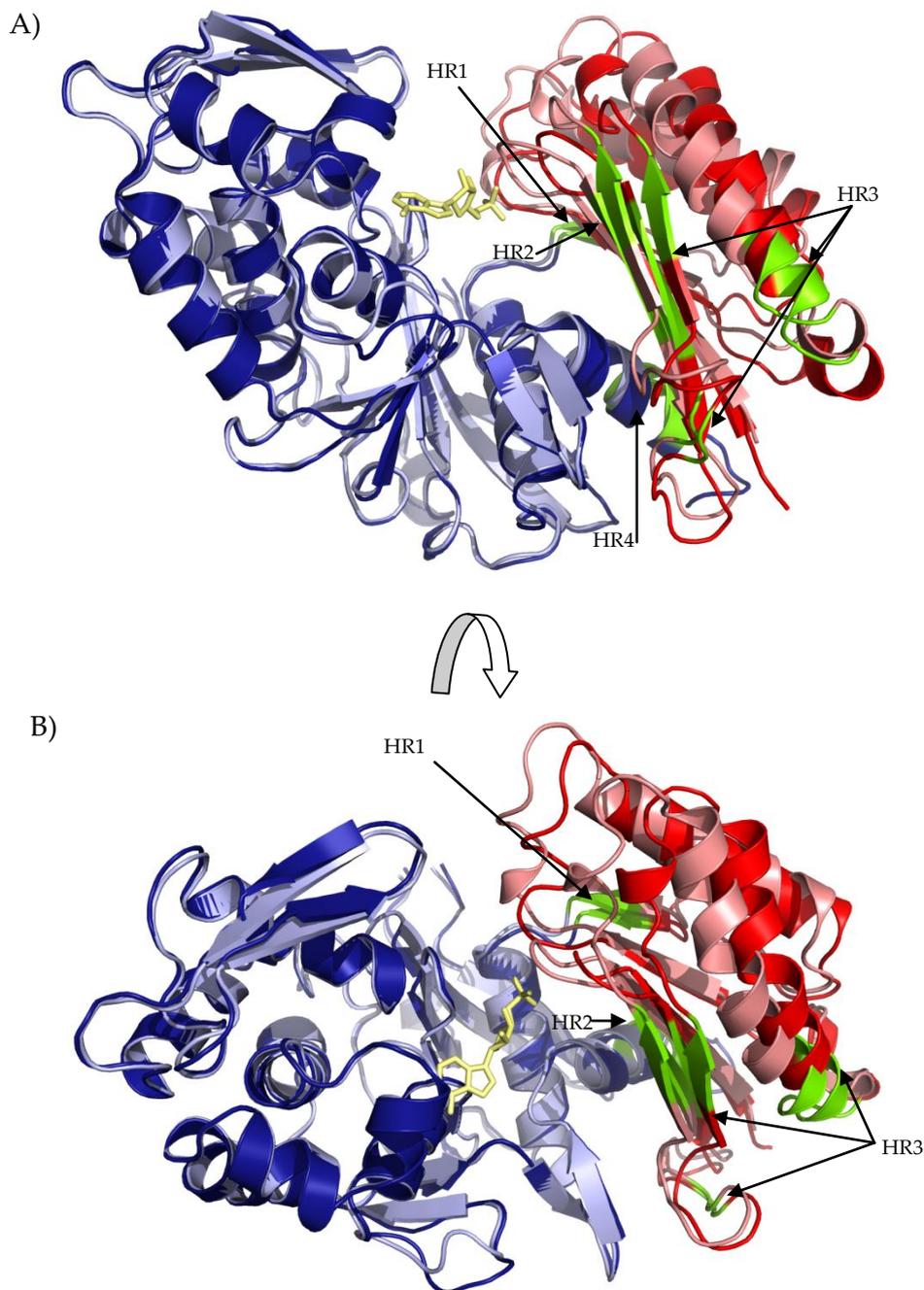


Figure 5-22 Visualisation of domain movement and hinge domains of 7WC upon AMP binding.

The C-terminal domain of each structure was fixed to demonstrate the movement of the N-terminus. Dark and light blue represent the C-terminal domain of Apo and AMP bound 7WC, respectively. Red and pink represent the N-terminal domain of the apo and AMP bound 7WC. The green regions represent the hinge domains determined by DynDom. The arrows indicate the main hinge domains B) is the structure rotated 90 ° about the x axis.

Also, with the hinge regions Leu30-His32, Lys41-Ile43 and Glu69-Pro73 (Figure 5-22-HR3) a relatively large domain movement ($\sim 5 \text{ \AA}$) between Phe40 and Glu75

occurs. In addition to the N-terminal hinge region one C-terminal hinge region was identified at Glu403-Val404 in the ARC5 helix (Figure 5-22-HR4). This is likely to be a consequence of the intrinsic flexibility of terminal residues rather than a consequence of AMP binding.

The closest structural homologue that has shown domain movement is the *A. aeolicus* PPX/GPPA enzyme (section 4.7). This work identified an 11.8° rotation about a flexible region that is present between Arg121-Tyr124, nine amino acids up-stream of the main hinge region observed for 7WC (Kristensen, et al., 2004; Kristensen, et al., 2008). The hinge region predictions for both structures are consistent with earlier work on actin domain movement (Schuler, 2001). Such similarity between these structures suggests a conservation of the domain movement mechanism between members of the actin superfamily.

5.4.2 Hinge domain analysis using apo-7WC and DbLnP co-crystallised with AMPPNP

Due to the high sequence identity (65 %) and structural homology between apo-7WC and the DbLnP structures (section 5.2.12) it was possible to perform a hinge domain analysis using the Dyndom server (<http://fizz.cmp.uea.ac.uk/dyndom/>) (Hayward, et al., 2002) using the AMPPNP co-crystallised DbLnP structure and the apo-7WC structure to determine the hinge region.

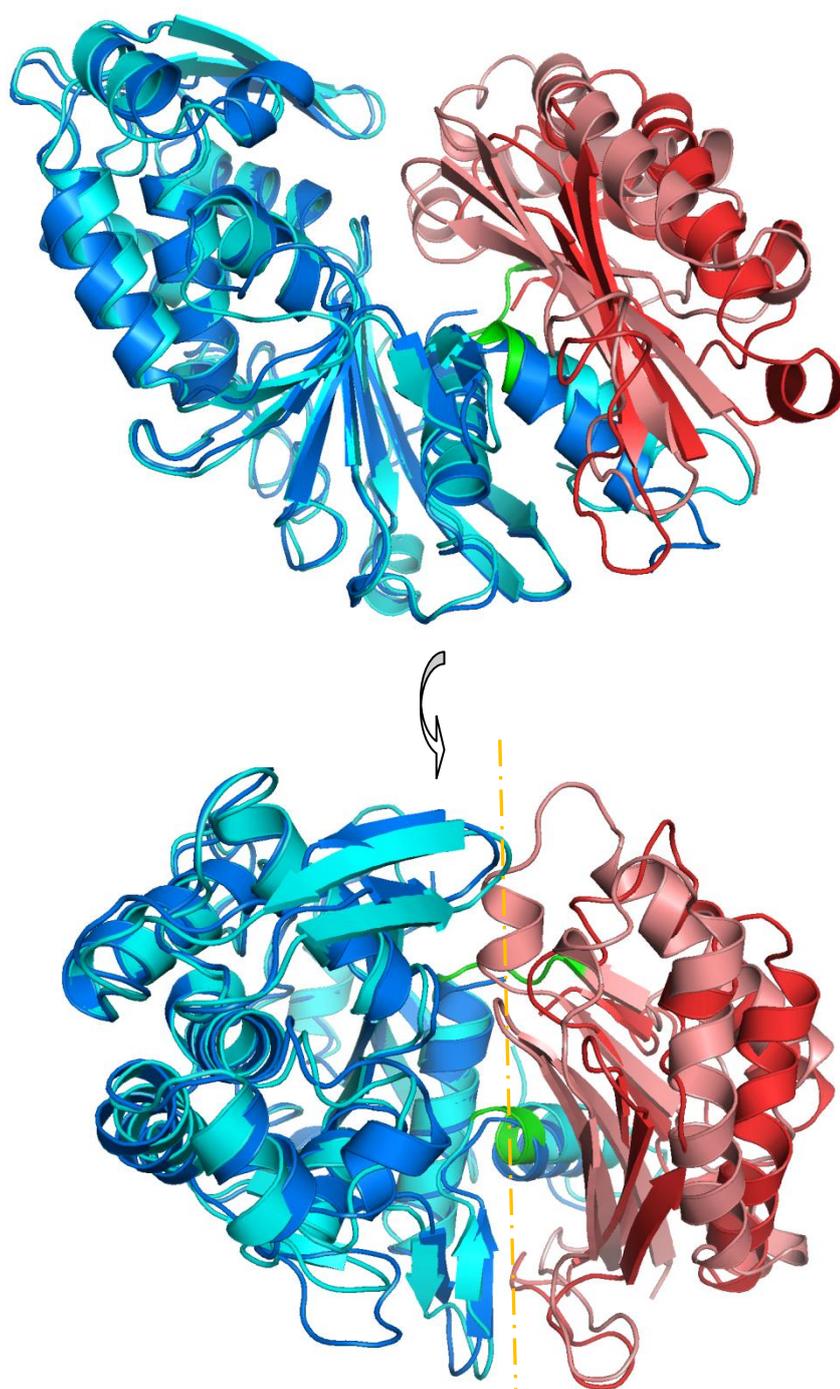


Figure 5-23 Visualisation of the hinge region using apo-7WC and AMPPNP co-crystallised with DbLnP.

The C-terminal domain of each structure was fixed to demonstrate differences in the inter-domain angle. Dark and light blue represent the C-terminal domain of the apo-7WC and DbLnP, respectively. The red and pink represent the N-terminal domain of the apo-7WC and DbLnP. The green regions represent the hinge domains determined by DynDom. Model B) is the structure rotated 90° about the x-axis. The yellow dashed line indicates the mirror axis.

Observation of the Dyndom analysis reveals two hinge regions (Figure 5-23). The first is 3 residues centred on Ile127 of DbLnP and apo-7WC. This result is consistent with the DynDom analysis of the apo-7WC and the AMP bound 7WC (section 5.4.1) and that identified in the *A. aeolicus* PPX/GPPA enzyme (Kristensen, et al., 2004; Kristensen, et al., 2008). The second hinge contains three residues centred on Trp394, in DbLnP and Trp397 in 7WC, respectively. This hinge is part of the ACR5 and contains the highly conserved tryptophan that is involved in coordinating the metal associated waters and, that when mutated to an alanine, alters the catalytic mechanism in human NTPDase2 (Smith, et al., 1999a). The two hinge domains create a plane of symmetry between the C-terminal domain and the N-terminal domain and the C-terminal α -helix that associates with the β -sheet of the N-terminal domain. The angle of rotation between that of the N- and C-terminal domains is 22.6° which is consistent with the conformational change in *A. aeolicus* PPX/GPPA enzyme reported as 22.5° (Kristensen, et al., 2008).

Taking the Dyndom and the TLS analysis results, the DbLnP structures with substrate and metal bound represents a “closed” conformation that has little dynamic movement. The apo-7WC structure that contains no metal bound represents the “open” form. The transition from the open to closed state requires the metal ion to associate with the enzyme since no movement was observed in the AMPPNP 7WC structure that did not contain a metal ion. Further, the closed conformation of DbLnP contains a metal ion and either substrate or phosphate and was sufficient to induce the domain movement. Since one of the identified hinge regions contains a metal associating residue one could speculate that metal ion induced conformational change occurs through the interaction between the Trp394 located in the ACR5 and the metal ion. This phenomenon was observed with another member of the actin superfamily, G-actin that only if substrate and metal ion bound, did conformational changes occur (Frieden, et al., 1988; Schuler, 2001).

In comparison to the domain movement between the apo-7WC and the 7WC bound with AMP (which showed a rotation angle of $\sim 12^\circ$) this analysis between

the apo-7WC and the AMPPNP bound DbLnP is likely to represent the conformational change of the NTPDases. By comparing the crystal structure of the co-crystallised DbLnP the movement is not retarded by crystal packing constraints.

5.4.3 The active site of the AMPPNP co-crystallised DbLnP.

The non-hydrolysable substrate, AMPPNP was unable to be successfully modelled into the active site, rather a de-phosphorylated adenosine 5' [β -imido] diphosphate (AMPPN) and a phosphate molecule could be successfully modelled using COOT directed by the $2|F_o|-|F_c|$ and $|F_o|-|F_c|$ maps contoured to 1σ and 3σ , respectively (Figure 5-24). Therefore the structure will be referred to as DbLnP+AMPPN for hereon. In addition, there was density for a manganese ion. This may represent a hydrolysis product bound within the active site. If this is the case it would provide information as to the binding mechanism immediately after hydrolysis of ATP.

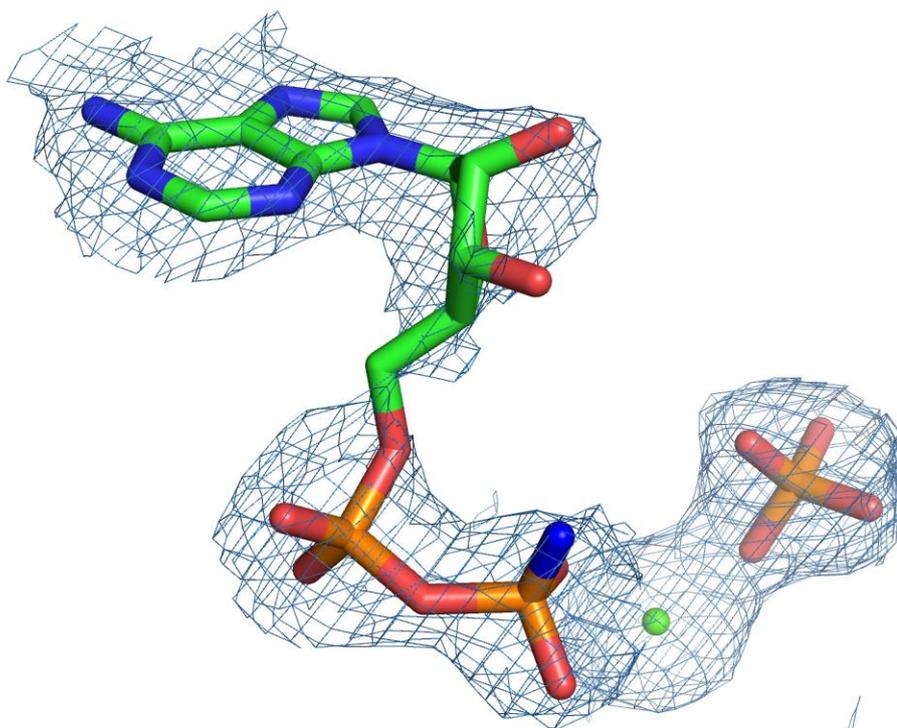


Figure 5-24 AMPPN and phosphate ion conformation in the active site of DbLnP. The green sphere represents the manganese. Blue mesh represents the $2|F_o|-|F_c|$ map contoured to 1σ .

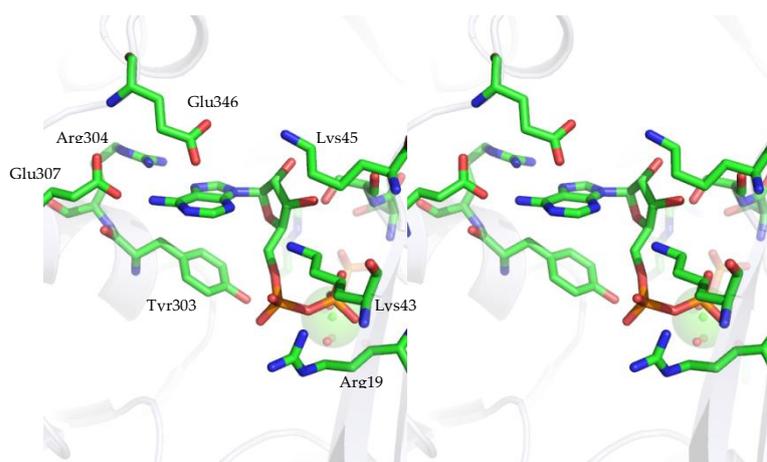


Figure 5-25 Stereo diagram of the nucleobase binding site within the active site of DbLnP+AMPPN.

5.4.3.1 Interactions that AMPPN makes with DbLnP side-chains

The adenosine base of AMPPN was located directly above the aromatic side-chain of Tyr303 (4-4.5 Å). The base is in an orientation that is not parallel with the Tyr303 rather, it is slightly twisted which may weaken the π -stacking associations between the base and the aromatic side-chain. It is likely that this is due to an association between the N1 on the adenine and the side-chain oxygen of Glu346 which is tilted towards the base 3.1 Å away. In contrast to 7WC, which has a Tyr303 and Phe350 that associates with the base, only one residue is likely to associate through π - π stacking interactions with the base in DbLnP. This feature is unique when compared with other published structures of NTPDases which show two hydrophobic residues in each case (Tanaka, et al., 2011b; Vivian, et al., 2010; Zebisch, et al., 2008), although a single residue binding occurs in the distantly related *Saccharomyces cerevisiae* actin (Vorobiev, et al., 2003). The adenosine base also interacts through the amine group on the C6 carbon and Glu307. In contrast to 7WC which has an Asp307 substitution, the Glu307 creates a smaller nucleobase binding pocket which would account for DbLnP having a higher k_{cat}/K_M for the smaller pyrimidines. Adjacent to the Glu307 is Arg304 that is orientated in a fashion that if the base contained negatively charged side groups, such as guanine, uracil and thymine, these could associate with the

Arg304 without altering the substrates orientation significantly. This is supported by the kinetic data which showed that DbLnP had no specificity for bases that contain negative or positive side groups (section 6.3.3).

The ribose of AMPPN exists in a C3'-endo conformation where the hydroxyl groups of ribose associate with DbLnP through a number of hydrogen bonds to side chains of Lys43, Lys45 and Ser16. The orientation of the hydroxyl groups could not be clearly defined in the electron density map so the following distances to these side chains are for hydroxyls toward and away from the base, respectively. The Lys43 is 2.6/3.1 Å from the C3 hydroxyl and Lys45 is 2.9/3.6 Å and 3.3/4.1 Å from the C2 and C3 hydroxyl of the ribose, respectively. The Ser16 is 4.5/3.1 Å and 3.8/2.6 Å from the C2 and C3 hydroxyl of the ribose, respectively. The α -phosphate of AMPPN is able to interact with DbLnP through Arg19 and Lys43, 3.5 Å and 2.9 Å from the associated oxygen. Comparing the position of Arg19 and the Lys43 between the AMPPN bound in 7WC and DbLnP showed that in the open conformation of 7WC the Arg19 side-chain orientates to associate with the γ -phosphate. However in the closed conformation DbLnP the Arg19 side-chain is rotated $\sim 90^\circ$ to associate with the α -phosphate of AMPPN.

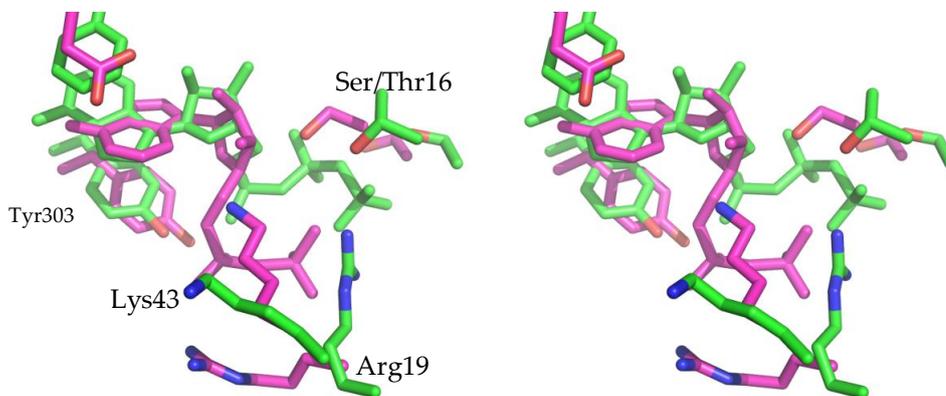


Figure 5-26 Stereo diagram of an overlay of 7WC and DbLnP + AMPPN. Green and pink residues/substrates represent 7WC and DbLnP respectively.

Also, Lys43 which has shown to associate with both the α -phosphate and the ribose in DbLnP is orientated in a way that does not contribute to the active site in 7WC. Lysine or arginine are highly conserved within a large number of NTPDases (such as mammalian NTPDase1-3 and *Glycine soya* GS52) and has

been shown in *L. pneumonia* to participate in substrate binding (Vivian, et al., 2010).

The β -phosphoramidate of AMPPN associates only through interactions between the nitrogen and the oxygen of Gly17. Previous reports of NTPDase structures all show either a phosphate or a substrate phosphate making a myriad of associations with the phosphate binding site residues (Vivian, et al., 2010; Zebisch, et al., 2008). In DbLnP this includes Ser15, Ser16 and Arg19. It is plausible that the substitution of the oxygen to nitrogen on the β -phosphate may give rise to these unexpected results. However, not all the oxygens are coordinated by the protein enabling the β -phosphoramidate to rotate freely about the phosphoanhydride bond.

In apo-7WC a free phosphate binds in this position and the α -phosphate of AMP bound to 7WC (section 4.6) and *L. pneumonia* NTPDase1 bound AMPPNP structure showed the γ -phosphate in this site (Vivian, et al., 2010). The rat NTPDase2 AMPPNP structure showed the β -phosphoramidate in the phosphate pocket (Zebisch, et al., 2008). The β -phosphoramidate is between 5.2 Å and 7.7 Å from the side-groups of these residues which is too far to associate. So it seems peculiar that a substrate phosphate, particularly the β -phosphoramidate does not associate with this phosphate binding pocket. The displacement of the phosphates is most likely due to the orientation of the Arg19 which is moved away from the Ser15 and Ser16 demonstrating the importance of Arg19 in the phosphate binding site.

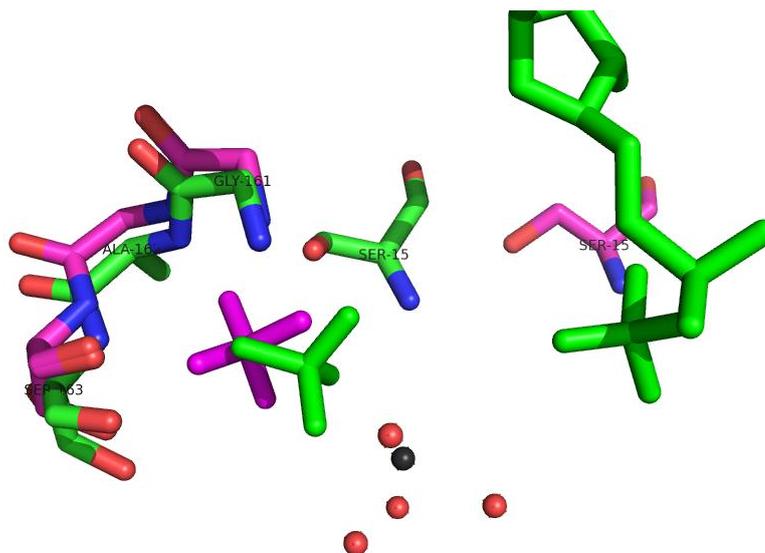


Figure 5-28 Overlay of the free phosphate of the apo-7WC and DbLnP+AMPPN structures.

7WC and DbLnP are shown in magenta and green, respectively. The black and red spheres represent the waters and metal ion present in the DbLnP structure.

5.4.3.3 *Hydrolysis of the AMPPNP in the DbLnP crystal*

The presence of the de-phosphorylated AMPPNP or adenosine 5' [β -imido] diphosphate, which will be termed AMPPN, was unexpected. The AMPPN could be a product of hydrolysis however in theory this should not occur. However there are many examples of ATP hydrolysing enzymes having the ability to hydrolyse the β - γ imido bond of AMPPNP. For example snake venom phosphodiesterase and the sarcoplasmic reticulum ATPase both showed the ability to hydrolyse AMPPNP albeit at a lower rate than the native substrate ATP (Taylor, 1981; Yount, et al., 1971). There was no evidence that DbLnP cleaved AMPPNP in the activity assays however the liberation of phosphate was measured after 2-3 minutes so slow hydrolysis of AMPPNP would not be detected (section 6.3.3). The crystals of DbLnP were grown in the presence of AMPPNP and were not used for collection for up to 4 weeks. This could, assuming that the crystalline DbLnP was active, be sufficient time for AMPPNP to be hydrolysed. Assuming that the pH optimum is similar to ATP, DbLnP would be active as the enzyme has 75 % of the maximal activity at pH 5.0 in which the crystals were grown (200 mM MgCl₂, 50 mM acetate, pH 5.0,

9-12 % MPD, 10-20 % PEG 3350, and 5 mM AMPPNP). Self hydrolysis of AMPPNP could also contribute to the results observed. This is likely as the crystals of DbLnp were grown in an acidic solution and that AMPPNP has been reported to be unstable in acidic conditions resulting in hydrolysis to a phosphate and the AMPPN (Gresser, et al., 1984).

AMPPN was also detected in the substrate stock solutions as determined by mass spectrometry of the AMPPNP in sodium hydroxide. The substrate solution used in the crystallisation revealed peaks that corresponded with molecular weight of AMPPNP minus hydrogen (505.2 Da), AMPPNP minus hydrogen with potassium (543.0 Da), also various peaks of AMPPNP with 1 or 2 lithium ions (6 Da each). Also present was a peak that corresponded with AMPPN minus hydrogen at 425.3 Da and 431.2 Da, respectively. It is possible that DbLnp had a higher affinity for AMPPN than AMPPNP, which would account for AMPPN being present in the active site.

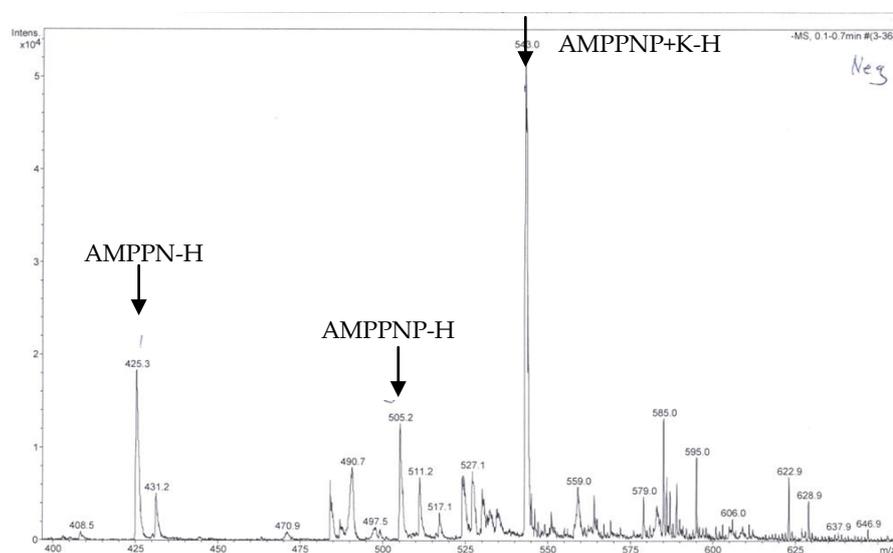


Figure 5-29 Mass spectrometry plot of AMPPNP. Arrows indicate the mass of compounds of interest.

5.4.3.4 Metal ion binding in the active site

In addition to the substrate in the active site, there was electron density to support the presence of a metal ion. It was assumed that this was a manganese

ion based on the X-ray fluorescence scanning on DbLnP+PO₄ crystal (Figure 5-14). The metal ion was positioned in the same place as observed for the previous DbLnP structure. The most notable feature of the metal ion is that these data suggest that the manganese is orientated in a pyramidal coordination, where the manganese is coordinated by 4 waters and the oxygen from the free phosphate. This is unlike the previous DbLnP structures which both show the metal ion is octahedrally coordinated by 5 waters and the phosphate. The missing water in the DbLnP bound to three phosphates (section 5.2.8) was not coordinated by any residues rather by two phosphates. It is likely that the missing water molecule could not be resolved due to the low resolution data (2.9 Å). However deleting the waters and performing Refmac5.0 resulted in a positive 2|F_o|-|F_c| map that only supported 4 waters. The distances between the manganese ion and the waters and the phosphate range between 2.1 Å and 2.3 Å which is consistent with other reports of O-Mn distances which are typically between 2.15-2.20 Å (Harding, 2004). The phenomenon of manganese coordinated in a square pyramidal fashion is common. At press, Harding's survey of the protein data base reported that 28.4 % of metalloproteins containing manganese were coordinated by 5 atoms (Harding, 2001, 2004).

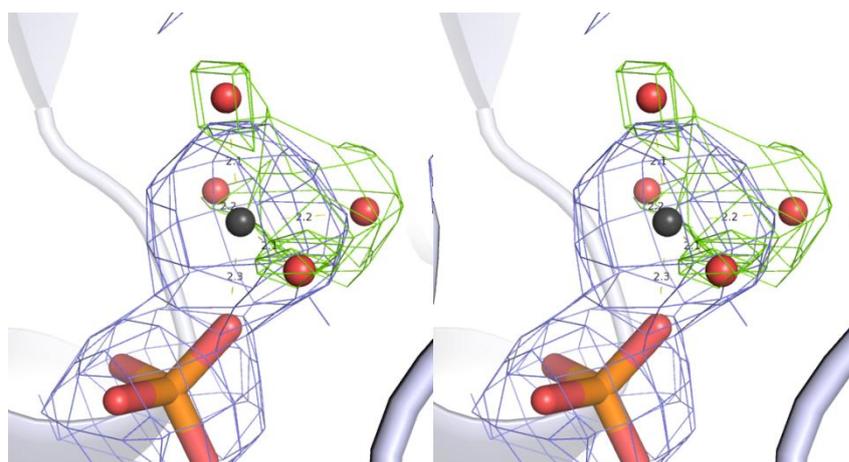


Figure 5-30 Stereo diagram of manganese (black sphere) bound in the active site of DbLnP+AMPPN.

Blue mesh represents the |F_o|-|F_c| contoured to 1 σ . The 2|F_o|-|F_c| map (contoured to 3 σ) in green was produced with Refmac5.0 with the waters (red spheres) removed. Waters were placed afterwards to illustrate their position.

5.4.3.5 *DbLnP co-crystallised with AMPPNP does not contain the nucleophilic water.*

An important feature of the catalysis of NTPDases is the presence of the nucleophilic water. It is commonly agreed that the catalytic mechanism is based on the nucleophilic attack of the phosphoanhydride bond by a de-protonated water located $\sim 3 \text{ \AA}$ from the γ -phosphate (Cohn, et al., 1956; Flaherty, et al., 1994; Vorobiev, et al., 2003; Zebisch, et al., 2008). The catalytically active water is orientated by Ser163 and Glu131 which was observed in the DbLnP+PO₄ structure (section 5.2.11).

Despite the importance of the nucleophilic water there was no evidence of a water molecule that is positioned near the phosphate that could perform nucleophilic attack of the phosphate in the DbLnP co-crystallised with AMPPNP. The absence of the water molecule supports the idea that AMPPNP is hydrolysed (section 5.4.3.3)

5.4.3.6 *Superimposition of the AMPPN with the rat NTPDase2 model*

The AMPPNP bound in the active site of the rat NTPDase2 demonstrated the most likely position of the substrate that could be hydrolysed by an NTPDase. Comparison of the orientation of the substrates was performed by superposition of this structure with DbLnP with AMPPN + phosphate bound (Figure 5-31).

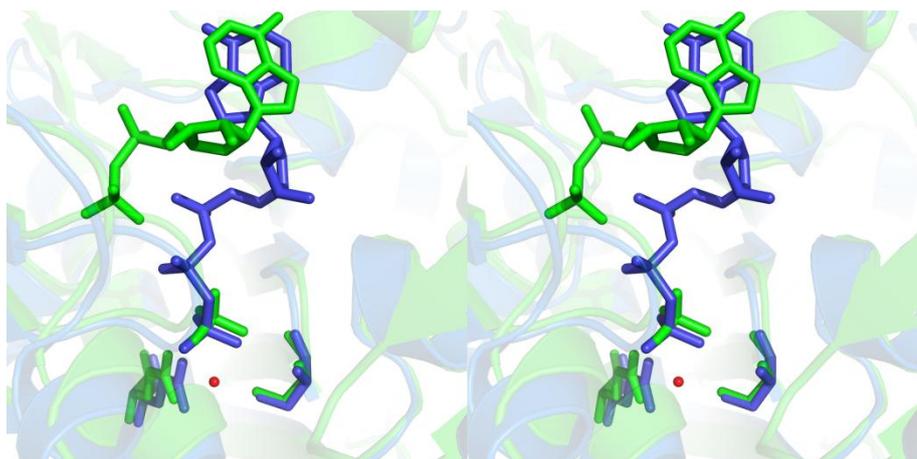


Figure 5-31 Stereo diagram of the superposition of rat bound AMPPNP (purple) and DbLnP bound AMPPN and phosphate (green). Red sphere represents the nucleophilic water in the NTPDase2 model.

Comparison of the two structures, the position of the glutamate, serine and the metal ion coordinated phosphate are located in equivalent positions. However there are many significant differences. For example, the position of the adenine in the rat NTPDase2 structure is twisted ($\sim 30^\circ$). The adenine in this model is not coordinated by any residues in the active site, in contrast the C6 amide of adenine in the DbLnP model is hydrogen bonded with Glu307 (section 5.4.3.1) The differences in the amino acid sequence may account for the differences in how the base is bound in the active site. The consequence of this is that the ribose interacts with different residues; in NTPDase2 the ribose interacts with C-terminal domain residues Arg245 and Asp246 and in DbLnP the ribose associates with N-terminal domain residues Lys43, Lys45 and Ser16. However, the most notable difference is the position of the phosphate tail. In NTPDase2 the phosphate tail is in a stretched conformation that extends towards the nucleophilic water, Glu166 and Ser206 whereas in DbLnP the phosphate tail is shifted away from the catalytic residues. As discussed earlier, the phosphate tail of AMPPN did not bind well in the phosphate binding site, rather the phosphates interacted with the side-groups of Lys43 and Arg19 which had shifted away from the phosphate binding site (section 5.4.3.1).

5.5 Discussion

The DbLnP structures in this chapter were found to be in the closed conformation. Compared with the open conformation of the apo-7WC the two domains have a much lower propensity for movement. By using both the open 7WC and closed DbLnP hinge domains could be inferred. This showed that two hinges are present in the regions between the two domains which differ to that of previous accounts of domain movement in members of the actin superfamily, where only one hinge region was observed. Both the hinge regions are within the boundaries of the ACRs (ACR3 and 5), and thus expand the roles of these regions.

The movement confirms hypotheses proposed by others, that domain movement occurs in a butterfly motion much like others in the actin superfamily and is the first direct evidence that such a movement occurs within NTPDases.

The AMP-bound structure of 7WC showed a smaller degree of trans-domain shift, and a larger number of hinges. These localised movements are likely due to crystal packing constraints probably not the true representation of the domain movement.

Attempts to solve the DbLnP structure with AMPPNP bound found that AMPPNP was hydrolysed within the active site. Thus, the model of DbLnP with AMPPN bound along with phosphate potentially represents the binding mode of the product after hydrolysis. The consequence of the reaction has effects on the conformation of the product and residue side-chains, the metal ion coordination number and the loss of the nucleophilic water.

During the nucleophilic attack of the substrate the nucleophilic hydroxyl forms a covalent bond with the γ -phosphate. If this model does represent the transition state during hydrolysis, it is expected that this nucleophilic water is not visible in this model. However the loss of water from the metal ion is unusual. It is most likely that this can be explained as a consequence of poor resolution of the data. The Mn^{2+} or potentially Mg^{2+} or Cu^{2+} ions (which are present in the crystallisation buffer and picked up on the X-ray fluorescence scan, respectively) are all very stable in octahedral coordination and in that state, the loss of water seems unlikely as the metal would quickly coordinate water that is present in the vicinity. However, the missing water molecule is not coordinated by any residues in the DbLnP structure. For the phosphate-bound model of DbLnP the water is coordinated by two phosphates and in the rat NTPDase2 model bound to AMPPNP and ADP the equivalent oxygen from the β -phosphoramidate contributes to the coordination (Zebisch, et al., 2008). This reduced coordination number of the metal ion is likely to be restored once the product is removed and new substrate binds.

After the hydrolysis, the orientation of the phosphate tail has moved away from the binding site. In particular the α - and β -phosphates are pulled from the

phosphate binding site 1. This movement is facilitated by the shift in the position of the Arg19 that pulls the α - and β -phosphates away from the catalytic residues. This movement causes the twist of the nucleobase away from being parallel with the Try303 which would result in the weakening of the π -stacking interaction. This destabilised binding mechanism would reduce the affinity and contribute to the expulsion of the product. This complements the assumption that the soluble legume NTPDases release the NDP after hydrolysis, observed for the soluble potato apyrase (Chen, et al., 2001).

It is conceivable that the full release of the hydrolysis product is facilitated by the movement of domains. If this is the case it would explain why the intermediate was not released in the DbLnP structure with the AMPPN bound, as the structure is constrained in the closed form by crystal packing constraints.

This thesis shows two very important features of the catalytic mechanism that has yet to be shown for the NTPDases. 1) That trans-domain shift does occur during substrate binding and 2) The putative binding mode for the hydrolysed substrate.

6 Kinetic parameters of 7WC and DbLnP

6.1 Introduction

7WC and DbLnP are NTPDases which hydrolyse the phosphoanhydride bond of nucleotide di- and triphosphates. This reaction requires the presence of a divalent cation (Chen, 2008; Etzler, et al., 1999). To gain a full understanding of the catalytic mechanism of these enzymes the kinetic parameters were determined for both 7WC and DbLnP.

7WC and DbLnP were tested against a range of nucleotide di- and triphosphates including purine and pyrimidine bases. In these experiments the kinetic parameters for the substrates ADP, ATP, CDP, CTP, GDP, GTP, TDP, TTP, UDP and UTP were characterised.

6.2 Kinetic analysis of 7WC

6.2.1 Divalent cation co-factor dependent 7WC catalysis.

7WC catalysis is dependent on a divalent cation (Chen, 2008). In the presence of 5 mM EDTA there was no activity (data not shown). To ensure that the co-factor was not the limiting factor in further activity assays the enzyme activity was measured for ATP and ADP with 0-800 μ M of the preferred divalent cation, CaCl₂. The activity assays were performed at the preferred pH for hydrolysis of each substrate as described in section 2.7.7.1.

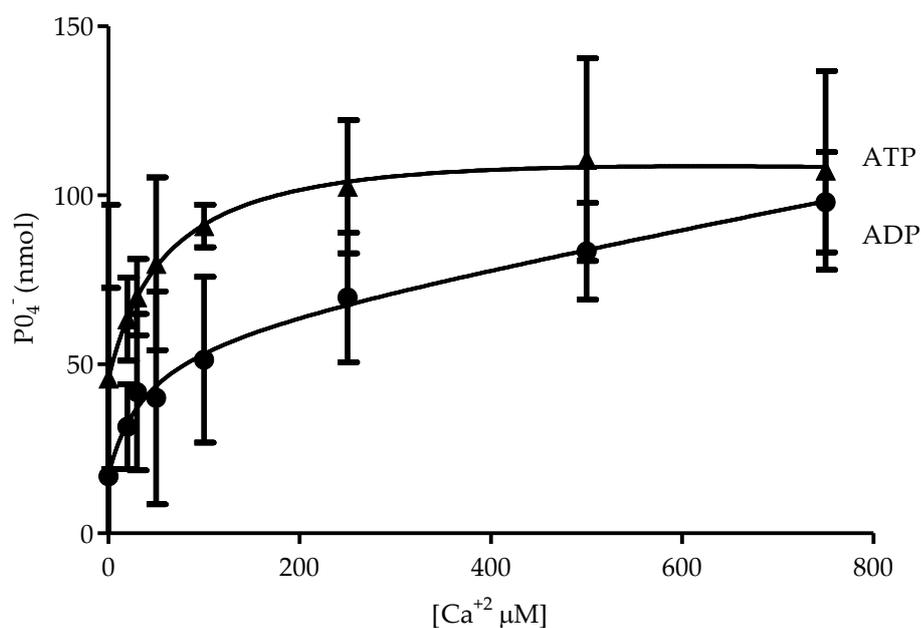


Figure 6-1 Calcium saturation curve for 7WC with ADP and ATP as substrate.

The EC₅₀ was found to be 36.6 µM and 48.4 µM for ADP and ATP respectively. Maximal activity was reached with at least 500 µM for both ADP and ATP and so a value of 750 µM CaCl₂ was used in further enzyme assays for 7WC.

6.2.2 Influence of pH on the activity of 7WC.

For each substrate characterised, the enzyme activity was measured from pH 3.5 to 10.5 as described in section 2.7.7.2. The activity was measured in triplicate and subtracted from a blank, which was important as the pH had a slight effect on the colourimetric change of the developing solution.

The results show that all the tested substrates, except for the cytosine nucleotides, have a pH optimum between pH 8-8.5. Cytosine nucleotides optimally hydrolyse at pH 6.5. The pH optimum is similar in all cases for both the nucleotide di- and triphosphate; observation of the pH profiles show similarities between each nucleotide di-/triphosphate which show that adenosine nucleotides, cytosine nucleotides and uracil nucleotides have a pronounced peak at the pH optimum whereas guanine nucleotides and thymine nucleotides are catalytically optimal over a broad range.

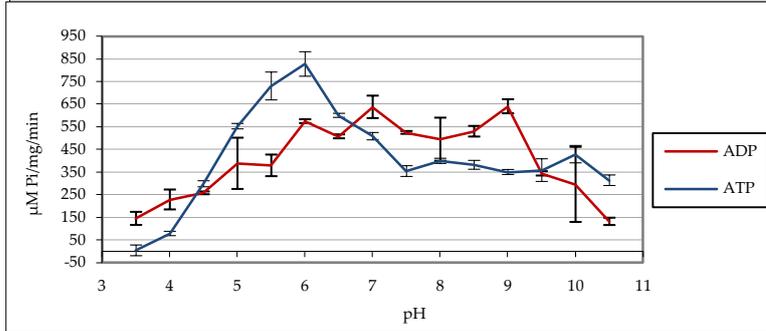
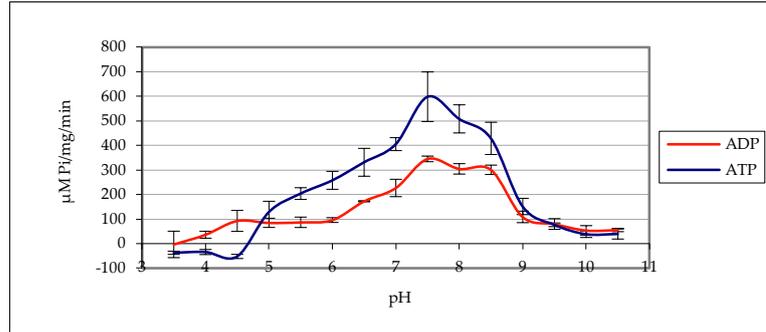
These data suggests that the optimal hydrolysis is dictated by the nucleobase rather than the number of phosphates each nucleotide has (Table 6-1). The determined pH optimum for each nucleotide was used in further experiments.

6.2.3 Substrate specificity

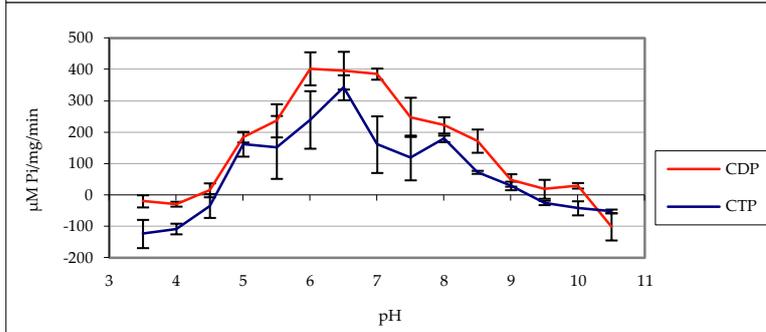
Specific activity was measured for all common di- and trinucleotides. The ability for 7WC to turn over AMP the non-hydrolysable substrate AMPPNP, and pyrophosphate was also measured. Figure 6-3 shows that the 7WC has a similar NDPase:NTPDase ratio for all the nucleotides and that ADP and ATP are the preferred substrates.

7WC has a negligible activity for AMP, AMPPNP and pyrophosphate which is consistent with other reports for homologous mammalian NTPDases (Wang, et al., 1996; Zebisch, et al., 2007). To be classed as an NTPDase, the enzyme must be divalent cation dependent and have the ability to hydrolyse a range of di- and trinucleotides but not mono-nucleotides, thus 7WC fulfils the requirements for being an NTPDase.

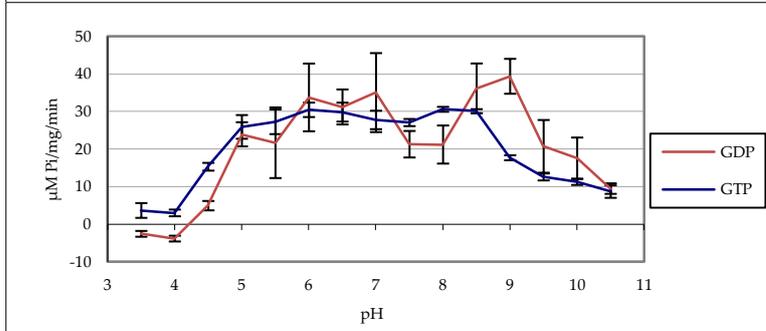
A)



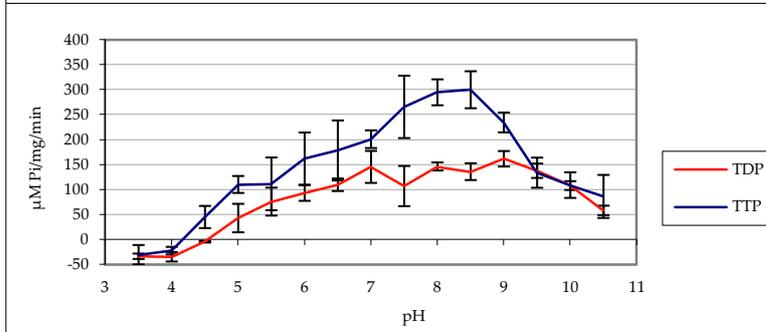
B)



C)



D)



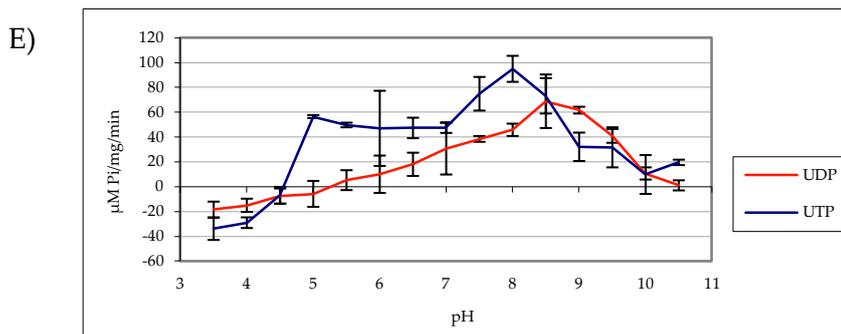


Figure 6-2 7WC enzyme activity over a range of pH with various nucleotides. A) adenine nucleotides, B) cytosine nucleotides, C) guanine nucleotides, D) thymine nucleotides, E) uracil nucleotides. Points are mean \pm standard deviation of three independent experiments.

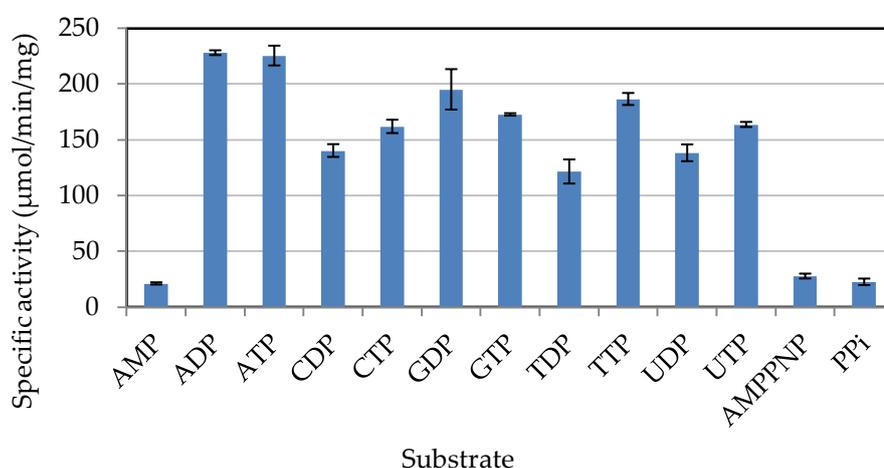


Figure 6-3 Specific activity of various substrates for 7WC

6.2.4 Determination of the initial rate

Since the potato apyrase, 7WC and DbLnP are all soluble enzymes and have a high sequence similarity (~45 %) it was assumed for this study that the catalytic mechanism for 7WC and DbLnP is the same as the soluble potato NTPDase represented in (Equation 1-3) (Etzler, et al., 1999; Roberts, et al., 1999).

Because of the release of NDP which can bind and be hydrolysed by NTPDases, considerations were made to ensure that the concentration of NDP is negligible when measuring the NTP kinetic parameters. It is essential that when kinetic assays are measured that the initial rate is measured. This is to ensure that the concentration of the NDP remained low so not to affect the measured activity.

The rate was measured when ~10 % of the total substrate was hydrolysed. Since this assay is a stop reaction assay, a time course assay was performed to determine the time at which 10 % of the substrate was hydrolysed. Similar experiments were performed with ADP so that the product inhibition from AMP did not occur (Figure 6-4).

The liberated phosphate ion was being measured in the assay. The total phosphate ion in the assay is 50 and 100 μ Mole for NDP and NTP, respectively.

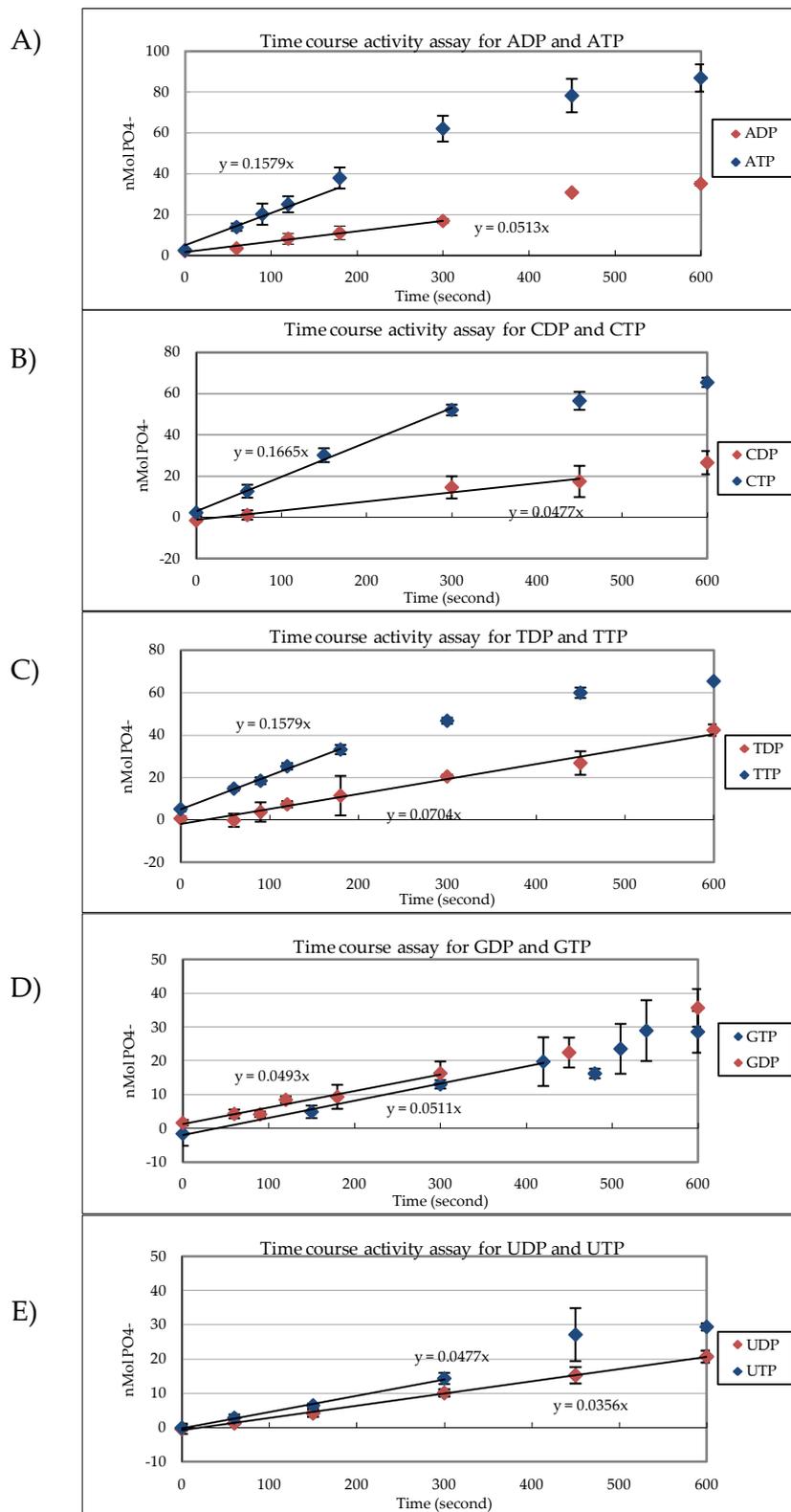


Figure 6-4 Time course activity assay of 7WC with various nucleotides. A) adenine nucleotides, B) cytosine nucleotides, C) guanine nucleotides, D) thymine nucleotides, E) uracil nucleotides. Points show the mean \pm standard deviation of three independent assays. The equations show the gradient of the linear region for each data series.

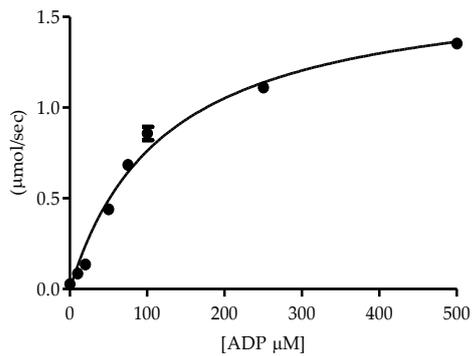
The amount of substrate hydrolysed was determined at various time points using optimal conditions for all nucleotides. The protein concentration was optimised to $0.002 \mu\text{Mol}\cdot\text{L}^{-1}$ to allow adequate time to perform the assay accurately. Generally the nucleotide triphosphates have a 2-3 fold faster initial rate compared to that with the nucleotide diphosphates. However this trend does not hold true for the guanine and uracil nucleotides which have an equal rate for both di- and tri phosphates. Using the initial rate, the time to perform the assays for further kinetic analysis was determined using Equation 5. This gave a time between 60-209 seconds, the times for each nucleotide are summarised in Table 6-1.

6.2.5 Michaelis-Menten kinetics for 7WC

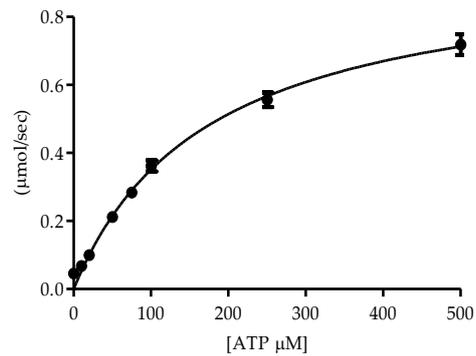
By measuring the Michaelis-Menten kinetics it is possible to compare the catalytic efficiency for various substrates and with other NTPDases. For each substrate the apparent K_M , k_{cat} and V_{max} were calculated by measuring the initial velocity with a range of substrate concentrations which when plotted exhibited normal Michaelis-Menten kinetics for all tested substrates.

Experiments were performed in triplicate and line of best fit was made using a non-linear regression curve and analysed by Graphpad prism software.

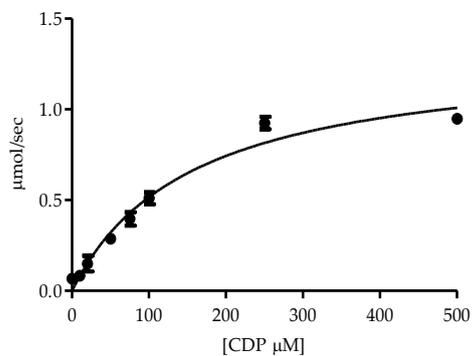
A) Michaelis-Menten data for 7WC with ADP



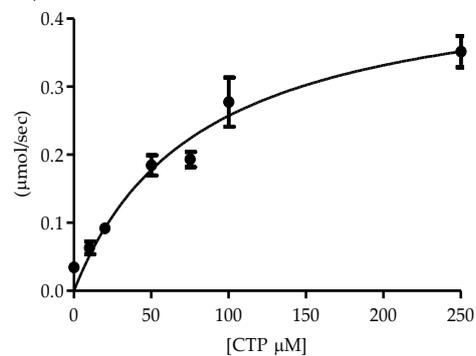
B) Michaelis-Menten data for 7WC with ATP



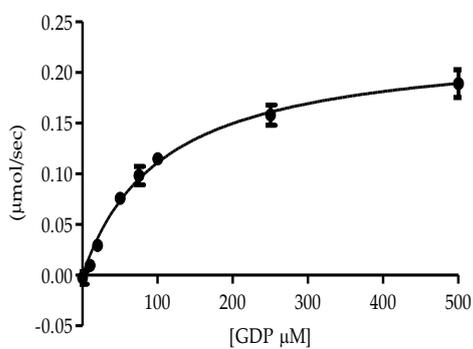
C) Michaelis-Menten data for 7WC with CDP



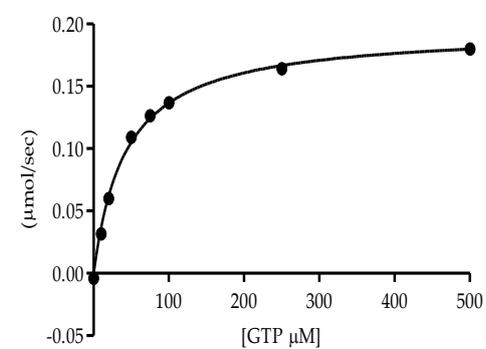
D) Michaelis-Menten data for 7WC with CTP



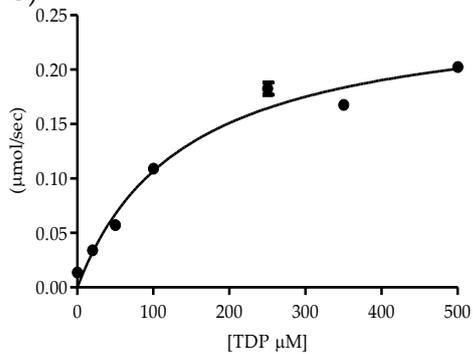
E) Michaelis-Menten data for 7WC with GDP



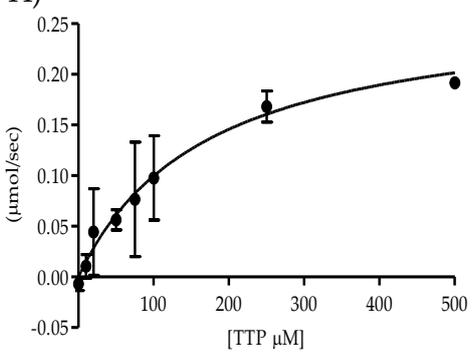
F) Michaelis-Menten data for 7WC with GTP



G) Michaelis-Menten data for 7WC with TDP



H) Michaelis-Menten data for 7WC with TTP



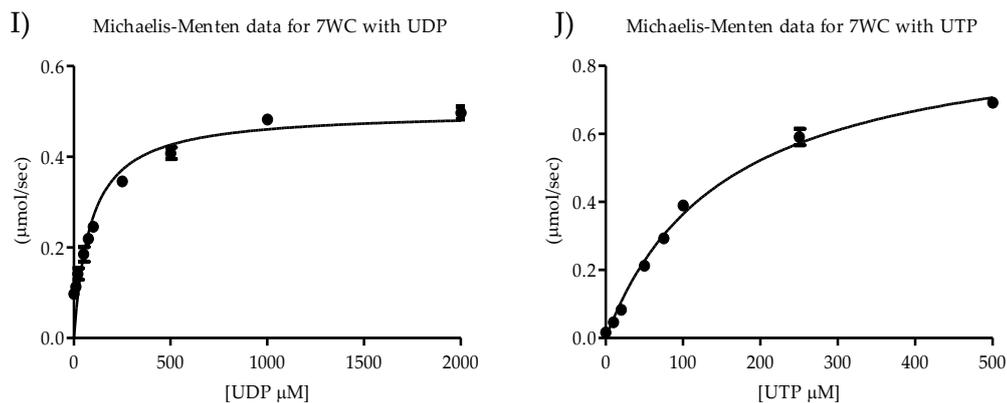


Figure 6-5 Michaelis-Menten plots for 7WC for various nucleotides. A) ADP, B) ATP, C) CDP, D) CTP, E) GDP, F) GTP, G) TDP, H) TTP, I) UDP, J) UDP. Points represent the mean of three replicates \pm standard error.

Using these data, key kinetic parameters were determined. The apparent K_M which reflects the binding of the substrate to the active site was calculated by determining the substrate concentration where half the active sites are occupied. The determined K_M values generally ranged between \sim 100-150 μmol for most substrates, exceptions are GTP, CTP and UTP which have lower values of 42 ± 2 , 80 ± 14 and 87 ± 12 μmol , respectively. TTP has a considerably higher K_M value of 232 ± 32 . Suggesting that 7WC has the highest affinity for GTP. The V_{max} were divided by the E_0 or the amount of active sites in solution to determine the turnover rate (k_{cat}). 7WC was found to have one active site, thus E_0 is equal to the protein concentration expressed in $\text{moles} \cdot \text{L}^{-1}$ (Chen, 2008). Since the E_0 is constant in all assays, k_{cat} follow the same trend as V_{max} and ranged from 219 to 797 sec^{-1} . The k_{cat} were used to calculate the total catalytic efficiency, k_{cat}/K_M . This ratio reveals that all substrates can be hydrolysed within the range of $1 \times 10^{+6} \text{M}^{-1}\text{s}^{-1}$ to $1 \times 10^{+7} \text{M}^{-1}\text{s}^{-1}$. The highest k_{cat}/K_M ratios were determined for ADP and CDP with values of 6.5×10^6 and $3.9 \times 10^6 \text{M}^{-1}\text{s}^{-1}$, respectively. The lowest ratio was observed for TTP, TDP and GDP with values of 7.4×10^5 , 7.8×10^5 and 1.1×10^6 , respectively.

Comparing the k_{cat}/K_M of all substrates shows no obvious trend pertaining to the preference for NDPase and NTPase activity, although differences between nucleobases may provide information on the binding and catalytic mechanism.

	optimum pH	time of assay (sec)	k_{cat} (s^{-1})	K_M (μmol)	V_{max} (μmol sec^{-1})	k_{cat}/K_M ($M^{-1}s^{-1}$)
ADP	8	90.1	800 ± 30	120 ± 10	1.7	6.5×10^6
ATP	8	51.8	450 ± 20	170 ± 16	0.96	2.6×10^6
CDP	6.5	108.7	620 ± 30	160 ± 20	1.32	3.9×10^6
CTP	6.5	60	220 ± 20	80 ± 10	0.47	2.7×10^6
GDP	8.5	195	110 ± 3	100 ± 10	0.23	1.1×10^6
GTP	8	101	97 ± 10	40 ± 2	0.2	2.3×10^6
TDP	8	71	110 ± 6	140 ± 21	0.22	7.8×10^5
TTP	8	60	140 ± 10	170 ± 30	0.27	7.9×10^5
UDP	8.5	140	230 ± 10	90 ± 10	0.5	2.7×10^6
UTP	8	209	430 ± 13	150 ± 10	0.92	2.8×10^6

Table 6-1 Summary of kinetic parameters for 7WC.
The k_{cat} and K_M values are \pm standard error.

6.3 Kinetic parameters of DbLnP

6.3.1 Divalent cation co-factor dependent DbLnP catalysis.

To ensure that the divalent cation concentration was not a limiting factor identical experiments were conducted with ATP and ADP for CaCl_2 as described in section 6.2.1. Analyses of the results show an EC_{50} of 74.4 μM and 84.5 μM for ADP and ATP, respectively (Figure 6-6). Both nucleotides reached maximum activity of at least 500 μM CaCl_2 , so in further experiments using 750 μM CaCl_2 was carried out.

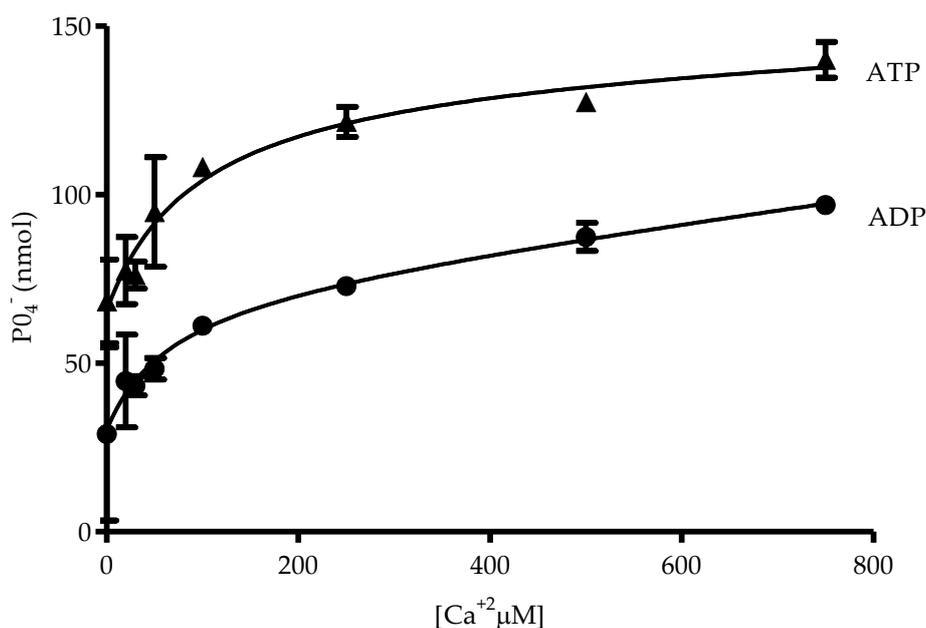


Figure 6-6 Calcium saturation curve for DbLnP with ADP and ATP as substrate. Points are the mean \pm standard error of three independent experiments.

6.3.2 Influence of pH profile on the activity of DbLnP.

Using the same pH range used for 7WC (section 6.2.2) the pH profile for DbLnP activity was performed. A summary of the pH optimum for all tested substrates are shown in Table 6-2. The data reveal that for all tested nucleotides except UDP and UTP, the hydrolysis of NDP is achieved optimally at slightly alkaline conditions, pH 8.0-8.5, whereas the NTP hydrolysis is optimal at pH 5.5 (Figure 6-7,A-D). In many cases DbLnP is able to hydrolyse nucleotides over a broad range of pH and so pH optimum was taken from the mid-point of this range. Hydrolysis of both UDP and UTP occur at pH 9.5 (Figure 6-7,E).

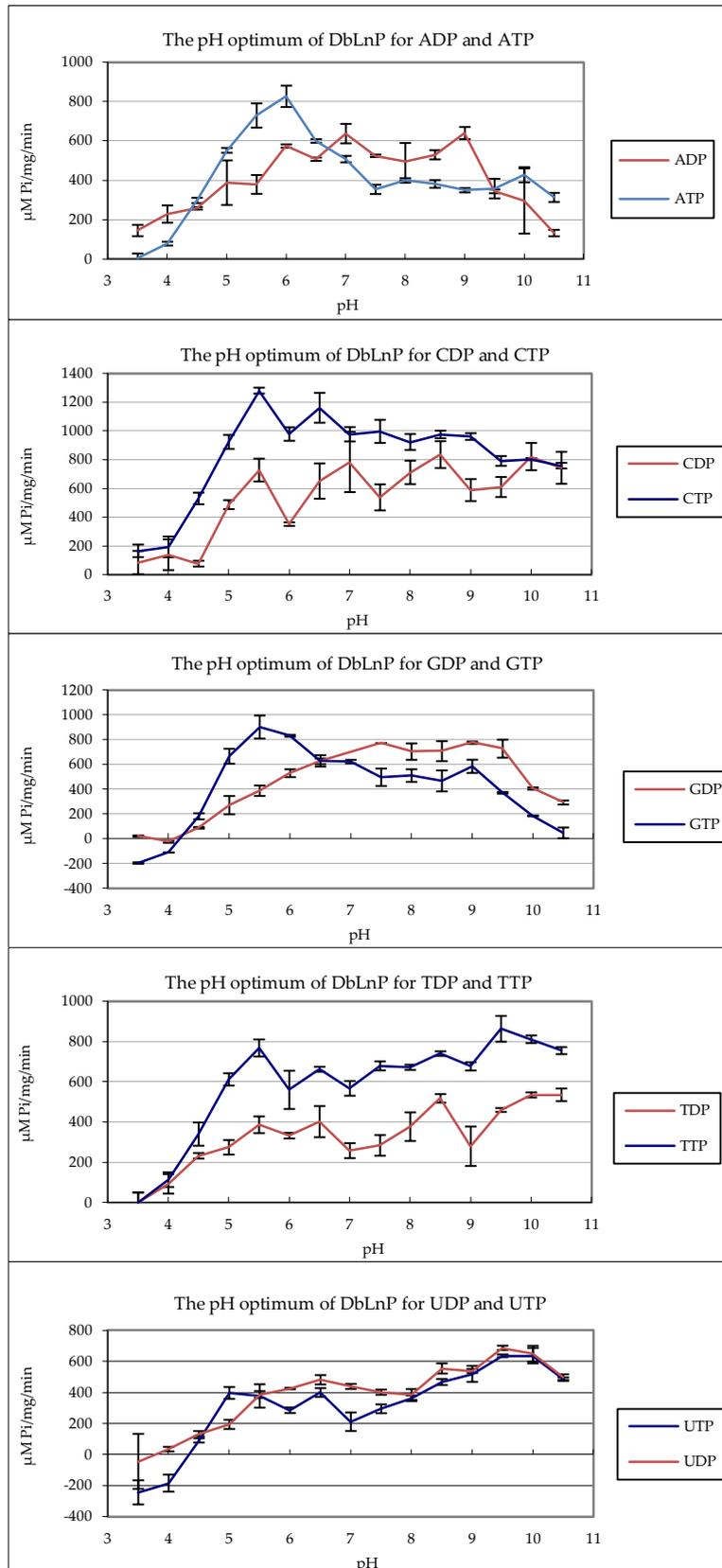


Figure 6-7 DbLnP enzyme activity over a range of pH with various nucleotides. A) adenine nucleotides, B) cytosine nucleotides, C) guanine nucleotides, D) thymine nucleotides, E) uracil nucleotides. Points are mean \pm standard deviation of three independent experiments.

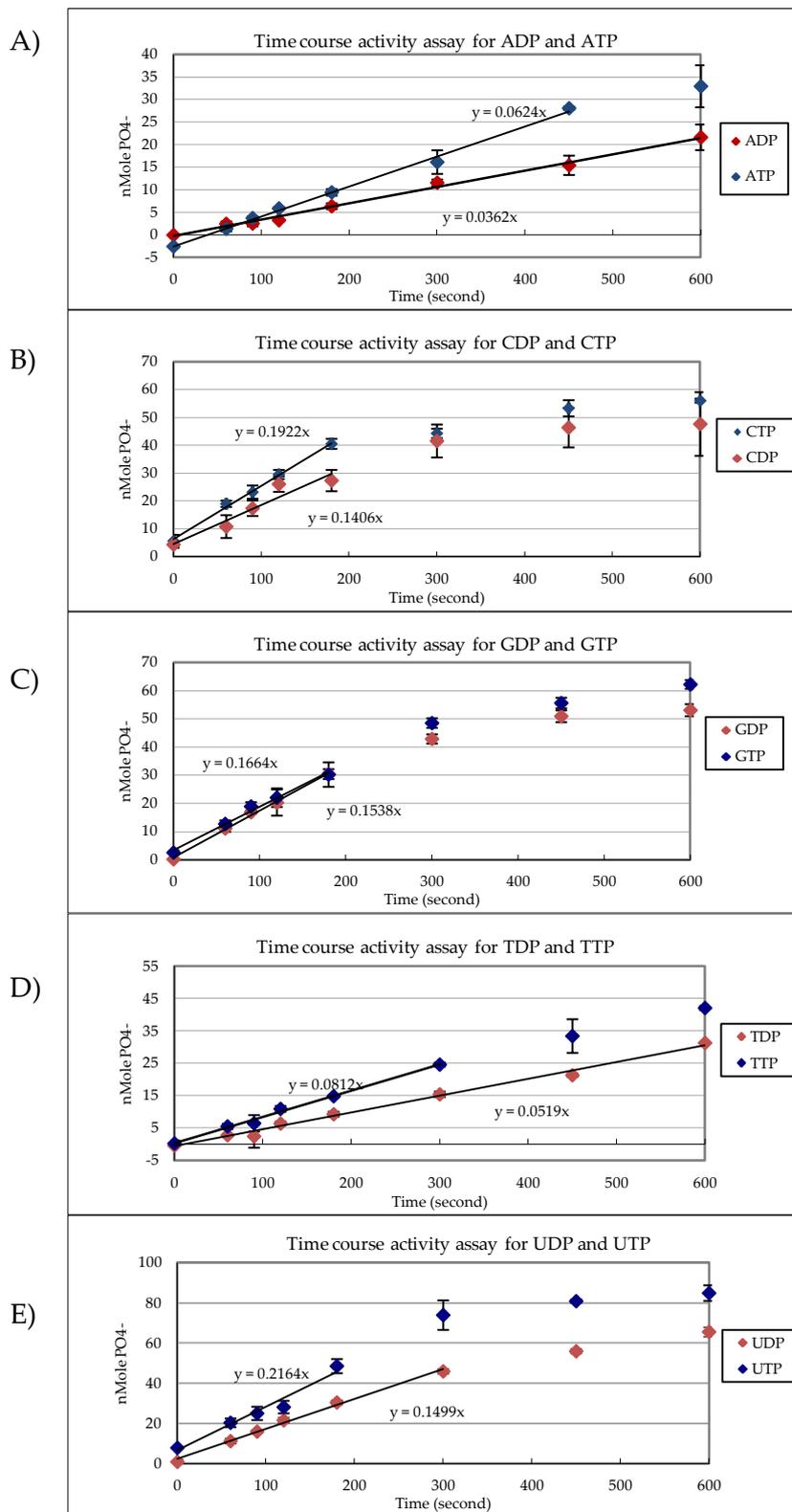


Figure 6-8 Time course activity assay of DbLnP with various nucleotides. A) adenine nucleotides, B) cytosine nucleotides, C) guanine nucleotides, D) thymine nucleotides, E) uracil nucleotides. Points show the mean \pm standard deviation of three independent assays. The equations show the gradient of the linear region for each data series.

6.3.3 Substrate specificity of DbLnP

The specific activity was measured for the rate of DbLnP catalysis with various substrates. The results show that DbLnP can hydrolyse each nucleotide effectively with CTP, TTP and uridine nucleotides hydrolysed the most effectively. The nucleotides, ADP and GTP are hydrolysed the least effectively. Generally DbLnP prefers the pyrimidine nucleotides (containing cytosine, uracil or thymine groups) over the purine nucleotides (containing guanine or adenine groups). In all cases except for guanine nucleotides, the hydrolysis is higher for NTPDase than NDPase activity, although the difference is insignificant between uridine nucleotides. Negligible activity was measured for both AMP and AMPPNP.

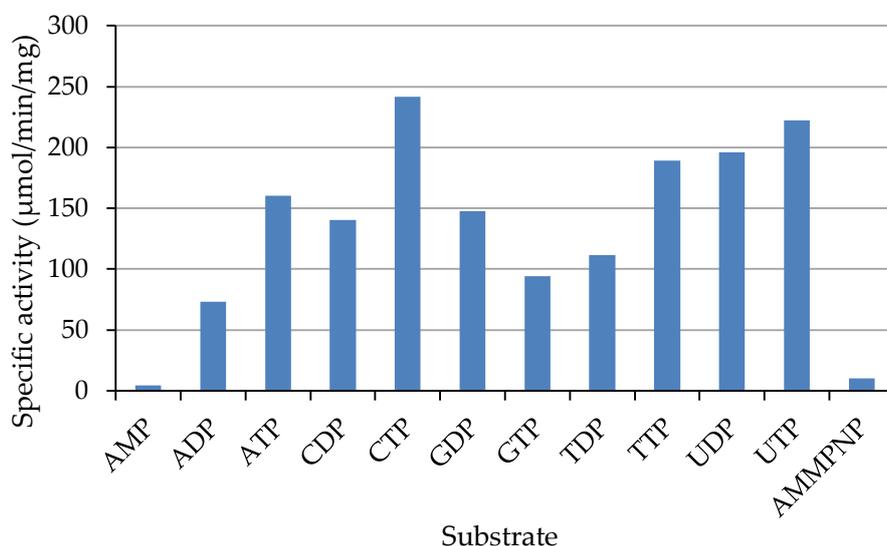


Figure 6-9 Specific activity of various substrates for DbLnP.

6.3.4 Determination of the initial rate

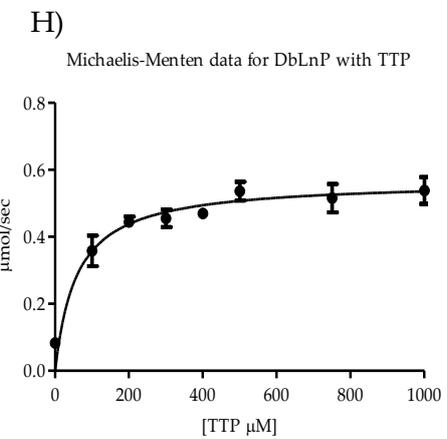
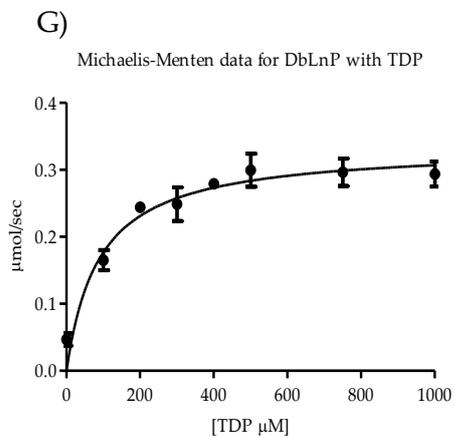
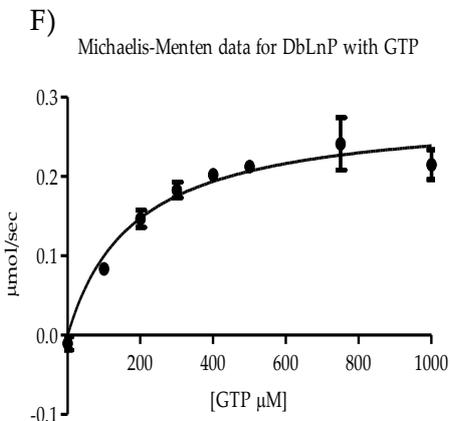
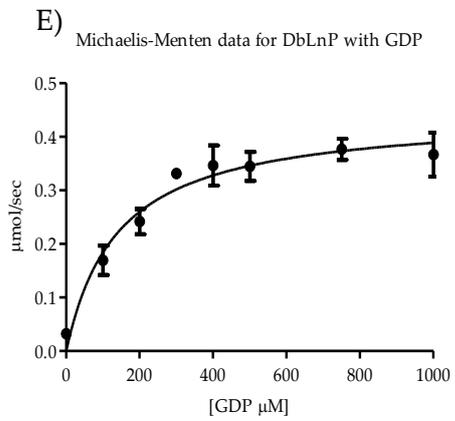
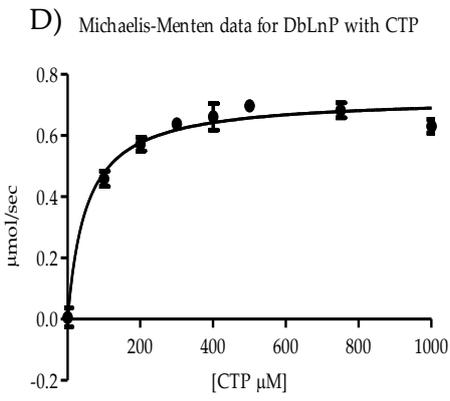
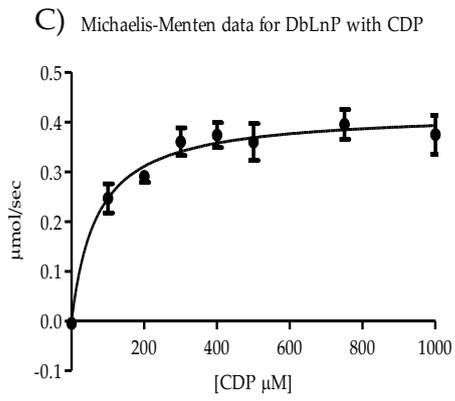
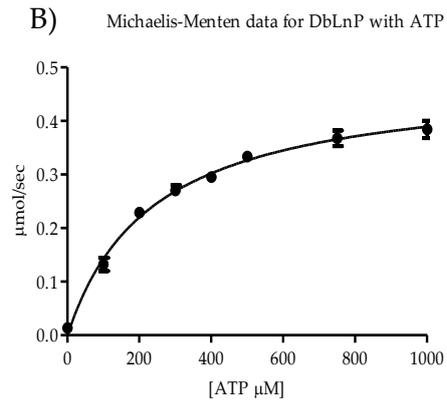
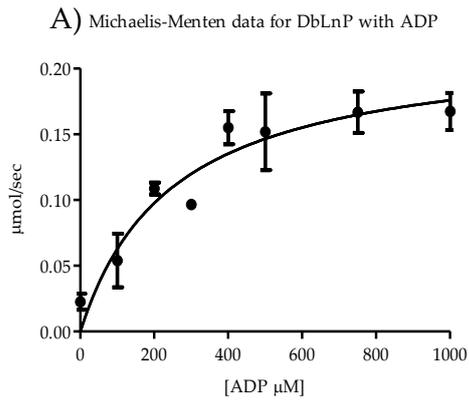
For DbLnP it was assumed that the catalytic mechanism was the same as 7WC. That is that DbLnP hydrolyses the γ -phosphate and the β -phosphate sequentially and that ADP is released. Thus the initial rate was very important for reasons mentioned in section 6.2.4. The protein concentration was adjusted to $0.001 \mu\text{Mol}\cdot\text{L}^{-1}$ to allow enough time for the assay to be accurately performed. NDPase and NTPase activities were similar for each nucleotide however they varied between nucleotides. The results suggest two groups; with adenine and thymine ~2-3 fold lower than the other tested nucleotides.

The time determined for 10 % of the total substrate to be hydrolysed ranged between 28.5 and 150 seconds representing TDP and ADP, respectively. These times were used for subsequent kinetic characterisation of DbLnP.

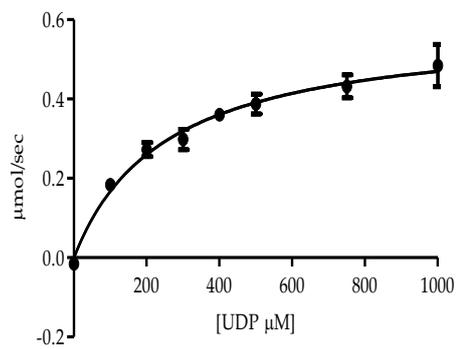
6.3.5 Michaelis-Menten kinetics for DbLnP

The Michaelis-Menten kinetics was determined as outlined in section 6.2.5 for DbLnP. Analysis of the data reveals an apparent K_M range between 51.7 and 252.9 μM for all the tested substrates (Table 6-2) with the preferred substrates being the pyrimidines cytosine and thymidine. The highest K_M values are shown for adenosine and uridine di-triphosphates. Generally the K_M values for NDPase and NTPase activity are similar for each nucleobase type. In past publications the K_M of DbLnP for ADP was shown to be 615 μM . This value is higher than the values that I determined. There are a number of reasons that may explain the discrepancies. Firstly the assay was conducted at pH 6.8, not 8.0 and that DbLnP was isolated from *D. biflorus* and therefore may have post-translational modifications effecting binding affinity (Etzler, et al., 1999).

The V_{max} of DbLnP ranged from 0.22 to 0.73 $\mu\text{mol sec}^{-1}$ which match the same trend as the catalytic constant, k_{cat} (219.8 and 725.3 s^{-1}). The highest k_{cat} values are seen for the NTPase activity on pyrimidine nucleotides (k_{cat} value for CTP = 725.3 ± 22.7 , UTP = 666.8 ± 53.2 , TTP = 567.1 ± 28.5) which although UDP was shown with a high K_{cat} value of 588.8 ± 53.2 .



I) Michaelis-Menten data for DbLnP with UDP



J) Michaelis-Menten data for DbLnP with UTP

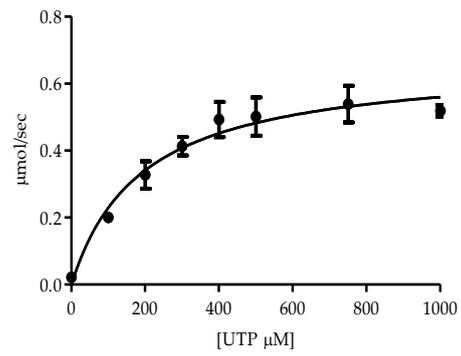


Figure 6-10 Michaelis-Menten plots for DbLnP for various nucleotides.

A) ADP, B) ATP, C) CDP, D) CTP, E) GDP, F) GTP, G) TDP, H) TTP, I) UDP, J) UDP.

Points represent the mean of three replicates \pm standard error.

	Optimal pH	Time of assay (sec)	k_{cat} (s^{-1})	K_M (μmol)	V_{max} ($\mu\text{mol s}^{-1}$)	k_{cat}/K_M ($M^{-1}s^{-1}$)
ADP	8	150	220±30	250±70	0.22	8.8 × 10 ⁶
ATP	5.5	138	480±20	240±20	0.48	2.0 × 10 ⁶
CDP	8.5	35	420±20	70±20	0.42	6.0 × 10 ⁶
CTP	5.5	52	720±20	50±10	0.73	1.4 × 10 ⁷
GDP	8.5	30	440±30	140±40	0.44	3.1 × 10 ⁶
GTP	5.5	47.4	282±20	180±40	0.28	1.5 × 10 ⁶
TDP	8.5	28.5	330±20	90±20	0.33	3.8 × 10 ⁶
TTP	5.5	33	570±30	60±20	0.57	9.6 × 10 ⁶
UDP	9.5	33	590±40	250±50	0.59	2.3 × 10 ⁶
UTP	9.5	46	670±50	190±50	0.67	3.5 × 10 ⁶

Table 6-2 Summary of the kinetic parameters of DbLnP

The k_{cat}/K_M value gives an indication to the overall catalytic efficiency, which takes into account the catalytic constant and the binding affinity. The values obtained vary between 8.8×10^5 and $1.4 \times 10^7 M^{-1}s^{-1}$. The highest values are for the pyrimidines; cytosine triphosphate and thymidine triphosphate suggesting that DbLnP may have a specific role in catalysing pyrimidine triphosphates. In comparison pyrimidine rings are significantly smaller than the purine bases, this would allow for the substrate to bind deeper into the binding pocket than the pyrimidine nucleotide bases. If this was the case, either the phosphate tail would be in a stretched conformation or that DbLnP would need be in a more closed conformation to ensure binding.

6.4 Discussion

Comparisons of the specific activity between 7WC and DbLnP show that the specific activity for each is within a similar range generally between 100 and 250 $\mu\text{mol min}^{-1} \text{mg}^{-1}$. These values are generally lower than that of other plant NTPDases. For example the host purified potato apyrase, exhibited values as high as 10,000 $\mu\text{mol min}^{-1} \text{mg}^{-1}$ for ATP and the pea purified NTPDase has an observed specific activity of 428 $\mu\text{mol min}^{-1} \text{mg}^{-1}$ for ATP (Chen, et al., 1987; Handa, et al., 1996). The specific activities measured for 7WC and DbLnP are however similar to that of the recombinant rat NTPDase1-3 which exhibit specific activities between 42.6 and 223.8 $\mu\text{mol min}^{-1} \text{mg}^{-1}$ for ATP (Zebisch, et al., 2007).

It is generally accepted that the phosphates from the NTP molecule associates strongly with the divalent cation. It is seen that the metal ion is important in binding to the phosphates of ATP within the active site of NTPDases and contributes to the catalytic mechanism by polarising the hydrolysed phosphoanhydride bond (Zebisch, et al., 2008). It is not surprising that both 7WC and DbLnP both require a divalent cation co-factor for nucleotide binding and coordination. The EC_{50} values for both 7WC and DbLnP are within the similar range to that determined in similar experiments performed on mammalian NTPDase1-3 (Drosopoulos, 2002; Vorhoff, et al., 2005; Zebisch, et al., 2007) which showed EC_{50} values for Ca^{2+} that range from 3.7-90 μM using ADP and ATP as the substrate.

These data show that the EC_{50} is lower for ADPase activity than ATPase for both 7WC and DbLnP, possibly due to either steric hindrance within the active site or differences to how the di-/triphosphates coordinate with the metal ion. The metal ion that resides in the active site has an impact in some cases on the NTPase:NDPase ratio in the mouse NTPDase3 and -8. In addition in the *Dictyostelium* actin it was observed that different metal ions had an impact on the distance of the terminal phosphate to the nucleophilic water and the angle of the terminal phosphate, the β - γ - bridging oxygen and the catalytic water. With this in mind it seems plausible that an exchange of metal ions is required before both

efficient NTP and NDP catalysis occurs (Kukulski, et al., 2005; Vorobiev, et al., 2003). A more comprehensive analysis of alternative divalent metal cations would give further clues as to the binding mechanism of the metal ions.

Recently it was found that extracellular ATP and other nucleotides can induce an increase in cytosolic calcium ions, which impact greatly on plant growth (Roux, et al., 2007). The relationship between NTPDases, ATP and calcium ions is unclear but it is possible that calcium ions act directly on NTPDases. Extracellular calcium ions have been measured as high as 10 μM in growing pollen tubes of *Lillium longiflorum* so it is possible that spikes in the calcium concentration could activate 7WC and DbLnP activity.

In addition to calcium ions acting directly with NTPDases, they may play an indirect role by regulating key receptors involved in rhizobium infection of legumes during nodulation.

During the perception of NOD factors there is an influx of extracellular calcium (Ehrhardt, et al., 1996) and that this induces the calcium and calmodulin-dependent protein kinase signal pathway (Levy, et al., 2004; Mitra, et al., 2004; Oldroyd, et al., 2006). Emerging evidence has shown that NTPDases in plants associate with calmodulin (Chen, et al., 1987; Steinebrunner, et al., 2000). If this is the case then NTPDases could regulate this transduction pathway by either by inhibiting the catalytic mechanism of calmodulin-dependent protein kinase by inhibiting the protein interaction with calmodulin or by reducing the substrate (ATP) required for autophosphorylation possibly through NOD factor induced NTPDase activity (Etzler, et al., 1999). This intrinsic link between, NOD factor perception, calcium influx, calmodulin, calmodulin-dependent protein kinase and NTPDases during early perception of rhizobium may give clues as to the role of NTPDases in the NOD factor signal transduction pathway.

Between 7WC and DbLnP the pH optimum for each substrate varies. 7WC generally hydrolysed nucleotide di- and triphosphates in similar pH conditions, with each nucleotide having a pH optimum that is specific to the nucleobase.

This phenomenon observed for 7WC has also been reported for other NTPDases for various mammalian and the *Legionella pneumophila* NTPDases. Human and rat NTPDase1, and -3 has a pH optimum for both ADP and ATP at pH 8.0 (Kukulski, et al., 2005; Sansom, et al., 2008; Zebisch, et al., 2007).

The pH optimum results for DbLnP is unlike 7WC, which suggest that the optimum is dictated by whether it is a nucleotide di- or triphosphate and that unlike 7WC, is not influenced by the nucleobase. Such a trend has been reported for rat NTPDase2 (Zebisch, et al., 2007).

These differences between pH optimum 7WC and DbLnP may represent two different nucleotide binding classes within the NTPDase superfamily. Comparison of the sequence alignment and a secondary similarity match with 7WC, DbLnP and the rat NTPDase2 structures, no obvious components could be found that could explain these differences.

The K_M values determined for 7WC and DbLnP are within the a similar range for the tested substrates and are similar to those determined for two isoforms of *S. tuberosum* NTPDases which have a reported K_M for ADP and ATP at 20 °C and 30 °C to be between 24 and 350 μmol (Kettlun, et al., 1982). The K_M for 7WC and DbLnP are slightly higher than that demonstrated for mammalian NTPDases which typically range between 10-150 μmol (Kukulski, et al., 2005; Zebisch, et al., 2007). Differences between plant and mammalian NTPDases are not surprising as they will have evolved for different physiological roles. In higher plants exogenous application of nucleotides as low as 40 μM have been shown to induce physiological responses, however if the extracellular ATP is too high these responses do not occur (Song, et al., 2006; Wu, et al., 2007). A model proposed by Roux and Steinbrunner suggested that an optimal extracellular ATP concentration is required for growth and that the NTPDase expression would correlate with the optimal ATP concentration required for growth. It is likely then that K_M values determined for 7WC and DbLnP for nucleotides correlate with optimal nucleotide concentration for physiological function such as growth.

Without accurate measurements of the extracellular ATP concentration it is difficult to be certain.

Comparing the K_M values between 7WC and DbLnP show little similarity between the nucleotide preference. DbLnP seems have a much affinity for cytosine and thymidine and significantly lower for the other tested nucleotides. This specificity suggests a defined physiological role such as regulating the phosphatidylcholine biosynthetic pathway which requires CTP for the formation of phospholipids (Clement, et al., 1999).

The most striking difference is the catalytic efficiency for NDP and NTP where 7WC is more efficient with NDP and conversely DbLnP has higher k_{cat}/K_M for NTP. The components that dictate NDPase/NTPase ratio has been a recurring topic in the literature. Despite a number of point mutations on mammalian NTPDases which have identified potential residues involved in distinguishing between NTP and NDP, the components that fully explain these differences are yet to be determined (Drosopoulos, et al., 2000; Grinthal, et al., 2000; Smith, et al., 1999a).

The k_{cat}/K_M ratio is a good indicator for the overall catalytic function since it takes in to consideration the binding affinity, catalytic function and the dissociation of the product from the enzyme. For 7WC the optimal substrates are adenosine and cytosine diphosphate. Both ADP and CDP contain amine groups on the C6 and C4 of the nucleotide base, respectively. These amine groups are positioned in the AMP bound 7WC structure which associates with the negatively charged Asp307 3.12 Å away. This interaction most likely positions the substrate in the most preferred orientation for hydrolysis. It is also plausible, that bases that contain a hydroxyl group at the C6/C4 position such as guanine, thymine and uracil interact with a an Arg211 (~ 4 Å) that would flip the base 180 ° leaving the nucleotide in an orientation that is not optimal for hydrolysis.

DbLnP has the highest catalytic efficiency for the cytosine and thymidine. These are both pyrimidines but differ in the group on C4 of the base. Comparison of the active site of 7WC and DbLnP can identify key residues that explain

differences between these enzymes. Firstly, DbLnP has a Glu307 which corresponds with Asp307 that is thought to have an impact on nucleotide specificity. The consequence of this substitution is that the nucleotide cannot enter as deep as in 7WC, thus the smaller pyrimidines would be preferred over the large purines. This theory is further reinforced with the K_M values that illustrated higher affinity for cytosine and thymine (pyrimidine) nucleotides than the adenosine and guanosine (purines).

Above it is discussed for 7WC that the groups on the nucleobase are important in substrate selection however such a discrepancy is not observed for DbLnP. Observation of the active site reveals an Arg304, which is substituted for a Leu304 in the 7WC active site and is orientated away from the active site. The Arg304 is positioned next to the Glu307, $\sim 2.5 \text{ \AA}$ from the nucleotide, in a way that the nucleotide could associate with either the Arg304 or the Glu307 without significantly disrupting the conformation of the nucleotide.

The DbLnP k_{cat}/K_M value for CTP of $1.4 \times 10^7 \text{ M}^{-1} \text{ sec}^{-1}$ is in the lower range of being considered 'super efficient' (Stroppolo, et al., 2001). This high ratio may be considered high however second order kinetics above 1.4×10^7 has also been observed in *R. novogicus* NTPDase1 for ADP and *L. pneumophila* with ADP and ATP (Sansom, et al., 2008; Zebisch, et al., 2007).

Further kinetic characterisation of these enzymes includes determining whether 7WC and DbLnP release ADP during catalysis. This would be achieved by separating reaction products over time by a C18 HPLC column and identifying the products by mass spectrometry. Also, to test my hypothesis for the selection of nucleotides in 7WC point mutation analysis is required, specifically it would be interesting to see if I could change the preference nucleotides, for example; a D307E may induce a preference for cytosine over adenosine and a D307N may reduce the specificity for different nucleotides.

7 General discussion

7.1 General background

There is increasing evidence that an extracellular purinogenic pathway is present in higher plants. Such a pathway increases the cytosolic calcium ions that lead to a diverse range of physiological responses including cell growth and division. Like many other pathways it is essential that the purinogenic pathway is controlled by regulating the signal. Similar to that of mammalian purinogenic transduction pathway, the extracellular NTPDases are prime candidates to regulate the signalling molecules. These enzymes catalyse the hydrolysis of the γ - and β - phosphoanhydride bond from a range of nucleotide di- or triphosphates in a divalent cation dependent manner.

The NTPDases in legumes are of particular interest due to the link with the early stages of rhizobium symbiosis during the nodulation and the nitrogen fixing process. The projects aims were to investigate the biochemical and structural characterisation of two NTPDases of legume origin. The first was 7WC, which was isolated from the roots of white clover the second was DbLnP, from the roots of *Dolichos biflorus* which has been characterised as a carbohydrate binding NTPDase that is directly associated with the perception of rhizobium signalling molecules.

7.2 Expression, refolding and crystallisation

For biochemical and structural characterisation of these NTPDases they were expressed in a prokaryotic expression system. This was performed to produce large amounts of protein efficiently that could be easily purified using metal affinity chromatography. Unfortunately, this expression system resulted in the expression of the protein as insoluble inactive aggregates or inclusion bodies. Inclusion bodies represent proteins that have been misfolded in the *E. coli* that may have arisen from over expression or incorrect folding conditions in the host's cytosol. Through solubilisation of inclusion bodies with high concentration

urea the NTPDases could be refolded to an active state. This step is often seen as a bottleneck as there are a number of contributing factors to correctly refold proteins. By using a high throughput *in vitro* refolding technique to determine the correct refolding parameters for each of the NTPDases, hundreds of conditions could be tested simultaneously and could be evaluated using either enzyme activity or protein aggregation.

The best refolding conditions for 7WC included a glutathione oxido-shuffling system which helped form correct disulfide bonds. PAGE analysis of the refolded 7WC showed a number of multimeric forms presumably caused by intra-molecular disulfide bonds. Analysis of each of the multimers showed that only the monomeric form was active and this could be enriched through acid buffer exchange, cation exchange and gel filtration chromatography. The top refolding condition of DbLnP included arginine which promotes refolding by inhibiting aggregation of structured intermediates. PAGE analysis and in-gel activity assays showed only one band with enzyme activity migrated into the gel. This active DbLnP was purified from aggregated protein via gel filtration chromatography.

The expression of both 7WC and DbLnP in *E.coli* and the ability to isolate pure, active protein via refolding and chromatography methods allowed me to perform kinetic characterisation and crystallisation trials. Crystallisation trials were performed on 7WC and DbLnP using conditions based on previous NTPDase crystallisation trials or using a high throughput method using liquid dispensing systems. Techniques such as seeding from other crystals proved valuable.

7.3 Structural elucidation

Using the recently solved rat NTPDase2 it was possible to determine the protein structures for 7WC. For successful structure determination it was necessary that a modified model that contained only conserved residues was made and then split into an N-terminal and C-terminal domain. Molecular replacement and

automated building methods used these two domains independently to solve 7WC. The DbLnP structure was solved via molecular replacement with the N- and C-terminal domains of 7WC. 7WC was chosen as a model as it was the structure that had the highest sequence homology with DbLnP. The necessity to solve the structure using each domain separately gave an early indication of differences in the trans-domain angle. Examination of the 7WC and DbLnP structure showed that they were both made up of two domains that are each reminiscent of a modified RNase-H fold that are characteristic of members of the Actin/hsp70 superfamily. The two domains are linked by a short linker region that positions the domains in a way that they form a cleft. This cleft is lined with the 5 conserved regions or apyrase conserved regions (ACRs) that are characteristic of NTPDases. The ACRs contain residues that are required in the binding and hydrolysis of substrate.

To better understand the binding and catalytic mechanism of NTPDases techniques such as co-crystallisation and substrate soaks were employed with 7WC and DbLnP to determine how the substrate and the product bound. The AMPPNP was chosen to soak into the crystals, as this molecule is analogous to ATP and is unable to be hydrolysed. A number of structures with various substrates were solved and summarised in Table 7-1.

Protein	Substrate	Metal ion	Conformation
7WC	apo	no	'open'
7WC	AMP+PO ₄ ⁻	no	'closed?'
7WC	AMPPNP+PO ₄ ⁻	no	'open'
DbLnP	3xPO ₄ ⁻	yes	'closed'
DbLnP	AMPPN+PO ₄ ⁻	yes	'closed'

Table 7-1 Table of structures determined in this project.

By soaking AMPPNP into crystals, it was found to bind to 7WC without the metal ion co-factor. It was orientated so that the AMPPNP bridges the two domains at the rim of the catalytic cleft whereby the nucleobase binds through π -

interactions with aromatic side chains on the C-terminal domain and the terminal phosphate associates with a phosphate binding site on the N-terminal domain in a similar fashion observed in the *Legionella pneumonia* NTPDase1. Such a conformation is unable to be hydrolysed as the catalytic residues are deep within the binding pocket.

Using co-crystallisation techniques, the non-hydrolysable AMPPNP the binding mechanism was resolved in DbLnP. Examination of the active site revealed a binding mechanism of the substrate immediately after hydrolysis. Where the hydrolysed products AMPPN and a phosphate ion associated with a metal ion cofactor was observed within the active site. Analogous to the position of the γ -phosphate of AMPPNP in the rat NTPDase2 structure, the phosphate ion is deep within the binding pocket bound to the C-terminal domain phosphate binding site that positions the phosphate in close proximity to the catalytic base. The phosphates of AMPPN are loosely associated with the N-terminal domain phosphate site. Compared with 7WC, the side-chains of key residues involved in substrate binding have a dramatic change in their position. As a result, pulls the product away from the active site. This movement could represent early local conformational changes that occur after hydrolysis that are required for expulsion of the product.

The metal ion was determined using X-ray fluorescence scanning to be a manganese ion present in the DbLnP+ PO₄ structure (Figure 5-14). Consistent with the rat NTPDase2, the phosphate bound DbLnP showed an octahedral coordination with 5 waters and an oxygen from a phosphate ion bound in the C-terminal domain phosphate binding site. The hydrolysed AMPPNP structure showed that the coordination of the metal ion shifts from a 6 to 5 coordination that may show the state of the metal after hydrolysis. Although both orientations are common, due to criticism over the resolution of the structure, this finding is most likely irrelevant.

One of the most important features of the catalytic mechanism identified in the rat NTPDase2 is the presence of a water molecule deep in the binding pocket that is believed to be directly involved in the nucleophilic attack of the terminal

phosphate phosphoanhydride bond. The nucleophilic water is present in the DbLnP structure with phosphates, but not in the hydrolysed AMPPNP structure giving further evidence that hydrolysis had occurred.

Due to the domain movement observed in actin and other proteins within this superfamily it has been hypothesised that upon binding of the substrate, the NTPDases go about a conformational change. By comparing the different binding states of 7WC and DbLnP through structural similarity matching, it was revealed that these NTPDases exist in two conformations. The NTPDases are either in an 'open' conformation or a 'closed' conformation. Results from this research suggest that the transition from 'open' and 'closed' is induced by the binding of the metal ion co-factor and the substrate. Binding of the substrate alone does not induce a conformational change; rather it binds in a non-catalytically active binding site (as shown in the 7WC/AMPPNP structure). The addition of the metal ion presumably encourages the substrate to shift down into the cleft towards the catalytic residues where hydrolysis can occur (shown in the DbLnP/AMPPNP structure). It is postulated that the domain movement expels the product and the metal ion, leaving the enzyme in the substrate receptive 'open' state.

Using the results of AMPPNP binding to 7WC without metal bound, results obtained using DbLnP and previously published *R. novogicus* NTPDase2 structure, the following schematic of the catalytic mechanism can be produced (Figure 7-1).

With reference to (Figure 7-1); The NTPDase is present in an apo state that contains no metal ion or substrate (I) that is receptive to substrate binding. The substrate binds to the NTPDase in a non-catalytically active site (II) and that the binding of a metal ion to the NTPDases induces a conformational change from an 'open' to a 'closed' form that creates a catalytically active binding site that the substrate moves into (III). Hydrolysis occurs through nucleophilic attack of the terminal phosphate (IV) and the product of the reaction moves from a stretched to a bent conformation facilitated by the movement of active site residues away

from the active site (V). Domain movement, from a 'closed' to an 'open' conformation expels the product from the enzyme (possibly is facilitated by the metal ion being released, although this has not been seen) (VI). The NTPDase is left in the substrate receptive apo-form for subsequent substrate binding (I).

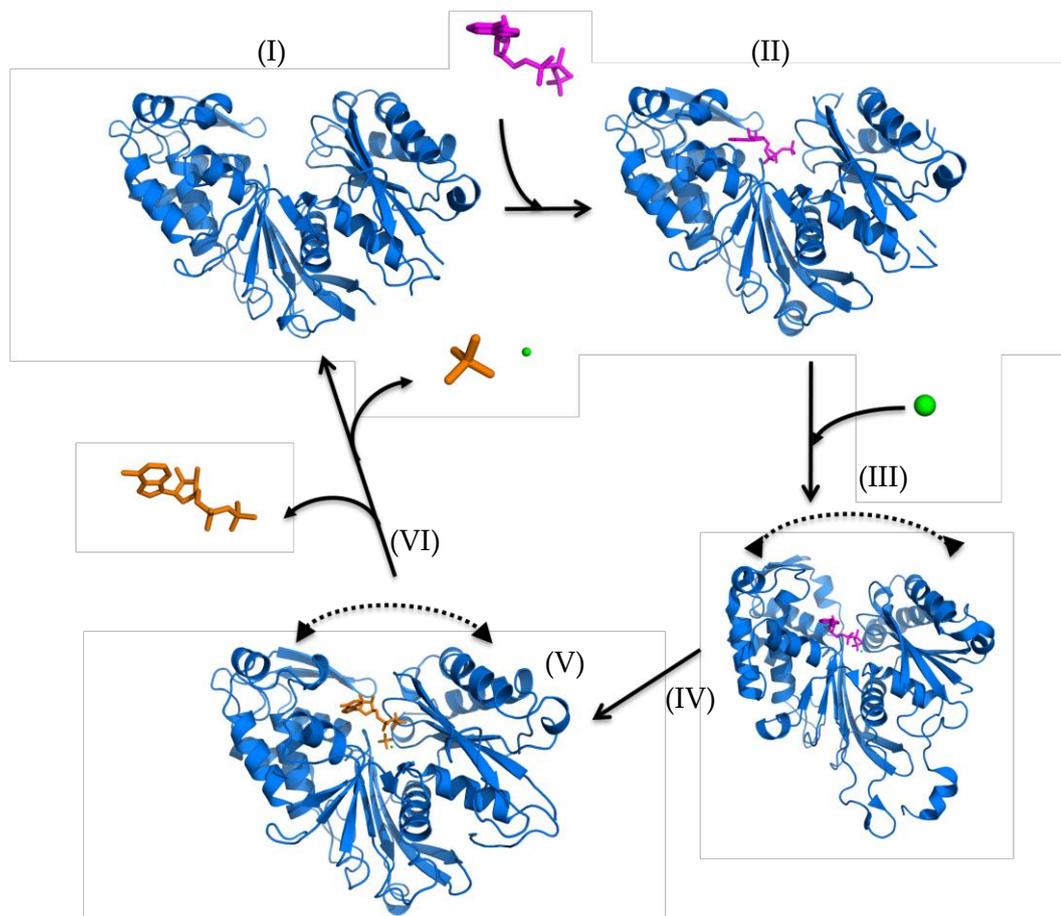


Figure 7-1 Schematic of the binding and catalytic mechanism of NTPDases. See text for details. (I) and (II) are 7WC, (III) is rat NTPDase2 and (V) is DbLnP. ATP, ADP and divalent metal ion are magenta, orange and green, respectively.

7.4 Kinetic characterisation

The NTPDases catalyse the hydrolysis of various nucleotides di- and tri-phosphates. The specificity and the rate of catalysis differ for each substrate. By determining the kinetic properties for both 7WC and DbLnP and comparing them with the solved structures it was possible to identify structural elements that confer substrate specificity.

Catalytic characterisation, such as Michaelis-Menten kinetics, pH profile, metal cofactor requirement and substrate specificity was performed for 8 different nucleotides. Both 7WC and DbLnP could catalyse all the tested substrates efficiently due to the π -interactions with aromatic side chains. Probing into the Michaelis-Menten kinetics data revealed that particular substrates had higher affinity and catalytic rates. These preferences were attributed to the size and decorations of the nucleobase and the residues within the nucleobase binding pocket. Despite the structural information, no elements could be attributed to explain the NDPase:NTPase ratio. Structural research into NTPDases that have a strong preference for one over the other, such as 6RG previously characterised by Chung Hong Chen, would give a clearer depiction of structural elements that contribute to differences in NDPase and NTPase activity (Chen, 2008). In general, the affinities for nucleotides are within the higher range of extracellular nucleotide concentrations in plants. Which suggests that efficient hydrolysis only occurs during transient increases in extracellular nucleotides. This seems logical as nucleotides play a number of essential roles within the plant and that an NTPDase with high affinity for nucleotides would no doubt be detrimental to many metabolic pathways.

Results from these analyses showed that the NTPDases are activated with calcium concentrations that are within the physiological range. This may be a mode of regulating NTPDase activity. Also, due to the relationship of NTPDases with calcium and calmodulin, the proposed role of NTPDases as a regulatory factor in the calcium/calmodulin-dependent kinase pathway during perception of NOD factors provides an intriguing area for future research.

7.5 Future research

Research presented in this thesis provides a platform for a number of research directions. Further kinetic characterisation addressing whether NDP is released from the enzyme using HPLC and mass spectrometry would give an interesting insight into the catalytic mechanism of these NTPDases. Also, point mutation

analysis on the residues that are hypothesised to confer substrate specificity and structural characterisation of NTPDases with different kinetic profiles could lead to engineering of substrate specific NTPDases.

7.6 Conclusion

This project investigated two NTPDases from legumes. The aims were to produce these enzymes in a prokaryotic expression system, perform a kinetic characterisation and to determine the structures. The first aim was essential before further research could be conducted and after the development of refolding protocols crystallisation and kinetic trials were achieved. The structures solved were of 7WC and DbLnP with various substrates in the 'open' and 'closed' conformations. This is the first report that a conformation change does occur during substrate binding in NTPDases. This collection of structural data complemented the kinetic data to give a comprehensive understanding on the binding mechanism of substrate to these NTPDases. This body of work will contribute to the emerging area of purinogenic signalling during nodulation in legumes and higher plants in general. Also, this research could be applied to human NTPDases to develop therapeutics for stroke and transplant patients.

8 Appendices

8.1 Nucleotide and protein sequences

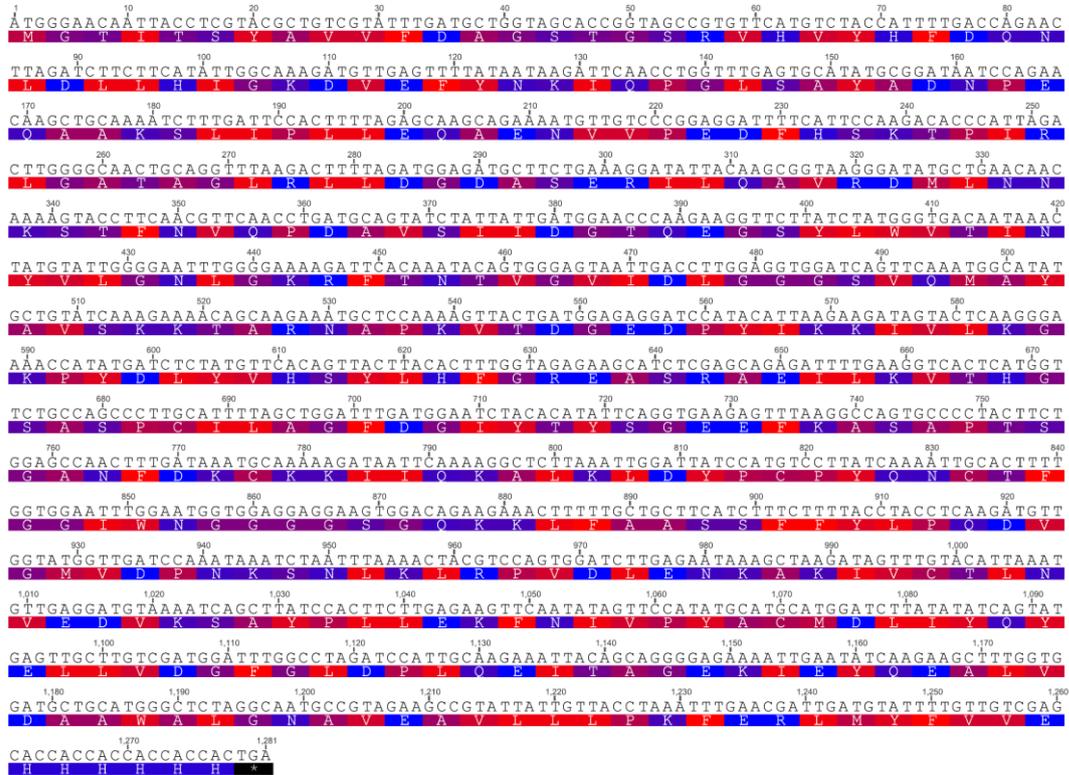


Figure 8-1 Gene and protein sequence of the 7WC expression construct. Protein sequence is highlighted based on hydrophobicity. The gradient from red to blue represent hydrophobic to hydrophilic.

1
 ATGGGCAGCAGCCATCATCA TCATCATCACAGCAGCGGCC TGGTGCCGCGCGGCAGCCATATGTCCATGGAAC TGTGACCCAGC
 M G S S H H H H H S S G L V P R G S H M S M B L L L S
 90
 TATGCGGTGATTTTGTATGCCGGCAGCAGCGGCAGCCG TGTTCATGTGTTTAACTTCGATCAGAACTGGATCTGCTGCA TATT
 Y A V I P D A G S S G S R V H V F N P D Q N L L L H I
 170
 GGCAACGATCTGGAA TTTACCAAAAAA TCAAAACCGGCGCTGAGCAGCTACGCCGATAAAACCGGAAAAAGCGGGAAAGCCTG
 G N D L E E F F K K I K P G L S S Y A D K F E K A A E S L
 250
 ATTCCGCTGCTGGAAGAAGCGGAGGACGTTGTTCCGGAAGA ACTGCA TCCGAAAACCCCGCTGAAACTGGGTGCCACCGCGCGT
 I P L L E E A E D V V P E E L H P K I P L K L L G A T A G
 330
 CTGCGTCTGCTGGATGGCGATGCCCGCAAAAAA TTTGCAAGCGGTTCTGTGAAA TGTTCGTAACCGTAGCAGCCTGAGCGTT
 L R L L D G D A A E K I L Q A V R E M F R N R S S L S V
 410
 CAGCCGGATGCCGTGAGCGTTATTGATGGCACCCAAGAAGGCAGCTA TCTGTGGGTGACCGTGAAC TATCTGCTGGGCAAACTG
 Q P L L E S V I D V V P E E L H P K I P L K L L G A T A G
 490
 GGTAAAAAA TTTACCAAAAAACCGTGGCGTGATGATCTG GCGCGGTGCCAGCGTTCAGATGGCGTATGCCG TGTCTAGAAAACCC
 G K K F T K T V G V I D L G G A S V O M A Y A V S R N T
 570
 GCCAAAAACCGCCGAAACCGCCGAGGGCGAAGATCCGTA TATGAAAAAACTGGTGCTGAAAGGCAAAAAA TACGATCTGT TAC
 A K N A P P O P Q P Y M K K L V L K G K Y D L Y
 650
 GTTCA TAGCTATCTGCGTTATG GCAACGATGCCGCCCGTGTGAAAA TTTTAAAAACCCAGGATGG TGCCGCCAGCCGCTGTCTG
 V H S Y L R Y G N D A A R V K I F K T T D G A A S P C L
 730
 CTGCGGGGCTATGAAAGATATTTATCGCTATAGCGGCGAAAGC TATAACATTTATGGCCCGACCA GCGCGCCAACTTTAACGAA
 L A G Y E D F Y R Y S G E S Y N I Y G P T S G A N F N E
 810
 TGCCGTGATCTGGCGCTGC AAAATTC TCGCTCTGAACGAACCGTGTAGCCATGAAAA TGTACCTTTGGCGGCATTTGGGATGGC
 C R D L A L O I L R L N E P C S H E N C T F G G I W D G
 890
 GGCAAAAGGCAGCGCCAGAAAAACCTGG TGGTGACCAAGCGCC TTTTATATCGTAGCAGCGAAGTGGGCTTTGTGACCCCGCCG
 G K G S G Q K N L V V T S A F Y Y R S S E V G F V T P P
 970
 AACAGCAAAAACCGTCCGCTGGATTTTGA AAAACCGCCGCAAAACAGGCGTGTAGCCCTGACCTTTGAAGAAGCGAAAAAGCACCTTT
 N S K N R P L D F E T A A K O A C S L T F E E A K S T F
 1,050
 CCGAACGTTGAAAAAGATAAAC TGCCGTTTCGTGTGTGTTGGATTTTCACTATCAGTATACCCCTGCTGGTTGATGGCTTTGGCC TG
 P N V E K D K L P F V C V D F T Y Q Y T L L V D G F G L
 1,130
 GACCCGGAACAGGAAATTACCGTTGCCGAAGGCATTGAA TATCAGGATGCCATTGTGGAAAACCGCCTGGCGCTGGGCACCGCC
 D P E Q E I T V A E G I E Y Q D A I V E T A W P L G T A
 1,210
 ATTTGAAGCGATTTTCAAGCTTGGCGAAATTTAACCGCC TGA TGTACTTCA TCTAATGA
 I E A I S S L P K F N R L M Y F I * *

Figure 8-2 Gene and protein sequence of the DbLnP expression construct. Protein sequence is highlighted based on hydrophobicity. The gradient from red to blue represent hydrophobic to hydrophilic.

8.2 Refolding screens

	A	B	C	D	E	F	G	H
Buffer	NaAc	NaAc	MES	MES	Tris	Tris	Tris	Tris
pH	5.5	6	6.5	7	7.5	8	8.5	9
1	arg.							
2	glut							
3	NaCl							
4	PEG							
5	arg, glut							
6	arg,NaCl							
7	arg,PEG							
8	glut, NaCl							
9	glut, PEG							
10	NaCl, PEG							
11	Buffer only							
12	NaCl, arg, glut, PEG							

Table 8-1 7WC refolding screen.

Abbreviations of the contents: NaAc; 50 mM acetate pH 5.0, 5.5. MES; 50 mM MES pH 6.5, 7.0. Tris; 50 mM Tris-HCl pH 7.5, 8.0, 8.5, 9.0. arg; 50mM L-arginine. glut; 5 mM:1 mM reduced:oxidised glutathione. NaCl; 150 mM NaCl. PEG; 0.2%(w/v) PEG 3350.

A	B	C	D	E	F	G	H
NaAc	NaAc	MES	MES	Tris	Tris	Tris	Tris
5.5	5.5	6	6	6.5	6.5	7	7
DTT, glycerol	GSH/GSSH, metals	DTT, metals	GSH/GSSH, glycerol	DTT, glycerol	EDTA, GSH/GSSH, urea	DTT, metals	GSH/GSSH, PEG ₃₃₅₀
DTT, L-arginine	EDTA, GSH/GSSH, L-arginine	EDTA, GSH/GSSH, L-arginine	cyclodextrin, DTT, glycerol	cyclodextrin, DTT, metals	cyclodextrin, GSH/GSSH, glycerol	cyclodextrin, DTT, urea	cyclodextrin, GSH/GSSH, EDTA
cyclodextrin, DTT	100 mM NaCl, GSH/GSSH, EDTA	100 mM NaCl, DTT, PEG ₃₃₅₀	100 mM NaCl, GSH/GSSH	100 mM NaCl, DTT	100 mM NaCl, GSH/GSSH, PEG ₃₃₅₀	100 mM NaCl, DTT, glycerol	100 mM NaCl, EDTA, GSH/GSSH, urea
cyclodextrin, DTT, urea	100 mM NaCl, cyclodextrin, GSH/GSSH, urea	100 mM NaCl, DTT, urea	100 mM NaCl, cyclodextrin, GSH/GSSH	100 mM NaCl, cyclodextrin, DTT	100 mM NaCl, cyclodextrin, GSH/GSSH, glycerol	100 mM NaCl, cyclodextrin, DTT, L-arginine	100 mM NaCl, cyclodextrin, GSH/GSSH, metals
100 mM NaCl, DTT, PEG ₃₃₅₀	250 mM NaCl, GSH/GSSH, PEG ₃₃₅₀	100 mM NaCl, cyclodextrin, DTT, metals	250 mM NaCl, GSH/GSSH, EDTA	150 mM NaCl, cyclodextrin, DTT, urea	250 mM NaCl, GSH/GSSH, EDTA	100 mM NaCl, cyclodextrin, DTT, PEG ₃₃₅₀	100 mM NaCl, cyclodextrin, GSH/GSSH, EDTA
100 mM NaCl, cyclodextrin, DTT	250 mM NaCl, cyclodextrin, GSH/GSSH, glycerol	250 mM NaCl, DTT, PEG ₃₃₅₀	250 mM NaCl, cyclodextrin, GSH/GSSH, urea	250 mM NaCl, DTT, urea	250 mM NaCl, EDTA, GSH/GSSH, PEG ₃₃₅₀	250 mM NaCl, DTT	250 mM NaCl, GSH/GSSH, glycerol
250 mM NaCl, DTT, metals	250 mM NaCl, cyclodextrin, GSH/GSSH, EDTA	250 mM NaCl, cyclodextrin, DTT, L-arginine	250 mM NaCl, cyclodextrin, GSH/GSSH, metals	250 mM NaCl, cyclodextrin, DTT	250 mM NaCl, cyclodextrin, GSH/GSSH, EDTA	250 mM NaCl, cyclodextrin, DTT	250 mM NaCl, cyclodextrin, GSH/GSSH, PEG ₃₃₅₀
L-arginine, EDTA	250 mM NaCl, cyclodextrin	cyclodextrin, L-arginine	250 mM NaCl, metals	L-arginine	250 mM NaCl, L-arginine	glycerol	250 mM NaCl
100 mM NaCl, PEG ₃₃₅₀	250 mM NaCl, cyclodextrin, urea	cyclodextrin, EDTA	250 mM NaCl, glycerol	cyclodextrin, EDTA	250 mM NaCl, metals	L-Arginine, EDTA	250 mM NaCl, L-arginine
100 mM NaCl, glycerol	250 mM NaCl, cyclodextrin, L-arginine	100 mM NaCl, urea	250 mM NaCl, cyclodextrin, PEG ₃₃₅₀	cyclodextrin, urea	250 mM NaCl, cyclodextrin, glycerol	cyclodextrin, metals	250 mM NaCl, cyclodextrin, metals
100 mM NaCl, cyclodextrin, metals	Buffer only	100 mM NaCl, EDTA	Buffer only	100 mM NaCl, metals	Buffer only	100 mM NaCl	Buffer only
250 mM NaCl	MilliQ	100 mM NaCl, cyclodextrin, glycerol	MilliQ	100 mM NaCl, cyclodextrin, PEG ₃₃₅₀	Empty	100 mM NaCl, cyclodextrin, urea	Empty

Table 8-2 Refolding plate for DbLnP.

Components concentrations: 50 mM acetate pH 5.0, 5.5 (NaAc), 50 mM MES pH 6.0, 50 mM Tris-HCl pH 6.5, 7.0, 50mM L-arginine, 5 mM:1 mM reduced:oxidised glutathione (GSH/GSSH), 100 or 250 mM NaCl, 0.2%(w/v) PEG 3350, 12.5 % cyclodextrin, 1mM DTT, 600 mM urea, 1 mM CaCl₂/MgCl₂ (metals), 20 % glycerol, 1 mM EDTA.

References

- Abbracchio, M. P., Burnstock, G., Boeynaems, J. M., Barnard, E. A., Boyer, J. L., Kennedy, C., Knight, G. E., Fumagalli, M., Gachet, C., Jacobson, K. A., & Weisman, G. A. (2006). International union of pharmacology LVIII: Update on the P2Y G protein-coupled nucleotide receptors: From molecular mechanisms and pathophysiology to therapy. *Pharmacological Reviews*, 58(3), 281-341.
- Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., & Zwart, P. H. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D-Biological Crystallography*, 66, 213-221.
- Afonine, P. V., Grosse-Kunstleve, R. W., & Adams, P. D. (2005). The Phenix refinement framework. *CCP4 Newsletter*, 42, 1-7.
- Arakawa, T., Ejima, D., Tsumoto, K., Obeyama, N., Tanaka, Y., Kita, Y., & Timasheff, S. N. (2007). Suppression of protein interactions by arginine: A proposed mechanism of the arginine effects. *Biophysical Chemistry*, 127(1-2), 1-8.
- Arakawa, T., & Tsumoto, K. (2003). The effects of arginine on refolding of aggregated proteins: not facilitate refolding, but suppress aggregation. *Biochemical and Biophysical Research Communications*, 304(1), 148-152.
- Aravind, L., & Koonin, E. V. (1999). Gleaning non-trivial structural, functional and evolutionary information about proteins by iterative database searches. *Journal of Molecular Biology*, 287, 1023-1040.
- Asai, T., Osullivan, W. J., & Tatibana, M. (1983). A potent nucleoside triphosphate hydrolase from the parasitic protozoan *Toxoplasma gondii*- Purification, some properties, and activation by thiol compounds. *Journal of Biological Chemistry*, 258(11), 6816-6822.
- Bailey, S. (1994). The CCP4 Suite - Programs for protein crystallography. *Acta Crystallographica Section D-Biological Crystallography*, 50, 760-763.
- Baykoz, A. A., Evtushenko, O. A., & Awaeva, S. M. (1988). A malachite green procedure for orthophosphate determination and its use in alkaline phosphatase-based enzyme immunoassay. *Analytical Biochemistry*, 171, 266-270.
- Baynes, B. M., Wang, D. I. C., & Trout, B. L. (2005). Role of arginine in the stabilization of proteins against aggregation. *Biochemistry*, 44(12), 4919-4925.
- Birnboim, H. C., & Doly, J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Research*, 7(6), 1513-1523.
- Bork, P., Sander, C., & Valencia, A. (1992). An ATPDase domain common to prokaryotic cell-cycle proteins, sugar kinases, actin, and HSP70 heat-

-
- shock proteins. *Proceedings of the National Academy of Sciences of the United States of America*, 89(16), 7290-7294.
- Burnstock, G. (2007). Physiology and pathophysiology of purinergic neurotransmission. *Physiological Reviews*, 87(2), 659-797.
- Caldwell, C. C., Hornyak, S. C., Pendleton, E., Campbell, D., & Knowles, A. F. (2001). Regulation of chicken gizzard ecto-ATPase activity by modulators that affect its oligomerization status. *Archives of Biochemistry and Biophysics*, 387(1), 107-116.
- Chen, C. (2008). *Structure function studies on lectin nucleotide phosphohydrolases (LNPs)*. University of Auckland.
- Chen, V. B., Davis, I. W., & Richardson, D. C. (2009). KiNG (Kinemage, Next Generation): A versatile interactive molecular and scientific visualization program. *Protein Science*, 18(11), 2403-2409.
- Chen, W., & Guidotti, G. (2001). Soluble apyrases release ADP during ATP hydrolysis. *Biochemical and Biophysical Research Communications*, 282(1), 90-95.
- Chen, Y. R., Datta, N., & Roux, S. J. (1987). Purification and partial characterization of a Calmodulin-Stimulated Nucleoside Triphosphatase from Pea Nuclei. *Journal of Biological Chemistry*, 262(22), 10689-10694.
- Chiang, W. C., & Knowles, A. F. (2008). Transmembrane domain interactions affect the stability of the extracellular domain of the human NTPDase 2. *Archives of Biochemistry and Biophysics*, 472(2), 89-99.
- Chivasa, S., Ndimba, B. K., Simon, W. J., Lindsey, K., & Slabas, A. R. (2005). Extracellular ATP functions as an endogenous external metabolite regulating plant cell viability. *Plant Cell*, 17(11), 3019-3034.
- Chow, M. K. M., Amin, A. A., Fulton, K. F., Whisstock, J. C., Buckle, A. M., & Bottomley, S. P. (2006). REFOLD: An analytical database of protein refolding methods. *Protein Expression and Purification*, 46(1), 166-171.
- Clark, E. D., Schwarz, E., & Rudolph, R. (1999). Inhibition of aggregation side reactions during in vitro protein folding. *Amyloid, Prions, and Other Protein Aggregates* (Vol. 309, pp. 217-236).
- Clark, G., & Roux, S. J. (2009). Extracellular nucleotides: Ancient signaling molecules. *Plant Science*, 177(4), 239-244.
- Clement, J. M., & Kent, C. (1999). CTP : phosphocholine cytidyltransferase: Insights into regulatory mechanisms and novel functions. *Biochemical and Biophysical Research Communications*, 257(3), 643-650.
- Cohn, M., & Meek, G. A. (1956). The mechanism of hydrolysis of adenosine di- and tri-phosphate catalysed by potato apyrase. *Biochemistry*, 66, 1957-1957.
- Colgan, S. P. (2006). Physiological roles for ecto-5' -nucleotidase (CD73). *Purinergic Signalling*, 2(351-360).
- Corpas, F. J., Barroso, J. B., Carreras, A., Quiros, M., Leon, A. M., Romero-Puertas, M. C., Esteban, F. J., Valderrama, R., Palma, J. M., Sandalio, L. M., Gomez, M., & del Rio, L. A. (2004). Cellular and subcellular localization of endogenous nitric oxide in young and senescent pea plants. *Plant Physiology*, 136(1), 2722-2733.

-
- Davis, I. W., Leaver-Fay, A., Chen, V. B., Block, J. N., Kapral, G. J., Wang, X., Murray, L. W., Arendall, W. B., Snoeyink, J., Richardson, J. S., & Richardson, D. C. (2007). MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Research*, *35*, W375-W383.
- Day, R. B., McAlvin, C. B., Loh, J. T., Denny, R. L., Wood, T. C., Young, N. D., & Stacey, G. (2000). Differential expression of two soybean apyrases, one of which is an early nodulin. *Molecular Plant-Microbe Interactions*, *13*(10), 1053-1070.
- de Jesus, J. B., Pinheiro, A. A. D., Lopes, A., & Meyer-Fernandes, J. R. (2002). An ectonucleotide ATP-diphosphohydrolase activity in *Trichomonas vaginalis* stimulated by galactose and its possible role in virulence. *Zeitschrift Fur Naturforschung C-a Journal of Biosciences*, *57*(9-10), 890-896.
- de Koning, H., & Diallinas, G. (2000). Nucleobase transporters. *Molecular Membrane Biology*, *17*(2), 75-94.
- DeLano, W. L. (2002). The PyMOL Molecular Graphics System. *DeLano Scientific*.
- DeMarco, R., Kowaltowski, A. T., Mortara, R. A., & Verjovski-Almeida, S. (2003). Molecular characterization and immunolocalization of *Schistosoma mansoni* ATP-diphosphohydrolase. *Biochemical and Biophysical Research Communications*, *307*(4), 831-838.
- Demidchik, V., Nichols, C., Oliynyk, M., Dark, A., Glover, B. J., & Davies, J. M. (2003). Is ATP a signaling agent in plants? *Plant Physiology*, *133*(2), 456-461.
- Demidchik, V., Shang, Z. L., Shin, R., Thompson, E., Rubio, L., Laohavisit, A., Mortimer, J. C., Chivasa, S., Slabas, A. R., Glover, B. J., Schachtman, D. P., Shabala, S. N., & Davies, J. M. (2009). Plant extracellular ATP signalling by plasma membrane NADPH oxidase and Ca²⁺ channels. *Plant Journal*, *58*(6), 903-913.
- Drosopoulos, J. H. F. (2002). Roles of Asp54 and Asp213 in Ca²⁺ utilization by soluble human CD39/ecto-nucleotidase. *Archives of Biochemistry and Biophysics*, *406*(1), 85-95.
- Drosopoulos, J. H. F., Broekman, M. J., Islam, N., Maliszewski, C. R., Gayle, R. B., & Marcus, A. J. (2000). Site-directed mutagenesis of human endothelial cell ecto-ADPase/soluble CD39: Requirement of glutamate 174 and serine 218 for enzyme activity and inhibition of platelet recruitment. *Biochemistry*, *39*(23), 6936-6943.
- Dwyer, K. M., Robson, S. C., Nandurkar, H. H., Campbell, D. J., Gock, H., Murray-Segal, L. J., Fisicaro, N., Mysore, T. B., Kaczmarek, E., Cowan, P. J., & d'Apice, A. J. F. (2004). Thromboregulatory manifestations in human CD39 transgenic mice and the implications for thrombotic disease and transplantation. *Journal of Clinical Investigation*, *113*(10), 1440-1446.
- Ehrhardt, D. W., Wais, R., & Long, S. R. (1996). Calcium spiking in plant root hairs responding to Rhizobium nodulation signals. *Cell*, *85*(5), 673-681.
- Emsley, P., & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallographica Section D-Biological Crystallography*, *60*, 2126-2132.

-
- Enyoji, K., Sevigny, J., Lin, Y., Frenette, P. S., Christie, P. D., Esch, J. S. A., Imai, M., Edelberg, J. M., Rayburn, H., Lech, M., Beeler, D. L., Csizmadia, E., Wagner, D. D., Robson, S. C., & Rosenberg, R. D. (1999). Targeted disruption of CD39/ATP diphosphohydrolase results in disordered haemostasis and thromboregulation. *Nature Medicine*, 5(9), 1010-1017.
- Etzler, M. E., Kalsi, G., Ewing, N. N., Roberts, N. J., Day, R. B., & Murphy, J. B. (1999). A nod factor binding lectin with apyrase activity from legume roots. *Proceedings of the National Academy of Sciences of the United States of America*, 96(10), 5856-5861.
- Failer, B. U., Aschrafi, A., Schmalzing, G., & Zimmermann, H. (2003). Determination of native oligomeric state and substrate specificity of rat NTPDase1 and NTPDase2 after heterologous expression in *Xenopus* oocytes. *European Journal of Biochemistry*, 270(8), 1802-1809.
- Fietto, J. L. R., DeMarco, R., Nascimento, I. P., Castro, I. M., Carvalho, T. M. U., de Souza, W., Bahia, M. T., Alves, M. J. M., & Verjovski-Almeida, S. (2004). Characterization and immunolocalization of an NTP diphosphohydrolase of *Trypanosoma cruzi*. *Biochemical and Biophysical Research Communications*, 316(2), 454-460.
- Fischer, B., Sumner, I., & Goodenough, P. (1993). Isolation, renaturation, and formation of disulfide bonds of eukaryotic proteins expressed in *Escherichia coli* as inclusion-bodies *Biotechnology and Bioengineering*, 41(1), 3-13.
- Flaherty, K. M., Wilbanks, S. M., Delucaflaherty, C., & McKay, D. B. (1994). Structural basis of the 70-KiloDalton Heat-Shock cognate protein ATP hydrolytic activity. 2. Structure of the active-site with ADP or ATP bound to Wild-type and mutant ATPase fragment. *Journal of Biological Chemistry*, 269(17), 12899-12907.
- Fountain, S. J., Cao, L. S., Young, M. T., & North, R. A. (2008). Permeation properties of a P2X receptor in the green algae *Ostreococcus tauri*. *Journal of Biological Chemistry*, 283(22), 15122-15126.
- Frieden, C. (1983). Polymerization of actin-Mechanism of the Mg²⁺-induced process at pH 8 and 20 degrees C. *Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences*, 80(21), 6513-6517.
- Frieden, C., & Patane, K. (1988). Mechanism for nucleotide exchange in monomeric actin. *Biochemistry*, 27, 3812-3820.
- Gaddie, K. J., & Kirley, T. L. (2010). Proline residues link the active site to transmembrane domain movements in human nucleoside triphosphate diphosphohydrolase 3 (NTPDase3). *Purinergic Signalling*, 6(3), 327-337.
- Gayle, R. B., Maliszewski, C. R., Gimpel, S. D., Schoenborn, M. A., Caspary, R. G., Richards, C., Brasel, K., Price, V., Drosopoulos, J. H. F., Islam, N., Alyonycheva, T. N., Broekman, M. J., & Marcus, A. J. (1998). Inhibition of platelet function by recombinant soluble ecto-ADPase/CD39. *Journal of Clinical Investigation*, 101(9), 1851-1859.
- Geigenberger, P., Riewe, D., & Fernie, A. R. (2009). The central regulation of plant physiology by adenylates. *Trends in Plant Science*, 15(2), 98-105.

-
- Goddette, D. W., & Frieden, C. (1985). The binding of Cytochalasin-D to monomeric Actin. *Biochemical and Biophysical Research Communications*, 128(3), 1087-1092.
- Gookin, T. E., Kim, J., & Assmann, S. M. (2008). Whole proteome identification of plant candidate G-protein coupled receptors in Arabidopsis, rice, and poplar: computational prediction and in-vivo protein coupling. *Genome Biology*, 9(7).
- Govindarajulu, M., Kim, S. Y., Libault, M., Berg, R. H., Tanaka, K., Stacey, G., & Taylor, C. G. (2009). GS52 Ecto-Apyrase Plays a Critical Role during Soybean Nodulation. *Plant Physiology*, 149(2), 994-1004.
- Gresser, M. J., Beharry, S., & Moennich, D. M. C. (1984). Inhibition of Mitochondrial F1-ATPase by adenylyl imidodiphosphate. *Current Topics in Cellular Regulation*, 24, 365-378.
- Grinthal, A., & Guidotti, G. (2000). Substitution of His59 converts CD39 apyrase into an ADPase in a quaternary structure dependent manner. *Biochemistry*, 39, 9-16.
- Grinthal, A., & Guidotti, G. (2002). Transmembrane domains confer different substrate specificities and adenosine diphosphate hydrolysis mechanisms on CD39, CD39L1, and chimeras. *Biochemistry*, 41(6), 1947-1956.
- Grinthal, A., & Guidotti, G. (2006). CD39, NTPDase 1, is attached to the plasma membrane by two transmembrane domains. Why? *Purinergic Signalling*, 2(2), 391-398.
- Grinthal, A., & Guidotti, G. (2007). Bilayer mechanical properties regulate the transmembrane helix mobility and enzymatic state of CD39. *Biochemistry*, 46(1), 279-290.
- Hamel, L. P., & Beaudoin, N. (2010). Chitooligosaccharide sensing and downstream signalling: contrasted outcomes in pathogenic and beneficial plant-microbe interactions. *Planta*, 232(4), 787-806.
- Handa, M., & Guidotti, G. (1996). Purification and cloning of a soluble ATP-diphosphohydrolase (apyrase) from potato tubers (*Solanum tuberosum*). *Biochemical and Biophysical Research Communications*, 218(3), 916-923.
- Harding, M. M. (2001). Geometry of metal-ligand interactions in proteins. *Acta Crystallographica Section D-Biological Crystallography*, 57, 401-411.
- Harding, M. M. (2004). Small revisions to predicted distances around metal sites in proteins. *Acta Crystallographica Section D-Biological Crystallography*, 62, 678-682.
- Hayward, S. (1999). Structural principles governing domain motions in proteins. *Proteins-Structure Function and Genetics*, 36(4), 425-435.
- Hayward, S., & Berendsen, H. J. C. (1998). Systematic analysis of domain motions in proteins from conformational change: New results on citrate synthase and T4 lysozyme. *Proteins-Structure Function and Genetics*, 30(2), 144-154.
- Hayward, S., & Lee, R. A. (2002). Improvements in the analysis of domain motions in proteins from conformational change: DynDom version 1.50. *Journal of Molecular Graphics & Modelling*, 21(3), 181-183.
- Holm, L., & Park, J. (2000). DaliLite workbench for protein structure comparison. *Bioinformatics*, 16(6), 566-567.

-
- Howlin, B., Butler, S. A., Moss, D. S., Harris, G. W., & Driessen, H. P. C. (1993). TLSANL - TLS Parameter-Analysis Program for Segmented Anisotropic Refinement of Macromolecular Structures. *Journal of Applied Crystallography*, 26, 622-624.
- Hurley, J. H. (1996). The sugar kinase heat shock protein 70 actin super family: Implications of conserved structure for mechanism. *Annual Review of Biophysics and Biomolecular Structure*, 25, 137-&.
- Iqbal, J., Vollmayer, P., Braun, N., Zimmermann, H., & Muller, C. E. (2006). A capillary electrophoresis method for the characterization of ecto-nucleoside triphosphate diphosphohydrolases (NTPDases) and the analysis of inhibitors by in-capillary enzymatic microreaction. *Purinergic Signaling*, 1(4), 349-358.
- Ivanenkov, V. V., Meller, J., & Kirley, T. L. (2005). Characterization of disulfide bonds in human nucleoside triphosphate diphosphohydrolase 3 (NTPDase3): Implications for NTPDase structural modeling. *Biochemistry*, 44(25), 8998-9012.
- Jaffe, M. J. (1973). Role of ATP in mechanically stimulated rapid closure of venus-flytrap. *Plant Physiology*, 51(1), 17-18.
- Javed, R., & Knowles, A. F. (2006). Mutagenesis of conserved lysine and arginine residues in Human Ecto-ATPase (E-NTPDase 2). *Faseb Journal*, 20(5), A901-A901.
- Javed, R., Yarimizu, K., Pelletier, N., Li, C., & Knowles, A. F. (2007). Mutagenesis of lysine 62, asparagine 64, and conserved region 1 reduces the activity of human ecto-ATPase (NTPDase 2). *Biochemistry*, 46(22), 6617-6627.
- Jeter, C., & Roux, S. (2006). Plant responses to extracellular nucleotides: Cellular processes and biological effects. *Purinergic signalling*, 2(3), 443-449.
- Jeter, C. R., Tang, W. Q., Henaff, E., Butterfield, T., & Roux, S. J. (2004). Evidence of a novel cell signalling role for extracellular adenosine triphosphates and diphosphates in Arabidopsis. *Plant Cell*, 16(10), 2652-2664.
- Kabsch, W., Mannherz, H. G., Suck, D., Pai, E. F., & Holmes, K. C. (1990). Atomic-Structure of the Actin-DNase-1 complex. *Nature*, 347(6288), 37-44.
- Kalsi, G., & Etzler, M. E. (2000). Localization of a nod factor-binding protein in legume roots and factors influencing its distribution and expression. *Plant Physiology*, 124(3), 1039-1048.
- Kansas, G. S., Wood, G. S., & Tedder, T. F. (1991). Expression, Distribution, and Biochemistry of human CD39-Role in activation-associated Homotypic adhesion of Lymphocytes. *Journal of Immunology*, 146(7), 2235-2244.
- Kawate, T., Michel, J. C., Birdsong, W. T., & Gouaux, E. (2009). Crystal structure of the ATP-gated P2X(4) ion channel in the closed state. *Nature*, 460(7255), 592-598.
- Kettlun, A. M., Espinosa, V., Garcia, L., & Valenzuela, M. A. (2005). Potato tuber isoapyrases: Substrate specificity, affinity labeling, and proteolytic susceptibility. *Phytochemistry*, 66(9), 975-982.
- Kettlun, A. M., Uribe, L., Calvo, V., Silva, S., Rivera, J., Mancilla, M., Valenzuela, M. A., & Traversocori, A. (1982). Properties of two apyrases from *Solanum tuberosum*. *Phytochemistry*, 21(3), 551-558.

-
- Kim, S. Y., Sivaguru, M., & Stacey, G. (2006). Extracellular ATP in plants. Visualization, localization, and analysis of physiological significance in growth and signalling. *Plant Physiology*, 142(3), 984-992.
- Kinosian, H., Selden, L. A., Estes, J. E., & Gershman, L. C. (1993). Nucleotide binding to actin. *The Journal of Biological Chemistry*, 268(12), 8683-8691.
- Kirley, T., Crawford, P., & Smith, T. (2006). The structure of the nucleoside triphosphate diphosphohydrolases (NTPDases) as revealed by mutagenic and computational modeling analyses. *Purinergic Signalling*, 2(2), 379-389.
- Kirley, T. L., Yang, F., & Ivanenkov, V. V. (2001). Site-directed mutagenesis of human nucleoside triphosphate diphosphohydrolase 3: The importance of conserved glycine residues and the identification of additional conserved protein motifs in eNTPDases. *Archives of Biochemistry and Biophysics*, 395(1), 94-102.
- Knowles, A. F. (2011). The GDA1_CD39 superfamily: NTPDases with diverse functions. *Purinergic Signalling*, 7, 21-45.
- Knowles, A. F., & Li, C. (2006). Molecular cloning and characterization of expressed human ecto-nucleoside triphosphate diphosphohydrolase 8 (E-NTPDase 8) and its soluble extracellular domain. *Biochemistry*, 45(23), 7323-7333.
- Koonin, E. V. (2010). New variants of known folds: do they bring new biology? *Acta Crystallographica Section F*, F66, 1226-1229.
- Kozakiewicz, A., Neumann, P., Banach, M., Komoszyński, M., & Wojtczak, A. (2008a). Modeling studies of potato nucleoside triphosphate diphosphohydrolase NTPDase1: an insight into the catalytic mechanism. *Acta Biochimica Polonica*, 55(1), 141-150.
- Kozakiewicz, A., Neumann, P., Banach, M., Komoszyński, M., & Wojtczak, A. (2008b). Modelling studies of potato nucleoside triphosphate diphosphohydrolase NTPDase1: an insight into the catalytic mechanism. *Acta Biochimica Polonica*, 55(1), 141-150.
- Krishnan, P. S. (1948). The preparation of apyrase from potato. *Archives of Biochemistry and Biophysics*, 16, 474-476.
- Krissinel, E., & Henrick, K. (2004). Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallographica Section D-Biological Crystallography*, 60, 2256-2268.
- Kristensen, O., Laurberg, M., Liljas, A., Kastrup, J. S., & Gajhede, M. (2004). Structural characterization of the stringent response related exopolyphosphatase/guanosine pentaphosphate phosphohydrolase protein family. *Biochemistry*, 43(28), 8894-8900.
- Kristensen, O., Ross, B., & Gajhede, M. (2008). Structure of the PPX/GPPA phosphatase from *Aquifex aeolicus* in complex with the alarmone ppGpp. *Journal of Molecular Biology*, 375(5), 1469-1476.
- Kukulski, F., & Komoszyński, M. (2003). Purification and characterization of NTPDase1 (ecto-apyrase) and NTPDase2 (ecto-ATPase) from porcine brain cortex synaptosomes. *European Journal of Biochemistry*, 270(16), 3447-3454.

-
- Kukulski, F., Lévesque, S. A., Lavoie, E. G., Lecka, J., Bigonnesse, F., Knowles, A. F., Robson, S. C., Kirley, T. L., & Sévigny, J. (2005). Comparative hydrolysis of P2 receptor agonists by NTPDases 1, 2, 3 and 8. *Purinergic Signaling*, 1(2), 193-204.
- Laliberte, J. F., & Beaudoin, A. R. (1983). Sequential hydrolysis of the gamma-phosphate and beta-phosphate groups of ATP by the ATP diphosphohydrolase from pig pancreas. *Biochimica et Biophysica Acta*, 742(1), 9-15.
- Langer, G., Cohen, S. X., Lamzin, V. S., & Perrakis, A. (2008). Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nature Protocols*, 3(7), 1171-1179.
- Lazarowski, E. R., Boucher, R. C., & Harden, T. K. (2003). Mechanisms of release of nucleotides and integration of their action as P2X- and P2Y-receptor activating molecules. *Molecular Pharmacology*, 64(4), 785-795.
- Leach, S. L., & Scheraga, H. A. (1960). Effect of light scattering on ultraviolet difference spectra. *Journal of the American chemical society*, 82(18), 4790-4792.
- Lee, R. A., Razaz, M., & Hayward, S. (2003). The DynDom database of protein domain motions. *Bioinformatics*, 19(10), 1290-1291.
- Leslie, A. G. W. (1992). Recent changes to the MOSFLM package for processing film and image plate data. *Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography*, 26.
- Levy, J., Bres, C., Geurts, R., Chalhoub, B., Kulikova, O., Duc, G., Journet, E. P., Ane, J. M., Lauber, E., Bisseling, T., Denarie, J., Rosenberg, C., & Debelle, F. (2004). A putative Ca²⁺ and calmodulin-dependent protein kinase required for bacterial and fungal symbioses. *Science*, 303(5662), 1361-1364.
- Li, C. S., Lee, Y., & Knowles, A. F. (2010). The Stability of Chicken Nucleoside Triphosphate Diphosphohydrolase 8 Requires Both of Its Transmembrane Domains. *Biochemistry*, 49(1), 134-146.
- Liebecq, C., Lallemand, A., & Degueldre-Guillaume, M. J. (1962). The apyrase activity of potato extract. *Archives of Biochemistry and Biophysics*, 97(609-610).
- Lovell, S. C., Davis, I. W., Adrendall, W. B., de Bakker, P. I. W., Word, J. M., Prisant, M. G., Richardson, J. S., & Richardson, D. C. (2003). Structure validation by C alpha geometry: phi, psi and C beta deviation. *Proteins-Structure Function and Genetics*, 50(3), 437-450.
- Maliszewski, C. R., Delespesse, G. J. T., Schoenborn, M. A., Armitage, R. J., Fanslow, W. C., Nakajima, T., Baker, E., Sutherland, G. R., Poindexter, K., Birks, C., Alpert, A., Friend, D., Gimpel, S. D., & Gayle, R. B. (1994). The CD39 Lymphoid-cell activation antigen-Molecular-cloning and structural characterisation. *Journal of Immunology*, 153(8), 3574-3583.
- Marcus, A. J., Broekman, M. J., Drosopoulos, J. H. F., Olson, K. E., Islam, N., Pinsky, D. J., & Levi, R. (2005). Role of CD39 (NTPDase-1) in thromboregulation, cerebroprotection, and cardioprotection. *Seminars in Thrombosis and Hemostasis*, 31(2), 234-246.
- Marcus, A. J., Broekman, M. J., Drosopoulos, J. H. F., Pinsky, D. J., Islam, N., Gayle, R. B., & Maliszewski, C. R. (2001). Thromboregulation by

-
- endothelial cells - Significance for occlusive vascular diseases. *Arteriosclerosis Thrombosis and Vascular Biology*, 21(2), 178-182.
- Mateo, J., Kreda, S., Henry, C. E., Harden, T. K., & Boyer, J. L. (2003). Requirement of Cys(399) for processing of the human ecto-ATPase (NTPDase2) and its implications for determination of the activities of splice variants of the enzyme. *Journal of Biological Chemistry*, 278(41), 39960-39968.
- Matthews, B. W. (1968). Solvent content of protein crystals *Journal of Molecular Biology*, 33, 491-497.
- Mayerhoff, O. (1945). The origin of the reaction of Harden and Young in cell-free alcoholic fermentation. *Journal of Biological Chemistry*, 157(105-119).
- McAlvin, C. B., & Stacey, G. (2005). Transgenic expression of the soybean apyrase in *Lotus japonicus* enhances nodulation. *Plant Physiology*, 137(4), 1456-1462.
- McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., & Read, R. J. (2007). Phaser crystallographic software. *Journal of Applied Crystallography*, 40, 658-674.
- Misawa, S., & Kumagai, I. (1999). Refolding of therapeutic proteins produced in *Escherichia coli* as inclusion bodies. *Biopolymers (Peptide science)*, 51, 297-307.
- Mitra, R. M., Gleason, C. A., Edwards, A., Hadfield, J., Downie, J. A., Oldroyd, G. E. D., & Long, S. R. (2004). A Ca²⁺/calmodulin-dependent protein kinase required for symbiotic nodule development: Gene identification by transcript-based cloning. *Proceedings of the National Academy of Sciences of the United States of America*, 101(13), 4701-4705.
- Moreland, N., Ashton, R., Baker, H. M., Ivanovic, I., Patterson, S., Arcus, V. L., Baker, E. N., & Lott, J. S. (2005). A flexible and economical medium-throughput strategy for protein production and crystallization. *Acta Crystallographica Section D-Biological Crystallography*, 61, 1378-1385.
- Moriyama, E. N., Strope, P. K., Opiyo, S. O., Chen, Z. Y., & Jones, A. M. (2006). Mining the *Arabidopsis thaliana* genome for highly-divergent seven transmembrane receptors. *Genome Biology*, 7(10), R96.
- Murphy-Piedmonte, D. M., Crawford, P. A., & Kirley, T. L. (2005). Bacterial expression, folding, purification and characterization of soluble NTPDase5 (CD39L4) ecto-nucleotidase. *Biochimica Et Biophysica Acta-Proteins and Proteomics*, 1747(2), 251-259.
- Murshudov, G. N., Vagin, A. A., & Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallographica Section D-Biological Crystallography*, 53, 240-255.
- Oldroyd, G. E. D., & Downie, J. A. (2006). Nuclear calcium changes at the core of symbiosis signalling. *Current Opinion in Plant Biology*, 9(4), 351-357.
- Pain, R. H. (1995). Overview of protein folding. *Current protocols in protein science*, 6(4), 1-7.
- Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF chimera - A visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, 25(13), 1605-1612.

-
- Plesner, L., Kirley, T. L., & Knowles, A. F. (Eds.). (1997). *Ecto-ATPases: Recent progress on Structure and Function*.
- Quinn, J. M., & Etzler, M. E. (1987). Isolation and characterization of a lectin from the roots of *Dolichos biflorus*. *Archives of Biochemistry and Biophysics*, 258(2), 535-544.
- Rangarajan, E. S., Nadeau, G., Li, Y. G., Wagner, J., Hung, M. N., Schrag, J. D., Cygler, M., & Matte, A. (2006). The structure of the exopolyphosphatase (PPX) from *Escherichia coli* O157 : H7 suggests a binding mode for long polyphosphate chains. *Journal of Molecular Biology*, 359(5), 1249-1260.
- Reichler, S. A., Torres, J., Rivera, A. L., Cintolesi, V. A., Clark, G., & Roux, S. J. (2009). Intersection of two signalling pathways: extracellular nucleotides regulate pollen germination and pollen tube growth via nitric oxide. *Journal of Experimental Botany*, 60(7), 2129-2138.
- Riewe, D., Grosman, L., Fernie, A. R., Wucke, C., & Geigenberger, P. (2008a). The potato-specific apyrase is apoplastically localized and has influence on gene expression, growth, and development. *Plant Physiology*, 147(3), 1092-1109.
- Riewe, D., Grosman, L., Fernie, A. R., Zauber, H., Wucke, C., & Geigenberger, P. (2008b). A Cell Wall-Bound Adenosine Nucleosidase is Involved in the Salvage of Extracellular ATP in *Solanum tuberosum*. *Plant and Cell Physiology*, 49(10), 1572-1579.
- Roberts, N. J., Brigham, J., Wu, B., Murphy, J. B., Volpin, H., Phillips, D. A., & Etzler, M. E. (1999). A Nod factor-binding lectin is a member of a distinct class of apyrases that may be unique to the legumes. *Molecular and General Genetics*, 262(2), 261-267.
- Robson, S. C. (2006). Hydrolysis of extracellular nucleotides by CD39/ENTPD family members: Prominent effects on thrombosis, vascular inflammation and immune reactions. *Acta Pharmacologica Sinica*, 27, 276-276.
- Roux, S. J., & Steinebrunner, I. (2007). Extracellular ATP: an unexpected role as a signalling molecule in plants. *Trends in Plant Science*, 12(11), 522-527.
- Sansom, F. M., Riedmaier, P., Newton, H. J., Dunstone, M. A., Muller, C. E., Stephan, H., Byres, E., Beddoe, T., Rossjohn, J., Cowan, P. J., d'Apice, A. J. F., Robson, S. C., & Hartland, E. L. (2008). Enzymatic properties of an ecto-nucleoside triphosphate diphosphohydrolase from *Legionella pneumophila* - Substrate specificity and requirement for virulence. *Journal of Biological Chemistry*, 283(19), 12909-12918.
- Schoemaker, J. M., Brasnett, A. H., & Marston, F. A. O. (1985). Examination of Calf Prochymosin accumulation in *Escherichia coli*-Disulfide linkages are a structural component of Prochymosin-containing Inclusion-bodies. *Embo Journal*, 4(3), 775-780.
- Schuler, H. (2001). ATPase activity and conformational changes in the regulation of actin. *Biochimica Et Biophysica Acta-Protein Structure and Molecular Enzymology*, 1549(2), 137-147.
- Silverman, J. A., Qi, H. L., Riehl, A., Beckers, C., Nakaar, V., & Joiner, K. A. (1998). Induced activation of the *Toxoplasma gondii* nucleoside triphosphate hydrolase leads to depletion of host cell ATP levels and rapid exit of

-
- intracellular parasites from infected cells. *Journal of Biological Chemistry*, 273(20), 12352-12359.
- Smith, T. M., Carl, S. A. L., & Kirley, T. L. (1999a). Mutagenesis of two conserved tryptophan residues of the E-type ATPases: Inactivation and conversion of an ecto-apyrase to an Ecto-NTPase. *Biochemistry*, 38(18), 5849-5857.
- Smith, T. M., & Kirley, T. L. (1999b). Site-directed mutagenesis of a human brain ecto-apyrase: Evidence that the E-type ATPases are related to the actin/heat shock 70/sugar kinase superfamily. *Biochemistry*, 38(1), 321-328.
- Song, C. J., Steinebrunner, I., Wang, X. Z., Stout, S. C., & Roux, S. J. (2006). Extracellular ATP induces the accumulation of superoxide via NADPH oxidases in Arabidopsis. *Plant Physiology*, 140(4), 1222-1232.
- Stein, N. (2008). CHAINSAW: a program for mutating PDB files used as templates in molecular replacement. *Journal of Applied Crystallography*, 41, 641-643.
- Steinebrunner, I., Jeter, C., Song, C., & Roux, S. J. (2000). Molecular and biochemical comparison of two different apyrases from Arabidopsis thaliana. *Plant Physiology and Biochemistry*, 38(12), 913-922.
- Steinebrunner, I., Wu, J., Sun, Y., Corbett, A., & Roux, S. J. (2003). Disruption of apyrases inhibits pollen germination in Arabidopsis. *Plant Physiology*, 131(4), 1638-1647.
- Stout, J. G., & Kirley, T. L. (1994). Purification and characterisation of the Ecto-Mg-ATPase of chicken gizzard smooth-muscle. *Journal of Biochemical and Biophysical Methods*, 29(1), 61-75.
- Stroppolo, M. E., Falconi, M., Caccuri, A. M., & Desideri, A. (2001). Superefficient enzymes. *Cellular and Molecular Life Sciences*, 58(10), 1451-1460.
- Surprenant, A., & North, R. A. (2009). Signalling at Purinergic P2X Receptors. *Annual Review of Physiology*, 71, 333-359.
- Tanaka, K., Gilroy, S., Jones, A. M., & Stacey, G. (2010). Extracellular ATP signalling in plants. *Trends in Cell Biology*, 20(10), 601-608.
- Tanaka, K., Nguyen, C. T., Libault, M., Cheng, J., & Stacey, G. (2011a). Enzymatic activity of the soybean Ecto-apyrase GS52 is essential for stimulation of nodulation. *Plant Physiology*, Ahead of print.
- Tanaka, K., Nguyen, C. T., Libault, M., Cheng, J. L., & Stacey, G. (2011b). Enzymatic Activity of the Soybean Ecto-Apyrase GS52 Is Essential for Stimulation of Nodulation. *Plant Physiology*, 155(4), 1988-1998.
- Taylor, J. S. (1981). Sarcoplasmic reticulum ATPase catalyzes hydrolysis of adenylyl-5'-yl imidodiphosphate. *Journal of Biological Chemistry*, 256(19), 9793-9795.
- Terwilliger, T. C., Grosse-Kunstleve, R. W., Afonine, P. V., Moriarty, N. W., Zwart, P. H., Hung, L. W., Read, R. J., & Adams, P. D. (2008). Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta Crystallographica Section D-Biological Crystallography*, 64, 61-69.
- Thomas, C., Rajagopal, A., Windsor, B., Dudler, R., Lloyd, A., & Roux, S. J. (2000). A role for ectophosphatase in xenobiotic resistance. *Plant Cell*, 12(4), 519-533.

-
- Tognoli, L., & Marrè, E. (1981). Purification and characterization of a divalent cation-activated ATP-ADPase from pea stem microsomes
Biochimica et Biophysica Acta, 642(1), 1-14.
- Traversocori, A., Chaimovich, H., & Cori, O. (1965). Kinetic studies and properties of potato apyrase. *Archives of Biochemistry*, 16, 173-184.
- Treuheit, M. J., Vaghy, P. L., & Kirley, T. L. (1992). Mg²⁺-ATPase from rabbit skeletal-muscle transverse tubules is 67-kiloDalton glycoprotein. *Journal of Biological Chemistry*, 267(17), 11777-11782.
- Tsumoto, K., Umetsu, M., Kumagai, I., Ejima, D., Philo, J. S., & Arakawa, T. (2004). Role of arginine in protein refolding, solubilization, and purification. *Biotechnology Progress*, 20(5), 1301-1308.
- Vagin, A., & Teplyakov, A. (1997). MOLREP: an automated program for molecular replacement. *Journal of Applied Crystallography*, 30, 1022-1025.
- Vasconcelos, E. G., Ferreira, S. T., deCarvalho, T. M. U., deSouza, W., Kettlun, A. M., Mancilla, M., Valenzuela, M. A., & VerjovskiAlmeida, S. (1996). Partial purification and immunohistochemical localization of ATP diphosphohydrolase from *Schistosoma mansoni* - Immunological cross-reactivities with potato apyrase and *Toxoplasma gondii* nucleoside triphosphate hydrolase. *Journal of Biological Chemistry*, 271(36), 22139-22145.
- Vincentelli, R., Canaan, S., Campanacci, V., Valencia, C., Maurin, D., Frassinetti, F., Scappucini-Calvo, L., Bourne, Y., Cambillau, C., & Bignon, C. (2004). High-throughput automated refolding screening of inclusion bodies. *Protein Science*, 13(10), 2782-2792.
- Vivian, J. P., Riedmaier, P., Ge, H. H., Le Nours, J., Sansom, F. M., Wilce, M. C. J., Byres, E., Dias, M., Schmidberger, J. W., Cowan, P. J., d'Apice, A. J. F., Hartland, E. L., Rossjohn, J., & Beddoe, T. (2010). Crystal Structure of a *Legionella pneumophila* Ecto-Triphosphate Diphosphohydrolase, A Structural and Functional Homolog of the Eukaryotic NTPDases. *Structure*, 18(2), 228-238.
- Vorhoff, T., Zimmermann, H., Pelletier, J., Sevigny, J., & Braun, N. (2005). Cloning and characterization of the ecto-nucleotidase NTPDase3 from rat brain: Predicted secondary structure and relation to other members of the E-NTPDase family and actin. *Purinergic Signalling*, 1(3), 259-270.
- Vorobiev, S., Strokopytov, B., Drubin, D. G., Frieden, C., Ono, S., Condeelis, J., Rubenstein, P. A., & Almo, S. C. (2003). The structure of nonvertebrate actin: Implications for the ATP hydrolytic mechanism. *Proceedings of the National Academy of Sciences of the United States of America*, 100(10), 5760-5765.
- Wang, T. F., & Guidotti, G. (1996). CD39 Is an Ecto-(Ca²⁺,Mg²⁺)-apyrase. *The journal of biological chemistry*, 271(17), 9898-9901.
- Wang, T. F., Ou, Y., & Guidotti, G. (1998). The transmembrane domains of ectoapyrase (CD39) affect its enzymatic activity and quaternary structure. *Journal of Biological Chemistry*, 273(38), 24814-24821.
- Weerasinghe, R. R., Swanson, S. J., Okada, S. F., Garrett, M. B., Kim, S. Y., Stacey, G., Boucher, R. C., Gilroy, S., & Jones, A. M. (2009). Touch induces ATP

-
- release in Arabidopsis roots that is modulated by the heterotrimeric G-protein complex. *Febs Letters*, 583(15), 2521-2526.
- Wu, J., Steinebrunner, I., Sun, Y., Butterfield, T., Torres, J., Arnold, D., Gonzalez, A., Jacob, F., Reichler, S., & Roux, S. J. (2007). Apyrases (nucleoside triphosphate-diphosphohydrolases) play a key role in growth control in Arabidopsis. *Plant Physiology*, 144(2), 961-975.
- Yang, F., Hicks-Berger, C. A., Smith, T. M., & Kirley, T. L. (2001). Site-directed mutagenesis of human nucleoside triphosphate diphosphohydrolase 3: The importance of residues in the apyrase conserved regions. *Biochemistry*, 40(13), 3943-3950.
- Yegutkin, G. G. (2008). Nucleotide- and nucleoside-converting ectoenzymes: Important modulators of purinergic signalling cascade. *Biochimica et Biophysica Acta*, 1783, 673-694.
- Yount, R. G., Babcock, D., Ballantyne, W., & Ojala, D. (1971). Adenylyl Imidodiphosphate, an Adenosine Triphosphate Analog Containing a P-N-P Linkage. *Biochemistry*, 10(13), 2484-2489.
- Zebisch, M., & Strater, N. (2007). Characterization of rat NTPDase1, -2, and -3 ectodomains refolded from bacterial inclusion bodies. *Biochemistry*, 46(42), 11945-11956.
- Zebisch, M., & Strater, N. (2008). Structural insight into signal conversion and inactivation by NTPDase2 in purinergic signaling. *Proceedings of the National Academy of Sciences of the United States of America*, 105(19), 6882-6887.
- Zhang, C., & McManus, M. T. (2000). Identification and characterisation of two distinct acid phosphatases in cell walls of roots of white clover. *Plant Physiology and Biochemistry*, 38(4), 259-270.
- Zhong, X. T., Buddha, M., Guidotti, G., Kriz, R., Somers, W., & Mosyak, L. (2008). Expression, purification and crystallization of the ecto-enzymatic domain of rat E-NTPDase1 CD39. *Acta Crystallographica Section F-Structural Biology and Crystallization Communications*, 64, 1063-1065.
- Zimmermann, H. (2000). Extracellular metabolism of ATP and other nucleotides. *Naunyn-Schmiedeberg's Archives of Pharmacology*, 362(4-5), 299-309.