

Molecular Characterisation of Recombinant Human Pyroglutamyl Peptidase (type I)

Thesis submitted for the degree of
Doctor of Philosophy

by

Paul-Roman Vaas, B.Sc.

Supervised by

Dr. Thecla Ryan

and

Dr. Brendan O'Connor

School of Biotechnology
Dublin City University
Ireland

September 2005

Declaration

I hereby certify that this material, which I now submit for assessment on the programme of study leading to the award of Doctor of Philosophy, is entirely my own work and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work.

Signed: Paul Roman Vaas

Paul-Roman Vaas

ID No.: 93531257

Date: 09/09/05

Dedications

This thesis, and the years used to accomplish it, I dedicate to...

... Mama und Papa, Ihr habt so viel für mich getan. Es ist schwierig meine endlose Dankbarkeit in einfache Worte zu fassen. Ich bin froh Euch zu haben.

... Dorothy, you've shared with me good times and bad. We tread life's path together.

... Thecla and Brendan, under your guidance I have learned so much (not least a great appreciation for the chilli pepper!), thank you both.

... my friends, it's been enjoyable.

Papers

Vaas, P-R., O'Connor, B. and Ryan, T. (2005) *Biochemical properties of purified human pyroglutamyl peptidase (type I), recombinantly expressed in E. coli.* (Manuscript in preparation)

Vaas, P-R., Kilbane, Z., O'Connor, B. and Ryan, T. (2005) *Structure-activity relationship studies of recombinant human pyroglutamyl peptidase (type I).* (Manuscript in preparation)

Vaas, P-R., Kilbane, Z., O'Connor, B. and Ryan, T. (2005) *Pyroglutamyl peptidase (types I & II): A review.* (Manuscript in preparation)

Posters

Vaas, P-R., O'Connor, B. and Ryan, T. *Molecular characterisation of a recombinant human neuropeptide-inactivating proteinase.* Biochemical Society, 57th Harden Conference, Oriel College, Oxford, UK. September 2003.

Presentations

Vaas, P-R., O'Connor, B. and Ryan, T. *Cloning, expression, purification, mutation and characterisation of a human neuropeptidase.* Biotechnology Seminar Series, Department of Biotechnology, Dublin City University, Dublin 9, Ireland. March 2003.

Vaas, P-R., O'Connor, B. and Ryan, T. *The use of bio-informatics software and biochemical assays for the study of Pyroglutamyl Peptidase (type I).* Society for General Microbiology Winter Meeting, Waterford Institute of Technology, Cork Road, Waterford, Ireland. January 2001.

Abbreviations

2-ME	2-mercaptoethanol
3D	Three-dimensional
A ₆₀₀	Absorbance at 600 nm
AA	Amino acid (see Appendix D)
AMC	7-amino-4-methylcoumarin
BCA	Bicinchoninic acid
bp	Base-pair
BSA	Bovine serum albumin
cDNA	complementary DNA
Conc.	Concentration
DAB	3,3'-Diaminobenzidine Tetrahydrochloride
DFP	Diisopropyl fluorophosphate
dH ₂ O	Deionised water
DMSO	Dimethyl sulphoxide
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
EC	Enzyme commission
EDTA	Ethylenediaminetetra acetic acid
IC ₅₀	Inhibitor conc. resulting in 50% loss of enzymatic activity
IEF	Isoelectric focusing
IMAC	Immobilised metal affinity chromatography
IPTG	Isopropyl-β-D-thiogalactopyranoside
k _{cat}	Turnover number
K _i	Dissociation constant
K _m	Michaelis constant
K _m ^{app}	Apparent K _m
LB	Luria Bertani
LHRH	Luteinising hormone-releasing hormone
Log	Logarithm
MCS	Multiple cloning site
MOPS	3-(N-Morpholino)propanesulfonic acid
mRNA	Messenger RNA

NTA	Nitrilotriacetic acid
OD	Optical density
ORF	Open reading frame
PAGE	Polyacrylamide gel electrophoresis
PAP	Pyroglutamyl peptidase
<i>p</i> -CMB	<i>p</i> -chloromercurybenzoate
PCR	Polymerase Chain Reaction
pGlu	Pyroglutamic acid
PIPES	Piperazine-1,4-bis(2-ethanesulfonic acid)
PMSF	Phenylmethanesulphonyl fluoride
RBS	Ribosome binding site
RNA	Ribonucleic acid
rRNA	Ribosomal RNA
RT-PCR	Reverse transcription PCR
SDS	Sodium dodecyl sulphate
TEMED	N, N, N, N'-tetramethyl ethylenediamine
TRH	Thyrotropin-releasing hormone
Tris	Tris (hydroxymethyl) amino methane
tRNA	Transfer RNA
UV	Ultraviolet light
v/v	Volume per volume
V_e	Elution volume
V_{max}	Maximal enzyme velocity
V_o	Void volume
w/v	Weight per volume
X-Gal	5-bromo-4-chloro-3-indolyl- β -D-galactopyranoside
β -NA	β -naphthylamide

Units

Da	Dalton
g	Gram
g	g-force (relative centrifugal force - rcf)
hr	Hour
L	Litre
m	Metre
M	Molar
min	Minute
sec	Second
pH	Logarithm of reciprocal hydrogen-ion concentration
pI	Isoelectric point
°C	Degrees celcius
Å	Angstrom (1.0×10^{-10} metres)
rpm	Revolutions per minute
kb	Kilo base
V	Volts
A	Ampere
Hz	Hertz

Prefixes

k	kilo (10^3)
c	centi (10^{-2})
m	milli (10^{-3})
μ	micro (10^{-6})
n	nano (10^{-9})
p	pico (10^{-12})

Declaration	I
Dedications	II
Publications	III
Abbreviations	IV
Table of Contents	VII
List of Figures	X1
List of Tables	XV
Abstract	XVI

Table of Contents

1.0 Introduction	1
1.1 Pyroglutamyl Peptidase	2
1.2 pGlu-peptides	2
1.3 PAP1	4
1.3.1 Physiological significance of PAP1	4
1.3.2 Classification of PAP1	6
1.3.3 Discovery and occurrence of PAP1	7
1.3.4 Detection of PAP1 activity	8
1.3.5 Purification of PAP1	9
1.3.6 Sequence data and analysis of PAP1	10
1.3.7 Catalytic residues of PAP1	17
1.3.8 Substrate specificity of PAP1	18
1.3.9 Characteristics of PAP1	22
1.3.10 Structure of PAP1	24
1.3.11 Mutational analysis of PAP1 structure and function	35
1.4 PAP2	36
1.5 Project aims & objectives	37
2.0 Materials & Methods	39
2.1 Bacterial strains, primers and plasmids	40
2.2 Media, solutions and buffers	45
2.3 Antibiotics	48

2.4 Isolation and purification of DNA and RNA	49
2.4.1 Isolation of plasmid DNA	49
2.4.1.1 Rapid boiling method	49
2.4.1.2 1-2-3 method	49
2.4.1.3 GenElute plasmid miniprep kit	50
2.4.2 Isolation of DNA from agarose gels	50
2.4.2.1 GeneClean	51
2.4.2.2 PerfectPrep kit	51
2.4.3 Purification and concentration of DNA samples	51
2.4.4 Isolation of RNA	52
2.4.5 Quantification of DNA and RNA	52
2.5 Agarose gel electrophoresis	52
2.6 Competent cells	53
2.6.1 Rubidium chloride method	53
2.6.2 TB method	54
2.6.3 Transformation of competent cells	55
2.6.4 Determining cell efficiency	55
2.7 Bacterial storage	55
2.8 Enzymes	55
2.8.1 Reverse transcription	56
2.8.2 Polymerase chain reaction	56
2.9 Gene manipulation	56
2.9.1 TA cloning of PCR products	56
2.9.2 Site-specific mutagenesis	57
2.10 DNA sequencing	58
2.11 Bio-Informatics	58
2.12 Protein expression	58
2.12.1 Standard expression culture	58
2.12.2 Preparation of cleared lysate	59
2.13 Protein purification	59
2.13.1 Standard IMAC procedure	59
2.13.2 Recharging of Ni-NTA resin	60
2.13.3 Preparation of dialysis tubing	60

2.14 Protein concentration	60
2.14.1 Quantitative determination by BCA assay	60
2.14.2 Quantitative determination by Coomassie assay	60
2.14.3 Qualitative determination by Coomassie assay	61
2.15 Fluorescence spectrometry	61
2.15.1 AMC standard curve	61
2.15.2 Quantitative fluorimetric PAP1 assay	61
2.15.3 Qualitative 96-well plate fluorimetric PAP1 assay	62
2.15.4 Fluorimetric colony assay	62
2.16 SDS-PAGE	62
2.16.1 Preparation of SDS gels	63
2.16.2 Sample preparation	63
2.16.3 Sample application	63
2.16.4 Gel staining	64
2.16.5 Gel analysis	65
2.17 UV Zymography	65
2.17.1 Native PAGE	65
2.17.2 UV zymogram development	65
2.18 Western blotting	65
2.19 PAP1 characterisation	66
2.19.1 Size exclusion chromatography	66
2.19.2 Determination of pH optimum	67
2.19.3 Determination of Michaelis constant	67
2.19.4 Determination of dissociation constant	67
2.19.5 Inhibition	68
2.19.6 Isoelectric focusing	68
2.19.7 Determination of crystallisation conditions	69
3.0 Cloning, expression and purification of recombinant human PAP1	72
3.1 Overview	73
3.2 Cloning of human PAP1 gene	73
3.3 Choice of expression host for recombinant human PAP1	76
3.4 Construction of <i>E. coli</i> expression system for recombinant human PAP1	77

3.4.1 Sub-cloning of human PAP1 gene into <i>E. coli</i> expression vector pQE-60	77
3.4.2 Sub-cloning of human PAP1 gene into <i>E. coli</i> expression vector pPC225	80
3.4.3 Construction of human PAP1 expression plasmid without His-tag	83
3.5 Preliminary characterisation of recombinant human PAP1	85
3.6 Optimisation of recombinant human PAP1 expression in <i>E. coli</i>	87
3.6.1 Expression of recombinant human PAP1 in <i>E. coli</i> XL10-Gold	92
3.6.2 Addressing codon bias of <i>E. coli</i> by site-specific mutagenesis	96
3.7 Purification of recombinant human PAP1	99
3.7.1 IMAC purification of recombinant human PAP1	100
3.8 Discussion	102
4.0 Characteristics of recombinant human PAP1	108
4.1 Biochemical properties of recombinant human PAP1	109
4.1.1 Relative molecular mass	109
4.1.2 Isoelectric point	110
4.1.3 Influence of pH	110
4.1.4 Influence of temperature	111
4.1.5 Influence of DTT and EDTA	112
4.1.6 Inhibition	113
4.2 Kinetic parameters of recombinant human PAP1	116
4.2.1 Competitive inhibition of recombinant human PAP1 by pGlu-peptides	121
4.3 Determining crystallisation conditions for recombinant human PAP1	123
4.4 Discussion	124
5.0 Structural analysis of recombinant human PAP1	128
5.1 Modelling the structure of human PAP1	129
5.1.1 Validation of the human PAP1 structure model	131
5.2 Functional analysis of recombinant human PAP1	133
5.2.1 Substrate binding pocket residues of human PAP1	136
5.2.2 Extended loop residues of human PAP1	137
5.2.3 C-terminal residues of human PAP1	138
5.2.4 Analysis of recombinant human PAP1 mutants	140
5.3 Discussion	145

5.3.1 Structure-based drug design	149
6.0 Summary & Recommendations	154
References	158
Appendices	176

List of Figures

Chapter 1

Figure 1.1 PAP cleavage of TRH	3
Figure 1.2 Catalytic mechanism of cysteine peptidases	7
Figure 1.3 PAP1 cleavage of pGlu-AMC	9
Figure 1.4 Amino acid sequence alignment of prokaryotic PAP1	12
Figure 1.5 Amino acid sequence alignment of eukaryotic PAP1	14
Figure 1.6 Amino acid sequence alignment of <i>Hsa</i> PAP1 and <i>Bam</i> PAP1	16
Figure 1.7 Location of conserved domains on <i>Bam</i> PAP1 monomer	17
Figure 1.8 Substitution analogues of $_L$ -pGlu	20
Figure 1.9 Structural analogues of pGlu	21
Figure 1.10 Active site-directed inhibitors of PAP1	21
Figure 1.11 Tetrameric arrangement of PAP1.	26
Figure 1.12 The A-D interface of <i>Bam</i> PAP1.	27
Figure 1.13 The A-C interface of <i>Bam</i> PAP1	27
Figure 1.14 Core structure of <i>Tli</i> PAP1	28
Figure 1.15 The A-B interface of <i>Tli</i> PAP1	28
Figure 1.16 The A-C interface of <i>Tli</i> PAP1	29
Figure 1.17 The A-D interface of <i>Pfu</i> PAP1	29
Figure 1.18 The A-B interface of <i>Pho</i> PAP1	30
Figure 1.19 Subunit monomer of <i>Bam</i> PAP1 with detail of active site region	31
Figure 1.20 Subunit monomer of <i>Tli</i> PAP1 highlighting active site	31
Figure 1.21 Subunit monomer of <i>Pfu</i> PAP1 highlighting active site	32
Figure 1.22 Subunit monomer of <i>Pho</i> PAP1 highlighting active site	32

Figure 1.23 Amino acid alignment of <i>Bam</i> PAP1, <i>Tli</i> PAP1, <i>Pfu</i> PAP1 and <i>Pho</i> PAP1	33
Figure 1.24 Structural comparison of <i>Bam</i> PAP1 with <i>Eco</i> PNP	34
 Chapter 2	
Figure 2.1 pCR2.1 vector	43
Figure 2.2 pQE-60 vector	44
Figure 2.3 pPC225 vector	44
Figure 2.4 pRARE vector	45
Figure 2.5 Reference DNA gel	53
Figure 2.6 Relative molecular weight protein marker (Mr)	64
Figure 2.7 Coloured molecular weight protein marker (CMr)	66
Figure 2.8 IEF Markers	69
 Chapter 3	
Figure 3.1 Cloning of human PAP1 gene	74
Figure 3.2 Human PAP1 putative mRNA sequence	75
Figure 3.3 Verification digest of pRV1	75
Figure 3.4 pRV1 plasmid map	76
Figure 3.5 Cloning strategy for pRV3 (pRV8)	79
Figure 3.6 Verification digest of pRV3	80
Figure 3.7 Deletion within pRV3	80
Figure 3.8 Cloning strategy for pRV5	81
Figure 3.9 Verification digest of pRV5	82
Figure 3.10 pRV5 sequence detail	83
Figure 3.11 Cloning strategy for pRV5_Δ6H	84
Figure 3.12 Verification digest of pRV5_Δ6H	85
Figure 3.13 Expression of <i>rHsa</i> PAP1 and <i>rHsa</i> PAP1 _{6H} in <i>E. coli</i>	86
Figure 3.14 Zymogram analysis of <i>rHsa</i> PAP1 _{6H}	87
Figure 3.15 Western blot analysis of <i>rHsa</i> PAP1 _{6H}	87
Figure 3.16 <i>E. coli</i> codon bias relative to <i>rHsa-pap1</i>	89
Figure 3.17 Growth curves for expression of <i>rHsa</i> PAP1 _{6H} in <i>E. coli</i>	90
Figure 3.18 Expression of <i>rHsa</i> PAP1 _{6H} in <i>E. coli</i> DH5α and INVαF'	91
Figure 3.19 Expression of <i>rHsa</i> PAP1 _{6H} in <i>E. coli</i> BL21 and Rosetta	91

Figure 3.20 Expression of rHsaPAP1 _{6H} in <i>E. coli</i> Nova Blue and RosettaBlue	92
Figure 3.21 Expression control in <i>E. coli</i> Nova Blue, RosettaBlue and XL10-Gold	92
Figure 3.22 Growth curve for expression of rHsaPAP1 _{6H} in <i>E. coli</i> XL10-Gold	94
Figure 3.23 Expression of rHsaPAP1 _{6H} in <i>E. coli</i> XL10-Gold	94
Figure 3.24 Expression control of rHsaPAP1 _{6H} in <i>E. coli</i> XL10-Gold	95
Figure 3.25 Effect of IPTG on rHsaPAP1 _{6H} expression in <i>E. coli</i> XL10-Gold	95
Figure 3.26 Effect of induction time on rHsaPAP1 _{6H} expression	96
Figure 3.27 Effect of sonication time on lysis	96
Figure 3.28 Confirmation of pRV5_R5 ^S K6 ^S by gradient PCR screening	98
Figure 3.29 Effect of Arg5/Lys6 codon substitution on rHsaPAP1 _{6H} expression	98
Figure 3.30 Effect of Arg5 codon substitution on rHsaPAP1 _{6H} expression	99
Figure 3.31 Level of transcription from pRV5, pRV5_R5 ^S a and pRV5_R5 ^S b	99
Figure 3.32 Imidazole elution profile	101
Figure 3.33 Purification of rHsaPAP1 _{6H}	101
Figure 3.34 Zymogram analysis of purified rHsaPAP1 _{6H}	102
Figure 3.35 Storage profile of rHsaPAP1 _{6H}	102
Figure 3.36 Potential structure of mRNA from pRV5, pRV5_R5 ^S a and pRV5_R5 ^S b	105

Chapter 4

Figure 4.1 Determination of rHsaPAP1 _{6H} native size	109
Figure 4.2 Iso-electric point of rHsaPAP1 _{6H}	110
Figure 4.3 Influence of pH on rHsaPAP1 and rHsaPAP1 _{6H} activity	111
Figure 4.4 Influence of temperature on rHsaPAP1 _{6H} activity	111
Figure 4.5 Temperature stability of rHsaPAP1 _{6H}	112
Figure 4.6 Influence of DTT on rHsaPAP1 _{6H} activity	112
Figure 4.7 Influence of EDTA on rHsaPAP1 _{6H} activity	113
Figure 4.8 Effect of L-pGlu on rHsaPAP1 _{6H} activity	114
Figure 4.9 Effect of 2-pyrrolidone on rHsaPAP1 _{6H} activity	114
Figure 4.10 Effect of DFP on rHsaPAP1 _{6H} activity	115
Figure 4.11 Effect of PMSF on rHsaPAP1 _{6H} activity	115
Figure 4.12 Effect of 1,10-phenanthroline on rHsaPAP1 _{6H} activity	116
Figure 4.13 Effect of 4,7-phenanthroline on rHsaPAP1 _{6H} activity	116
Figure 4.14 Michaelis-Menten curve for purified rHsaPAP1 _{6H}	117

Figure 4.15 Michaelis-Menten curve for <i>rHsaPAP1</i> _{6H} and <i>rHsaPAP1</i> in lysate	118
Figure 4.16 Lineweaver-Burk plot for purified <i>rHsaPAP1</i> _{6H}	118
Figure 4.17 Lineweaver-Burk plot for <i>rHsaPAP1</i> _{6H} and <i>rHsaPAP1</i> in lysate	119
Figure 4.18 Eadie-Hofstee plot for purified <i>rHsaPAP1</i> _{6H}	119
Figure 4.19 Eadie-Hofstee plot for <i>rHsaPAP1</i> _{6H} and <i>rHsaPAP1</i> in lysate	120
Figure 4.20 Hanes-Woolf plot for purified <i>rHsaPAP1</i> _{6H}	120
Figure 4.21 Hanes-Woolf plot for <i>rHsaPAP1</i> _{6H} and <i>rHsaPAP1</i> in lysate	121
Figure 4.22 K_i determination for selected peptides using Lineweaver-Burk plot	122
Figure 4.23 K_i determination for selected peptides using Eadie-Hofstee plot	122
Figure 4.24 K_i determination for selected peptides using Hanes-Woolf plot	123
Figure 4.25 Crystals observed while attempting crystallisation of <i>rHsaPAP1</i> _{6H}	124

Chapter 5

Figure 5.1 <i>HsaPAP1</i> 3D model compared with structural templates	130
Figure 5.2 Primary sequence alignment showing structural features	131
Figure 5.3 Primary sequence homology alignment	131
Figure 5.4 Comparison of <i>PAP1</i> active sites	132
Figure 5.5 Substitution of <i>HsaPAP1</i> residues Cys149 and His168	133
Figure 5.6 Primary sequence alignment showing residues targeted for mutagenesis	134
Figure 5.7 <i>HsaPAP1</i> 3D model showing residues targeted for mutagenesis	134
Figure 5.8 Verification of mutation using diagnostic restriction analysis	135
Figure 5.9 Substitution of <i>HsaPAP1</i> residues Phe13, Phe16 and Tyr147	137
Figure 5.10 Substitution of <i>HsaPAP1</i> residues Asp97 and Asn98	137
Figure 5.11 Cloning strategies for 3' deletion constructs	139
Figure 5.12 Expression of <i>rHsaPAP1</i> _{6H} mutants	141
Figure 5.13 RT-PCR analysis of pRV5 mutants	141
Figure 5.14 Comparison of <i>rHsaPAP1</i> _{6H} with <i>rHsaPAP1</i> _{ΔC}	142
Figure 5.15 Purification of <i>rHsaPAP1</i> _{6H} mutants	142
Figure 5.16 Effect of mutations on specific activity	143
Figure 5.17 Effect of mutations on K_m	144
Figure 5.18 Effect of mutations on k_{cat}	144
Figure 5.19 Western blot analysis of <i>rHsaPAP1</i> mutants	145
Figure 5.20 Zymogram analysis of <i>rHsaPAP1</i> mutants	145

Figure 5.21 Comparison of <i>Hsa</i> PAP1 and <i>Bam</i> PAP1 binding pockets	148
Figure 5.22 Structure-based design strategy for novel anti-HIV drugs	151

Chapter 6

Figure 6.1 Development of a fluorescence plate assay for PAP1 activity	156
--	-----

List of Tables

Chapter 1

Table 1.1 Some peptides with an N-terminal pGlu residue	2
Table 1.2 Prokaryotic PAP1 sequences	11
Table 1.3 Homology scores for prokaryotic PAP1 sequences	13
Table 1.4 Eukaryotic PAP1 sequences	13
Table 1.5 Homology scores for eukaryotic PAP1 sequences	15
Table 1.6 Characteristics of various PAP1 enzymes	23

Chapter 2

Table 2.1 Bacterial strains	40
Table 2.2 Primers	40
Table 2.3 Plasmids	42
Table 2.4 Preparation of SDS-PAGE gels	63
Table 2.5 Silver staining of SDS-PAGE gels	64
Table 2.6 Compounds tested for inhibition	68
Table 2.7 Crystal screen 1	70
Table 2.8 Crystal screen 2	71

Chapter 4

Table 4.1 Kinetic parameters for <i>rHsa</i> PAP1 _{6H} and <i>rHsa</i> PAP1	121
Table 4.2 K_i values for selected peptides	123

Chapter 5

Table 5.1 Mutant derivatives of <i>rHsa</i> PAP1 _{6H}	135
Table 5.2 Kinetic parameters of <i>rHsa</i> PAP1 _{6H} mutants	143

Abstract

Pyroglutamyl Peptidase I (PAP1, EC 3.4.19.3) hydrolytically cleaves pyroglutamic acid (pGlu) from the N-terminal of most pGlu-peptides. In higher organisms Thyrothropin Releasing Hormone is a notable biologically active substrate of PAP1. The sequence of human PAP1 was obtained from GenBank at NCBI (www.ncbi.nlm.nih.gov). Using suitable primers cDNA was synthesised using RNA isolated from a human cell line. Functionally active recombinant human PAP1 was expressed in Escherichia coli. To facilitate both high expression levels and ease of purification a C-terminal His-tag fusion was made. It was shown that this fusion did not affect enzymatic activity. Yields of up to 25 mg purified recombinant human PAP1 per litre of culture were achieved. Gel filtration chromatography and SDS-PAGE data showed that the recombinant human PAP1 exists as an active monomer. The Michaelis-Menten constant (K_m) for the fluorometric substrate pGlu-7-amino-4-methyl coumarin was determined as 51.8 μM and the turnover constant (K_{cat}) was determined as 3.75 s^{-1} . The activity of the enzyme displayed an absolute requirement for a thiol-reducing agent such as DTT. Optimal enzyme activity was observed at pH range 8.0-9.5 and temperature range 30-50 °C. The inhibitory effect of a range of compounds (e.g. 2-pyrrolidone) was measured. The three dimensional structure of human PAP1 was modelled on existing homologous PAP1 structures. Several amino acid residues were selected for site-directed mutagenesis. Kinetic analysis of the mutant PAP1 enzymes gave insight into the role these residues play within catalytic site, substrate binding pocket and other significant regions of PAP1. Preliminary determination of conditions under which recombinant human PAP1 forms protein crystals was carried out.

1.0 Introduction

1.1 Pyroglutamyl peptidase

Pyroglutamyl Peptidase (PAP) hydrolytically removes pyroglutamic acid (pGlu; 5-oxopyrrolidine-2-carboxylic acid) from the N-terminus of pGlu-peptides (Figure 1.1). To date two classes of PAP have been described, as reviewed by Cummins and O'Connor (1998). Type I (PAP1) is a widely distributed cytosolic cysteine peptidase with broad specificity for pGlu-substrates, found in a wide range of prokaryotic and eukaryotic organisms, including *Homo sapiens*. PAP1 is the topic of this study and is reviewed in Section 1.3. Type II (PAP2, Section 1.4) is highly specific toward the pGlu-His bond of Thyrotropin-releasing hormone (TRH; Table 1.1, Figure 1.1) or very closely related peptides. PAP2, a zinc-dependent metallopeptidase, exists in both a membrane-bound and serum form.

1.2 pGlu-peptides

pGlu is a cyclised derivative of glutamic acid (Figure 1.1), first described by Haitinger (1882) who reported that when glutamic acid was heated at 190°C it converted to a new compound that differed from glutamic acid in having lost a molecule of water. There are many reports of the enzymatic formation of pGlu from glutamic acid and glutaminyll peptides, as reviewed by Orłowski and Meister (1971), suggesting important physiological relevance. This relevance is highlighted by the fact that a number of bioactive peptides (e.g. neuropeptides) exhibit an N-terminal pGlu residue (Table 1.1).

Table 1.1 Some peptides with an N-terminal pGlu residue

Peptide ^a	Sequence
Thyrotropin-releasing Hormone	pGlu-His-Pro-NH ₂
Luteinizing Hormone-releasing Hormone	pGlu-His-Trp-Ser-Tyr-
Bombesin	pGlu-Gln-Arg-Leu-Gly-
Neurotensin	pGlu-Leu-Tyr-Glu-Asn-
Gastrin	pGlu-Gly-Pro-Trp-Leu-
Eledoisin	pGlu-Pro-Ser-Lys-

^a Further examples are given by Awadé *et al.* (1994).

Neuropeptides released in the autocrine, paracrine or endocrine contribute to intracellular communication which is essential in integrating the basic processes of life such as metabolic activity, cell differentiation and growth. Neuropeptides are typically short polypeptide chains with specific modifications to individual residues, such as an N-terminal pGlu residue. Such a modification can provide resistance against

degradation by various proteolytic enzymes and thus PAP possesses the unique ability to degrade or un-block members of this group of peptides.

An N-terminal pGlu residue can define the highly specific biological properties of neuropeptides, such as their ability to recognize and bind cell surface receptors. For example, TRH binds to its receptor via specific interactions between the pGlu ring and certain residues of the trans-membrane TRH receptor (Perlman *et al.*, 1994a; 1994b). Slight structural substitution in the pGlu ring results in an almost complete loss of receptor binding ability. TRH is one of the most widely documented biologically active pGlu-peptides. It was initially defined as a hormone for its ability to stimulate the release of Thyrotropin Stimulating Hormone (TSH) from the anterior pituitary gland and prolactin from the pituitary gland (Jackson, 1982). It has features that also qualify it as a neurotransmitter or “neuromodulator”, namely its extrahypothalamic distribution in the brain combined with its localisation and release at the synaptic level as well as its attachment to high affinity receptors, which show remarkable degree of anatomical localisation. It exhibits specific effects on neuronal activity and stimulates a wide range of centrally mediated behavioural effects. For comprehensive reviews of TRH see Griffiths (1985) and O’Leary and O’Connor (1995).

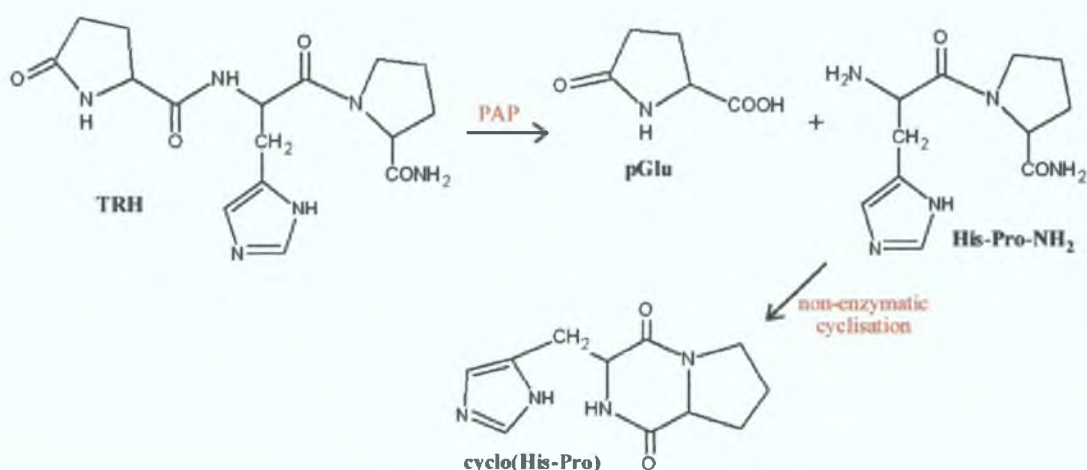


Figure 1.1 PAP cleavage of TRH

The cleavage of TRH by PAP liberates pGlu. The remaining His-Pro-NH₂ undergoes non-enzymatic cyclisation to cyclo(His-Pro). Illustrated using ChemSketch (Section 2.11).

The cleavage of TRH by PAP yields free pGlu and His-Pro-NH₂ (Figure 1.1). The subsequent non-enzymatic cyclisation of His-Pro-NH₂ generates cyclo(His-Pro) which has been shown to possess pharmacological, endocrine, electrophysiological and cardiovascular activity of its own (Prasad, 1995). Functions of free pGlu are less clear

but some pharmacological properties have been described and elevated plasma levels of free pGlu have been demonstrated in patients suffering from Huntington's disease and pyroglutamic academia (Cummins and O'Connor, 1998). He and Barrow (1999) and Russo *et al.* (2002) reported that amyloid- β peptides beginning with pGlu residues have a higher tendency of forming insoluble aggregates than other amyloid- β peptides. Alzheimer's disease is a neurodegenerative disorder characterised pathologically by the progressive deposition of amyloid- β peptides both intra-neuronally and extra-cellularly in the brain.

1.3 PAP1

As mentioned in Section 1.1 PAP1 is capable of cleaving pGlu from a variety of biologically active pGlu-peptides and pGlu-substrates (Figures 1.1 and 1.3), with exception of those commencing with pGlu-Pro (Section 1.3.8). PAP1 has been exploited commercially in peptide sequencing since N-terminal pGlu residues pose a problem in structural analysis of peptides (Martini *et al.*, 1985; Awadé *et al.*, 1994; Mozdzanowski *et al.*, 1998). Such a blocking group prevents application of the Edman degradation procedure.

1.3.1 Physiological significance of PAP1

At present, there is a varied view as to the exact physiological function of PAP1. Albert and Szewczuk (1972) suggested a possible role in the adsorption of peptides from the alimentary tract by virtue of PAP1 presence in the intestinal mucous membrane and duodenum (Pierro and Orsatti, 1970). Its wide distribution in such dissimilar tissues as skeletal muscle, brain, kidney, spleen, heart, liver and intestine suggests that PAP1 may contribute to the intracellular catabolism of peptides to free amino acids, thus being involved in regulating the cellular pool of free pGlu (Szewczuk and Kwiatkowska, 1970; Lauffart *et al.*, 1989; Mantle *et al.*, 1990; 1991). The physiological implications of this involvement were highlighted by Falkous *et al.* (1995), who found significantly increased levels of PAP1 in spinal cord of patients suffering from motor neuron disease. The identification of characteristic deposits of proteins in degenerating spinal cord neurons in motor neuron disease cases suggests that an abnormality of intracellular protein catabolism may contribute to the pathogenesis of this disorder.

The cytosolic location of PAP1 excludes a significant role in extracellular peptide degradation (Charli *et al.*, 1987; Abe *et al.*, 2004). Instead it is suggested that PAP1 may

contribute toward a means of returning pGlu terminating neuropeptides released from damaged or ageing vesicles back to the cellular amino acid pool (O’Cuinn *et al.*, 1990). PAP1 may be part of a system to degrade peptides that are produced in excess when secretion from neuropeptide-synthesising cells is suppressed. Supporting this are the findings by Faivre-Bauman *et al.* (1986) that the addition of specific PAP1 inhibitors to TRH-synthesising hypothalamic cells in culture, results in significant increase of TRH content and release from cells under both basal and K⁺ stimulated conditions.

A correlation between PAP1 activity and TRH levels in developing mammalian brain has been reported in a series of studies (De Gandarias *et al.*, 1992; 1994; 1998; 2000). As brain development progresses, a decrease in PAP1 activity coincides with increasing levels of TRH. The authors thus indicate that PAP1 plays a part in the normal development of mammalian brain. The same study shows that during earlier stages of development, high PAP1 activity is linked to elevated levels of cyclo(His-Pro) (Section 1.2), a correlation which was previously indicated (Prasad *et al.*, 1983).

The activity of PAP1 in the functionally connected retina and hypothalamus of rats exhibits a highly significant periodic fluctuation, coinciding with environmental light and dark conditions (Ramirez *et al.*, 1991; Sánchez *et al.*, 1996). A similar diurnal rhythm exhibited by TRH further demonstrates a possible association with PAP1.

Recent findings (Valdivia *et al.*, 2004) indicate that PAP1 participates in regulating the levels of seminal TRH and close analogues, and in mediating sperm death associated with necrozoospermia. However, the possibility that an increase in PAP1 activity may be due to sperm death in order to catalyse the remaining peptides must be considered. Further studies are ongoing to clarify if the increased levels of activity are a cause or consequence of sperm cell death.

In prokaryotic organisms PAP1 is likely to be a factor in the degradation and utilisation of pGlu-peptides as nutrients (Awadé *et al.* 1994; Gonzales and Robert-Baudouy, 1996). To this end Doolittle and Armentrout (1968) found that *Pseudomonas fluorescens* could grow on media having free pGlu as the sole carbon and nitrogen source (Section 1.3.3). PAP1 may also be involved in the detoxification of pGlu-peptides since high levels of such peptides would abnormally acidify the cell cytoplasm. Of note is the fact that PAP1 is not present in all bacteria (Section 1.3.3). Therefore PAP1 could contribute to such functions when present, rather than being uniquely responsible for them.

1.3.2 Classification of PAP1

Enzymes are classified by the Enzyme Commission (EC) list, which is maintained by the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (NC-IUBMB, <http://www.chem.qmul.ac.uk/iupac/jcbtn>). Proteolytic enzymes are of class 3, the hydrolases, and subclass 3.4, the peptide hydrolases or peptidases. Peptidases are either exopeptidases cleaving one or a few amino acids from the N- or C-terminal, or endopeptidases that cleave internally in polypeptide chains. Exopeptidases that act at a free N-terminal to liberate a single amino acid are known as aminopeptidases while those that act at a free C-terminal to liberate a single residue are known as carboxypeptidases. Exopeptidases that release several amino acid residues are known as oligopeptidases. Omega peptidases (EC 3.4.19) are exopeptidases that remove terminal residues which are substituted, cyclized (e.g. pGlu, Figure 1.1) or linked by iso peptide bonds.

Peptidases can also be grouped according to the residues responsible for catalysis. The main catalytic types are Cysteine, Serine, Aspartic and Metallo. In cysteine peptidases the catalytic mechanism requires a nucleophile and a proton donor. The nucleophile is the thiol group of a Cys residue. The proton donor is usually the imidazolium ring of a His residue. In some cases a third acidic residue is required to orientate the imidazolium ring of His (Barrett *et al.*, 1998). A generalised mechanism of cysteine peptidase catalysis is represented in Figure 1.2.

Over 40 families of cysteine peptidases have been shown to come from at least eight different evolutionary origins (Barrett and Rawlings, 2001). The notable MEROPS Peptidase Database (<http://merops.sanger.ac.uk>, Rawlings *et al.*, 2004) has grouped each of these origins as a clan (CA, CD, CE, CF, CH, CK, PA and PB) as an alternative classification system to the EC list. Each clan has distinctive functional, sequence and structural properties.

PAP1 was initially classified as EC 3.4.11.8 and has been referred to under several different names including pyrrolidonyl peptidase, pyrrolidone carboxyl peptidase, 5-oxoprolyl-peptidase, PYRase and pyroglutamyl aminopeptidase. PAP1 is currently classified as a cysteine omega peptidase (EC 3.4.19.3), which hydrolytically removes the amino terminal *L*-pyroglutamic acid (pGlu) residue from specific pGlu-substrates (Figures 1.1 and 1.2). In the MEROPS database PAP1 has been assigned as family C15,

clan CF. As yet, no other peptidases have been assigned to this family or clan (Barrett and Rawlings, 2001).

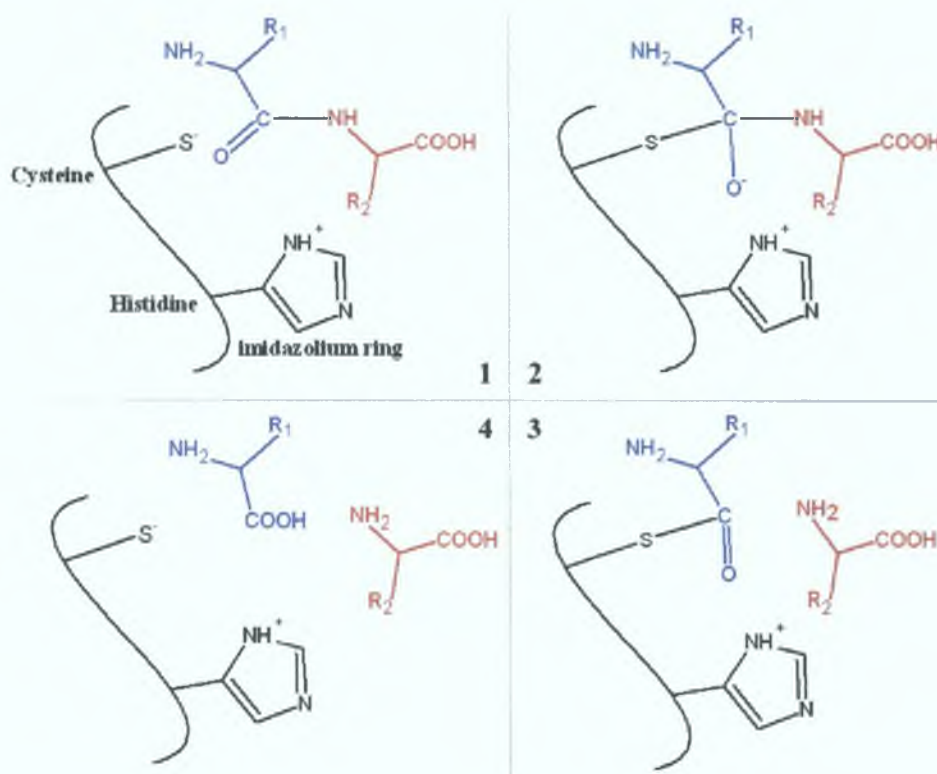


Figure 1.2 Catalytic mechanism of cysteine peptidases

The catalytic mechanism of cysteine proteases is represented via a generic dipeptide (blue-red). The thiolate ion of the Cys residue is the catalytic nucleophile and is stabilized through the formation of an ion pair with neighbouring imidazolium ring of a His residue (1). Catalysis proceeds through the formation of a covalent intermediate (2). The dipeptide is cleaved (3) and both fragments are liberated (4). For amino acid information see Appendix D. Illustrated using ChemSketch (Section 2.11).

1.3.3 Discovery and occurrence of PAP1

PAP1 activity was first identified in prokaryotes (Doolittle and Armentrout, 1968). The authors had actually intended to source an enzyme capable of opening the pyrrolidone ring of pGlu-peptides to benefit peptide sequencing (Section 1.3). An organism able to use free pGlu as sole carbon and nitrogen source was isolated from soil and identified as *Pseudomonas fluorescens*. A crude extract of *P. fluorescens* was able to liberate free pGlu from pGlu-alanine. This catalytic hydrolysis was attributed to the novel enzyme PAP1 (*Pfl*PAP1). Of a wide range of known peptidases, none were active toward this peptide. No PAP1 activity was found in *E. coli*.

PAP1 was subsequently documented in a wide range of prokaryotes including *Bacillus subtilis* (*Bsu*PAP1, Szewczuk and Mulczyk, 1969), *Klebsiella cloacae* (*Kcl*PAP1, Kwiatkowska *et al.*, 1974), *Streptococcus cremoris* (*Scr*PAP1, Exterkate, 1977),

Streptococcus faecium (SfaPAP1, Sullivan *et al.*, 1977), *Bacillus amyloliquefaciens* (BamPAP1, Tsuru *et al.*, 1978), *Streptococcus pyogenes* (SpyPAP1, Cleuziat *et al.*, 1992), *Staphylococcus aureus* (SauPAP1, Patti *et al.*, 1995), *Enterococcus faecalis* (EfaPAP1, Mineyama and Saito 1998), *Pyrococcus horikoshii* (PhoPAP1, Kawarabayasi *et al.*, 1998), *Pyrococcus furiosus* (PfuPAP1, Tsunasawa *et al.*, 1998), *Thermococcus litoralis* (TliPAP1, Singleton *et al.*, 1999a) and *Mycobacterium bovis* (MboPAP1, Kim *et al.*, 2001)

Szewczuk and Kwiatkowska (1970) found PAP1 activity in various vertebrates and plants. Among the vertebrates were human (*Homo sapiens*: HsaPAP1), bovine (*Bos Taurus*: BtaPAP1), porcine (*Sus scrofa*: SscPAP1), rabbit (*Oryctolagus cuniculus*: OcuPAP1), mouse (*Mus musculus*: MmuPAP1), Rat (*Rattus norvegicus*: RnoPAP1), guinea-pig (*Cavia porcellus*: CpoPAP1), pigeon (*Columba*: ColPAP1), chicken (*Gallus gallus*: GgaPAP1) and fish (*Cyprinus carpio*: CcaPAP1). Different tissues were tested including liver, kidney, spleen, lung, heart, brain, skeletal muscle and intestines. Among the plants were bean, oats, wheat, parsley, carrot, cabbage, potato and cauliflower. Tissues tested included leaves, seeds, sprouts, bulbs and roots. PAP1 activity was found in all samples screened. In vertebrates the highest activity was noted in liver and kidney tissues. PAP1 was later reported in hamster hypothalamus (*Mesocricetus auratus*: MauPAP1, Prasad and Peterkofsky, 1976) and feline brain (*Felis domesticus*: FdoPAP1, DeGandarias *et al.*, 1992).

1.3.4 Detection of PAP1 activity

PAP1 activity was initially determined by monitoring the release of alanine from pGlu-alanine (Doolittle and Armentrout, 1968). Mulczyk and Szewczuk (1970) developed a simple and rapid assay using the synthetic PAP1 substrate L-pyrrolidonal- β -naphthylamide (pGlu- β -NA). The assay exploits the release of β -NA by PAP1, which can be colourimetrically determined upon coupling with azotised *o*-dianisidine. A number of commercial kits involving pGlu- β -NA for the detection of PAP1 activity were confirmed as useful in preliminary and presumptive identification of enterococci and streptococci (Mitchell *et al.*, 1987; Kaufhold *et al.*, 1989; Dealler *et al.*, 1989).

A more direct assay for PAP1 activity, requiring fewer steps than with pGlu- β -NA, was developed using synthetic chromogenic (pGlu-*p*-nitroanilide) and fluorogenic (pGlu-7-amino-4-methylcoumarin; pGlu-AMC) substrates (Fujiwara and Tsuru, 1978). The former substrate is cleaved by PAP1, releasing *p*-nitroanilide which leads to an increase

in absorbance at 410 nm. pGlu-AMC is cleaved to yield free AMC, a fluorogenic compound with an emission of 440 nm upon excitation at 370 nm (Figure 1.3). The pGlu-AMC assay is used in this work and is one thousand fold more sensitive than the pGlu-*p*-nitroanilide assay.

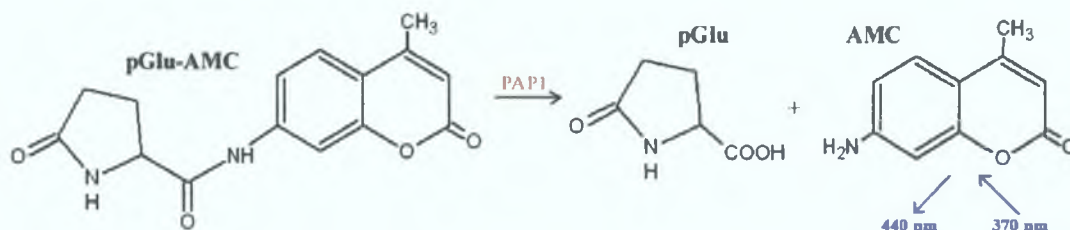


Figure 1.3 PAP1 cleavage of pGlu-AMC

pGlu-AMC is cleaved by PAP1 to yield free AMC, which fluoresces at 440 nm when excited at 370 nm. Illustrated using ChemSketch (Section 2.11).

Besson *et al.* (1994) have reported on the use of conductimetry to measure the activity of PAP1. Hydrolysis of pGlu-Ala by PAP1 was continuously monitored by measuring the variation in conductance. A drawback of this approach is the limitation to low ionic strength solutions, as sensitivity is defined by the relative change of conductance.

1.3.5 Purification of PAP1

Prior to the availability of a recombinant source, PAP1 was purified from various prokaryotic cell cultures and a wide range of eukaryotic tissues including human skeletal muscle, kidney and brain (Doolittle and Armentrout, 1968; Armentrout, 1969; Szewczuk and Mulczyk, 1969; Szewczuk and Kwiatkowska, 1970; Mudge and Fellows, 1973; Kwiatkowska *et al.*, 1974; Prasad and Peterkofsky, 1976; Exterkate, 1977; Sullivan *et al.*, 1977; Tsuru *et al.*, 1978; 1982; 1984; Browne and O'Cuinn, 1983; Lauffart and Mantle, 1988; Lauffart *et al.*, 1989; Mantle *et al.*, 1989; 1990; 1991; Cummins and O'Connor, 1996; Mineyama and Saito, 1998).

Various stages of ammonium sulphate fractionation, gel filtration, anion exchange chromatography, hydrophobic interaction chromatography and thiol-affinity chromatography were used to obtain pure enzyme. The PAP1 substrate analogue 2-pyrrolidone (Section 1.3.8) was frequently included during purification at a concentration of 100 mM. This competitive reversible inhibitor of PAP1 has a stabilising effect on activity and could consequently be removed by dialysis. Thiol-

reducing agents dithiothreitol (DTT) or 2-mercaptoethanol (2-ME) were also frequently used to aid in stabilisation.

The cloning of PAP1 genes (Section 1.3.6) allowed for the high level expression of several active recombinant prokaryotic PAP1 enzymes in *E. coli* (Gonzalès and Awadé, 1992; Awadé *et al.*, 1992; Yoshimoto *et al.*, 1993; Gonzalès and Robert-Baudouy, 1994; Tsunasawa *et al.*, 1998; Singleton *et al.*, 2000; Sokabe *et al.*, 2002). In some cases a fusion with six histidine codons allowed expression of PAP1 enzymes having a 6xHis tag on their N-terminal (Patti *et al.*, 1995; Kim *et al.*, 2001), facilitating one-step purification by nickel chelate affinity chromatography (an approach used in this work). High yields of purified PAP1, up to 45 mg per litre of recombinant *E. coli* culture, were obtained from these recombinant sources. At the time this project was initiated, no eukaryotic PAP1 had been expressed from a recombinant source.

1.3.6 Sequence data and analysis of PAP1

The genes for *Bsu*PAP1 (Awadé *et al.*, 1992) and *Spy*PAP1 (Cleuziat *et al.*, 1992) were the first PAP1 genes to be cloned and sequenced. Genomic libraries of *B. subtilis* and *S. pyogenes* were constructed and transformed into *E. coli*. Since this host is PAP1 negative, it was possible to develop a screen for PAP1 activity by a plate colony assay using pGlu- β -NA as described by Mulczyk and Szewczuk (1970). Positive clones were analysed by restriction mapping followed by DNA sequencing and the open reading frames (ORFs) of *Bsu*PAP1 (*Bsu-pap1*) and *Spy*PAP1 (*Spy-pap1*) were identified. The sequences showed no homology to any other data available on GenBank (Benson *et al.*, 1996) at that time. Later, by similar approaches, several other prokaryotic PAP1 genes were sequenced (Table 1.2).

Genomic digests of *B. subtilis*, *S. pyogenes* and *P. fluorescens* were southern blotted and probed with PCR fragments internal of *Bsu-pap1*, *Spy-pap1* and *Pfl-pap1* respectively (Awadé *et al.*, 1992; Cleuziat *et al.*, 1992; Gonzalès and Robert-Baudouy, 1994). In each case only one signal was detected concluding that the genes are present as single copies. *B. subtilis* and *S. pyogenes* total RNA samples were northern blotted and probed with the same *Bsu-pap1*, *Spy-pap1* internal PCR fragments respectively, establishing the mRNA sequences to be around 700 bp in length, consistent with the length of the ORFs.

Table 1.2 Prokaryotic PAP1 sequences

Organism	PAP1 Gene	GenBank Accession	ORF (bp)	PAP1 Enzyme	Amino Acids	Molecular Weight	Reference
<i>B. subtilis</i>	<i>Bsu-pap1</i>	X66034	645	<i>BsuPAP1</i>	215	23,774	Awadé <i>et al.</i> , 1992
<i>S. pyogenes</i>	<i>Spy-pap1</i>	X65717	645	<i>SpyPAP1</i>	215	23,132	Cleuziat <i>et al.</i> , 1992
<i>B. amyloliquefaciens</i>	<i>Bam-pap1</i>	D11035	645	<i>BamPAP1</i>	215	23,287	Yoshimoto <i>et al.</i> , 1993
<i>P. fluorescens</i>	<i>Pfl-pap1</i>	X75919	639	<i>PflPAP1</i>	213	22,438	Gonzalès and Robert-Baudouy, 1994
<i>S. aureus</i>	<i>Sau-pap1</i>	U19770	636	<i>SauPAP1</i>	212	23,227	Patti <i>et al.</i> , 1995
<i>P. furiosus</i>	<i>Pfu-pap1</i>	AB015291	624	<i>PfuPAP1</i>	208	22,822	Tsunasawa <i>et al.</i> , 1998
<i>T. litoralis</i>	<i>Tli-pap1</i>	Y13966	660	<i>TliPAP1</i>	220	24,745	Singleton <i>et al.</i> , 2000
<i>M. bovis</i>	<i>Mbo-pap1</i>	U91845	666	<i>MboPAP1</i>	222	23,193	Kim <i>et al.</i> , 2001
<i>P. horikoshii</i>	<i>Pho-pap1</i>	AP000002	618	<i>PhoPAP1</i>	206	22,640	Sokabe <i>et al.</i> , 2002

[Molecular weights were deduced from amino acid sequences]

An alignment of prokaryotic PAP1 amino acid sequences is shown in Figure 1.4. Sequence homology is represented in gray scale shading, with black being the highest homology. The four most highly conserved segments have been labelled domains 1 to 4. Domains 2 and 3 had previously been noted by Cummins and O'Connor (1998). As can be seen in Figures 1.6 and 1.7 these domains coincide closely with the active site region of PAP1. Table 1.3 summarises the homology scores corresponding to this alignment. The highest degree of sequence homology exists between *BsuPAP1* and *BamPAP1* (72% identity, 85% similarity) and the largest divergence exists between *MboPAP1* and *SauPAP1* (27% identity, 48% similarity).


```

BamPAP1 : MEKKLLTGFDPEGGGTNNSEWEAVKRNGAAEGP-ASTVSEQNTVVEYKSLAVIRE : 56
BsuPAP1 : MRKKLLTGFDPPDKETNNSEWEAAKRNGFETEE-AITTAEQNTVFRSALDTNRQ : 56
SpyPAP1 : M-S-ILVTGFDPEGGEAENSALAIKKIPATIHG--AELKCIEVETVQKSADVIQQ : 53
SauPAP1 : M-H-ILVTGAPPDNQDENSEWAVTQENII-GT-HTIDKLEETSEKKVDTIINK : 53
PflPAP1 : MR-ILLLTGFEPPDQDPNNSEWEAVRQDGVQLGSDVKTVARRLECAFATAGECTR : 56
PfuPAP1 : M-H-VLVTGEEPEGGKENSTERIAKD-DGIKIGD-AQVFGRVLEVVVGKAKEVIER : 54
TliPAP1 : M-RRKLLITGPEPPGGDSKNSTEQIAKYFDRKQIGN-AMVYGRVLEVSVKRATIEDKR : 55
PhoPAP1 : M-S-LLLTGFEPEGGDDKNSTMDIVEATSERI----PEVWGEILEVSEKRRARKLKL : 51
MboPAP1 : MSN-VLVTGEGPYGVTPNNIAQLTAEEIDGRTIAG-ATVVISRIIVNTTFESIAAAQQ : 55

```

domain 1

```

BamPAP1 : AAKKHQDDITLCVQAGTGMQTTSEPVAINIDAR---TEPNEENIPVGEDISQGE : 110
BsuPAP1 : AAKKHQDDITLCVQAGTGMQTTSEPVAINIDAR---TEPNECHPIDEETSPDGE : 110
SpyPAP1 : HIESFQDDAVLCIQQAGTGLTSEPVAINQDAR---TEPNEENIPIDTPRADCK : 107
SauPAP1 : TEASNHYDVVLATQQAGTGNATTEPVAINIDAR---TEPNDDEFPIDCAATHLQCA : 107
PflPAP1 : LIDELHAAVVIATLGGPESDTSVERVAINIDAR---TEPDLSELPIDTAIVADGF : 110
PfuPAP1 : TEEIKQDIAHYQLAPESATSTRTAVATDAR---TEPNEKKIEDEPTVPGAE : 108
TliPAP1 : YEEIKGEIVINQLAPETYSNITVERIAVHIDAR---TEPNDYSEIDEKTEEDAE : 109
PhoPAP1 : VLDDVRDDITINQLAPGRTHISVERVAINIDAR---TEPNDSELPKDEPTVEGGE : 105
MboPAP1 : AIAETESAIVIMLEPYGGRSMETVERLAQVNNCGRYGLACARVLGGPTDPAGE : 112

```

domain 2

```

BamPAP1 : AANTTGLTETRLVEEKKENIPAAVSYTAGTFVCHHLEFGMDETSRRHHPHIGGFI : 167
BsuPAP1 : AANTTRLPEHRTAKMKEHEIPAAVSYTAGTFVCHYLFYGLMDHISRTSPHIGGFI : 167
SpyPAP1 : AAYESTLPTKAMVAAHQAGIPASVNTAGTFVCHNHTMQATYLVDRKCPNARAGFI : 164
SauPAP1 : PRRSNLPVEAMTQSMINQSLGALNSAGTFVCHNVLIHLGYLQDKHYPHLRFGFI : 164
PflPAP1 : AAEFTTLPTKAMVKAREASIAASVQTAGTFVCHQVFFYLLQHALAGS--GYVSGFI : 165
PfuPAP1 : TATESTLPTKIMKKHEHESIPAYISNAGTLCHYVWYLSLHHSATKGYPKMSGFI : 165
TliPAP1 : LAKMATLPTVRAITKTERDNSEIPATISYASGYLCHYMPKTLHFSKIEGYPLKAGFI : 166
PhoPAP1 : AATPATIETREIVEEMKKNSIPAVLSYAGTYLCHFAMTLTHTSATKGYPKIAGFI : 162
MboPAP1 : VAYHATVPRAMVLAMRKACVPADVSDAAGTFVCHHMYGVVHHLAQKGLPVRAENT : 169

```

domain 3

```

BamPAP1 : HIFVIFEDTLOK-----SAPS-SLDHITKALKIAAVTAAVHED--DIETGGGELH : 215
BsuPAP1 : HIFVIFEDTIDK-----TAPSLSDITIVRAIRIAAVTAAQYDE--DVKSPGGTLH : 215
SpyPAP1 : HIFVIFEDTVDR-----PNTAANLDDITRGSEAAIFAIVDKDRSDLKRVGGATH : 215
SauPAP1 : HVFYIFEDVVGK-----SDTPSMFLEQIVAGTAAIEAISDHD---DLRIALGTE : 212
PflPAP1 : HVFLEEDVAGS-----QRPSMALDAMVAGQAANLTAWHTPV--DVKEAGGQVS : 213
PfuPAP1 : HVFYIFEDIDRIGK-GQVFPSSCYEMELEANVVAIEVALEELL----- : 208
TliPAP1 : HVFPTIDVNNFFLLGKNTPSMCLEAEKAEIADVKSLDYLEKDRDDIKIPL-- : 220
PhoPAP1 : HVFPTIDAVLEK-----KNTPSMSLDLEKCEIATRVQAQSALHSSQLR----- : 206
MboPAP1 : HLECLSESVAALDH---NLGVPSMSVQTAVAGTAGTEAAIRQSADIREPIPSRLQI : 222

```

domain 4

Figure 1.4 Amino acid sequence alignment of prokaryotic PAP1

Alignment was created using MultAlin with Blosum62-12-2 parameters and edited using GenDoc (Section 2.11). Sequence homology is represented by grey scale shading, with black being the highest homology. The most highly conserved domains have been labelled 1 to 4. For amino acid information see Appendix D.

Table 1.3 Homology scores for prokaryotic PAP1 sequences

	<i>Bsu</i> PAP1	<i>Spy</i> PAP1	<i>Sau</i> PAP1	<i>Pfl</i> PAP1	<i>Pfu</i> PAP1	<i>Tli</i> PAP1	<i>Pho</i> PAP1	<i>Mbo</i> PAP1
<i>Bam</i> PAP1	72 (85)	48 (68)	41 (62)	43 (63)	38 (57)	32 (57)	41 (59)	32 (52)
	<i>Bsu</i> PAP1	47 (65)	43 (62)	46 (65)	39 (57)	36 (59)	37 (56)	33 (50)
		<i>Spy</i> PAP1	49 (66)	41 (61)	36 (57)	34 (59)	39 (60)	31 (49)
			<i>Sau</i> PAP1	43 (60)	35 (50)	34 (54)	36 (51)	27 (48)
				<i>Pfl</i> PAP1	38 (56)	37 (55)	40 (57)	36 (51)
					<i>Pfu</i> PAP1	53 (71)	51 (73)	33 (54)
						<i>Tli</i> PAP1	53 (69)	32 (53)
							<i>Pho</i> PAP1	34 (47)

[Identity scores are given, with similarity in parenthesis]

A BLAST search (Section 2.11) of known PAP1 sequences against the current GenBank data (August 2004) reveals a number of confirmed and putative eukaryotic PAP1 gene sequences (Table 1.4) many of which have not yet been documented in publications. Sequence homology is also seen with several putative prokaryotic PAP1 sequences and un-identified protein products (not tabulated).

Table 1.4 Eukaryotic PAP1 sequences

Organism	Common Name	PAP1 Gene	GenBank Accession	ORF (bp)	PAP1 gene product	Amino Acids	Molecular Weight
<i>H. sapiens</i>	Human	<i>Hsa-pap1</i>	AJ278828 ^a	627	<i>Hsa</i> PAP1	209	23,138
<i>M. musculus</i>	Mouse	<i>Mmu-pap1</i>	AJ278829 ^a	627	<i>Mmu</i> PAP1	209	22,934
<i>R. norvegicus</i>	Rat	<i>Rno-pap1</i>	AB098134 ^b	624	<i>Rno</i> PAP1	208	22,800
<i>Tetraodon nigroviridis</i>	Green Pufferfish	<i>Tni-pap1</i> *	CAAE01014609	615	<i>Tni</i> PAP1*	205	22,376
<i>Takifugu rubripes</i>	Japanese Pufferfish	<i>Tru-pap1</i> *	AJ301641	633	<i>Tru</i> PAP1*	211	23,331
<i>Apis mellifera</i>	Honey Bee	<i>Ame-pap1</i> *	XM_392560	573	<i>Ame</i> PAP1*	191	21,941
<i>Drosophila melanogaster</i>	Fruit Fly	<i>Dme-pap1</i> *	NM_168616	672	<i>Dme</i> PAP1*	224	24,807
<i>Anopheles gambiae</i>	Mosquito	<i>Aga-pap1</i> *	XM_308793	627	<i>Aga</i> PAP1*	209	23,940
<i>Arabidopsis thaliana</i>	Cress	<i>Ath-pap1</i> *	NM_104547	657	<i>Ath</i> PAP1*	219	24,046
<i>Caenorhabditis elegans</i>	Nematode	<i>Cel-pap1</i> *	NM_060090	822	<i>Cel</i> PAP1*	274	31,007

[^a Dando *et al.*, 2003; ^b Abe *et al.*, 2003; * Putative sequences; Molecular weights were deduced from amino acid sequences]

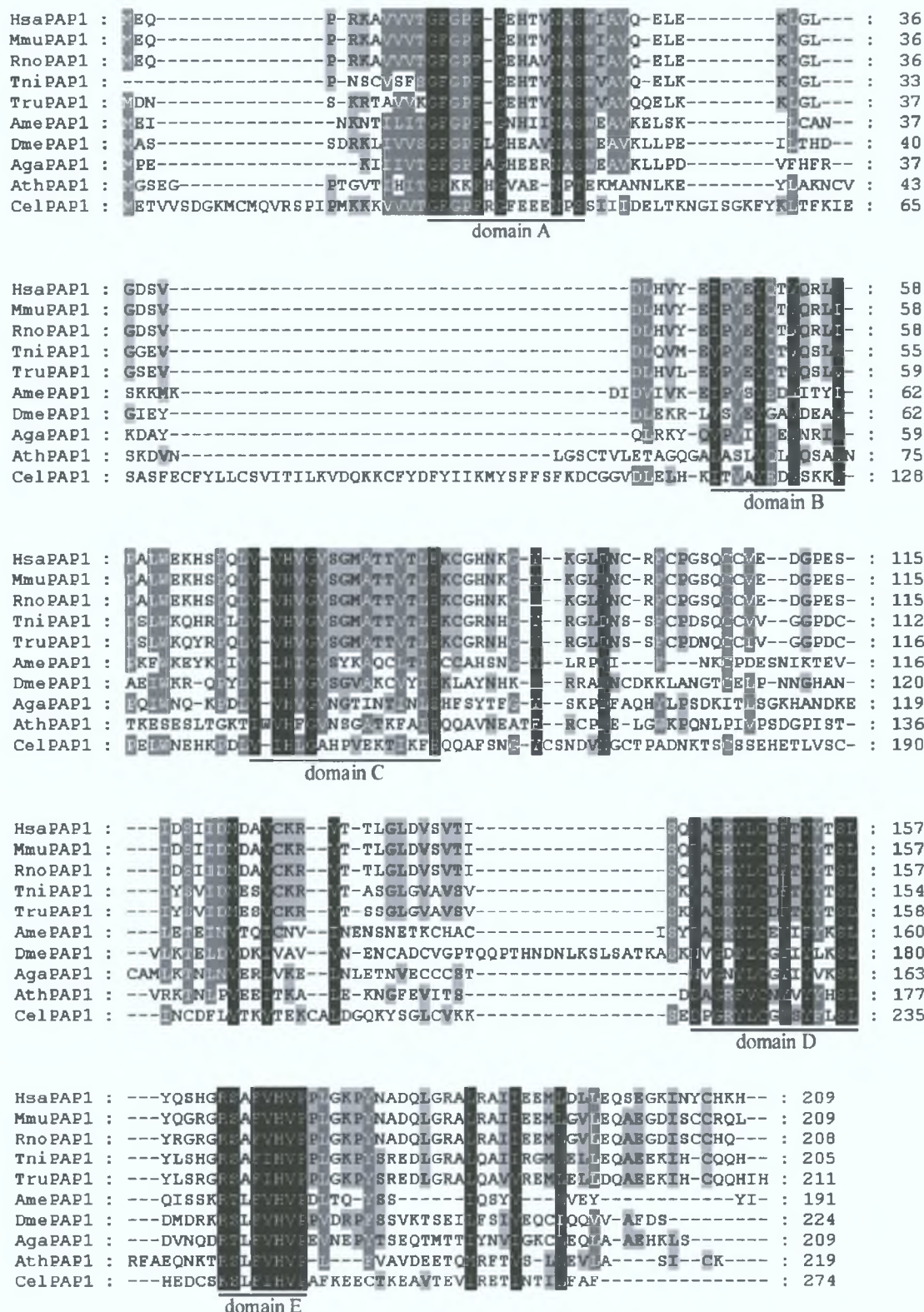


Figure 1.5 Amino acid sequence alignment of eukaryotic PAP1

Alignment was created using MultAlin with Dayhoff-8-0 parameters and edited using GenDoc (Section 2.11). Sequence homology is represented by grey scale shading, with black being the highest homology. The most highly conserved domains have been labelled A to E. For amino acid information see Appendix D.

Figure 1.5 shows an alignment of eukaryotic PAP1 amino acid sequences. On this alignment the regions of highest homology have been designated domains A to E. These domains have a degree of overlap with the prokaryotic segments of high homology (Figure 1.4) as can be seen in Figure 1.6.

The *Cel*PAP1 sequence has several distinct regions of added sequence. Table 1.5 has a summary of the homology scores for this alignment. Highest homology exists between the rodent sequences *Mmu*PAP1 and *Rno*PAP1 (98% identity/similarity). These also show closest homology to *Hsa*PAP1 (94% identity, 96% similarity). Such high homology is not observed among the insect sequences *Dme*PAP1, *Aga*PAP1 and *Ame*PAP1 (at best 35% identity, 56% similarity). Interestingly, the insect sequences show a similar degree of homology with *Hsa*PAP1 as they do with each other (~30% identity, ~50% similarity). The plant sequence *Ath*PAP1 and nematode sequence *Cel*PAP1 exhibit the greatest divergence from all other eukaryotic sequences.

Table 1.5 Homology scores for eukaryotic PAP1 sequences

	<i>Mmu</i> PAP1	<i>Rno</i> PAP1	<i>Tni</i> PAP1	<i>Tru</i> PAP1	<i>Ame</i> PAP1	<i>Dme</i> PAP1	<i>Aga</i> PAP1	<i>Ath</i> PAP1	<i>Cel</i> PAP1
<i>Hsa</i> PAP1	94 (96)	94 (96)	71 (84)	68 (85)	31 (50)	30 (46)	29 (51)	22 (38)	22 (36)
	<i>Mmu</i> PAP1	98 (98)	69 (81)	68 (83)	30 (49)	30 (47)	29 (50)	22 (38)	22 (36)
		<i>Rno</i> PAP1	69 (80)	67 (83)	30 (49)	31 (47)	29 (50)	22 (38)	22 (36)
			<i>Tni</i> PAP1	87 (91)	29 (50)	29 (44)	28 (47)	22 (37)	20 (34)
				<i>Tru</i> PAP1	32 (52)	31 (46)	29 (47)	23 (38)	20 (35)
					<i>Ame</i> PAP1	28 (46)	29 (48)	21 (38)	21 (35)
						<i>Dme</i> PAP1	35 (56)	20 (36)	18 (32)
							<i>Aga</i> PAP1	20 (36)	20 (36)
								<i>Ath</i> PAP1	14 (31)

[Identity scores are given, with similarity in parenthesis]

The prokaryotic sequence that *Hsa*PAP1 shows most homology to is *Bam*PAP1. Figure 1.6 shows an alignment of these two sequences. They exhibit an identity of 25% and a similarity of 45%, which is a higher score than seen among some of the eukaryotic sequences (Table 1.5). The conserved domains from Figures 1.4 and 1.5 have been shown in association with *Bam*PAP1 and *Hsa*PAP1 respectively. Also indicated are several residues of functional significance in *Bam*PAP1. Cys144, His168 and Glu81 constitute the catalytic triad of *Bam*PAP1 (as described in Section 1.3.7) while Phe10 and Phe13 contribute to the hydrophobic binding pocket of *Bam*PAP1 (as described in Section 1.3.10.2). Although the conserved eukaryotic and prokaryotic domains do not complement completely they do overlap at these key residues.

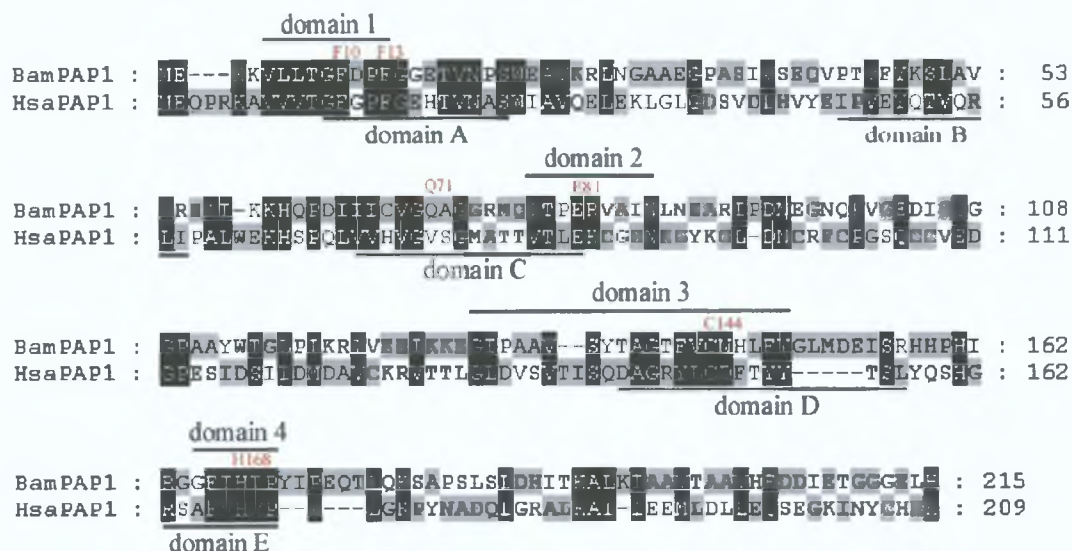


Figure 1.6 Amino acid sequence alignment of HsaPAP1 and BamPAP1
 Alignment was created using MultAlin with Blosum62-12-2 parameters and edited using GenDoc (Section 2.11). Sequence homology is represented by grey scale shading, with black being the highest homology. The domains introduced in Figures 1.4 and 1.5 are shown. Several functionally relevant residues have been marked in red (numbering applies to BamPAP1). For amino acid information see Appendix D.

Furthermore, these functional residues together with the prokaryotic conserved domains are shown on a ribbon diagram of the BamPAP1 monomer (described in more detail in Section 1.3.10.2) in Figure 1.7. It can be seen that the conservation is in correlation with the catalytic region of BamPAP1.



Figure 1.7 Location of conserved domains on *BamPAP1* monomer

The prokaryotic conserved domains introduced in Figure 1.4 are coloured in red on ribbon diagram of *BamPAP1* monomer. Catalytic residues Cys144, His168 and Glu81 (Section 1.3.6 and 1.3.10.2) and residues Phe10 and Phe13 contributing to hydrophobic pocket (Section 1.3.10.2) are shown. For amino acid information see Appendix D. Generated using DeepView (Section 2.11).

1.3.7 Catalytic residues of PAP1

The catalytic importance of a cysteine thiol group in PAP1 activity has been demonstrated by its absolute requirement for a thiol-reducing agent such as DTT or 2-ME. (Armentrout, 1969; Szewczuk and Mulczyk, 1969; Szewczuk and Kwiatkowska, 1970; Kwiatkowska *et al.*, 1974; Tsuru *et al.*, 1978; Tsuru *et al.*, 1982; Tsuru *et al.*, 1984; Prasad, 1987; Mantle *et al.*, 1991; Cummins and O'Connor, 1996; Tsunasawa *et al.*, 1998; Singleton and Littlechild, 2001, Dando *et al.*, 2003).

Further demonstrating this are the findings that PAP1 activity is highly sensitive to μM concentrations of thiol-blocking compounds such as iodoacetate, iodoacetamide, *p*-chloromercuribenzoate (*p*-CMB), *p*-mercuriphenylsulphonate, *N*-ethylmaleimide and sodium tetrathionate (Doolittle and Armentrout, 1968; Szewczuk and Mulczyk, 1969; Szewczuk and Kwiatkowska, 1970; Mudge and Fellows, 1973; Kwiatkowska *et al.*, 1974; Prasad and Peterkofsky, 1976; Tsuru *et al.*, 1978; Bauer and Kleinkauf, 1980; Tsuru *et al.*, 1982; Browne and O'Cuinn, 1983; Awadé *et al.*, 1992; Gonzalès and Robert-Baudouy, 1994; Patti *et al.*, 1995; Cummins and O'Connor, 1996; Tsunasawa *et al.*, 1998; Mineyama and Saito, 1998; Singleton *et al.*, 2000; Singleton and Littlechild, 2001; Dando *et al.*, 2003).

A serine peptidase inhibitor phenylmethylsulphonyl fluoride (PMSF) had no effect on the activity of PAP1 (Tsuru *et al.*, 1982; Lauffart *et al.*, 1989; Mantle *et al.*, 1990; Gonzalès and Robert-Baudouy, 1994; Singleton *et al.*, 2000).

PAP1 being established as a cysteine peptidase, this implicated the presence of a His residue as proton donor and perhaps a third acidic residue such as Glu or Asp to orientate the imidazolium ring (Section 1.3.2). Site-specific mutagenesis studies were conducted to identify the catalytic residues of PAP1.

The two cysteine residues of *Bam*PAP1, Cys68 and Cys144, were mutated to Ser (Yoshimoto *et al.*, 1993). The mutant *Bam*PAP1_C68mS had wild type activity while *Bam*PAP1_C144mS was completely inactive, implicating Cys144 as providing the active site thiol group. Also, by titration with 5,5'-dithio-bis-(2-nitrobenzoate), it was shown that the Cys68 is located internally.

Le Saux *et al.* (1996) investigated several residues as possible contributors to the active site of *Pfl*PAP1, based on homology alignments with known PAP1 sequences. Substitutions of Cys144 and His166 by Ala and Ser, respectively, completely abolished activity of *Pfl*PAP1 without affecting protein conformation as shown by gel electrophoresis. Similarly, one of Glu81, Asp89 or Asp94 was shown to be a likely third member of an active triad of *Pfl*PAP1.

Tsunasawa *et al.* (1998) substituted the Cys142 residue of *Pfu*PAP1 with Ser, resulting in loss of activity. In light of published data and by sequence analysis, these authors proposed the catalytic triad of *Pfu*PAP1 to consist of Cys142, His166 and Glu79, which correspond to Cys144, His168 and Glu81 of *Bam*PAP1 respectively. This catalytic triad was later confirmed by solution of the tertiary structure of *Bam*PAP1 and *Pfu*PAP1 as outlined in Section 1.3.10.2.

1.3.8 Substrate specificity of PAP1

Many studies have contributed to an accurate definition of the substrate specificity of PAP1 (Doolittle and Armentrout, 1968; Uliana and Doolittle, 1969; Szewczuk and Mulczyk, 1969; Armentrout, 1969; Mudge and Fellows, 1973; Kwiatkowska *et al.*, 1974; Sullivan *et al.*, 1977; Podell and Abraham, 1978; Fujiwara *et al.*, 1979; Tsuru *et al.*, 1982; Browne and O'Cuinn, 1983; Lauffart *et al.*, 1989; Mantle *et al.*, 1990; Mantle *et al.*, 1991; Cummins and O'Connor, 1996; Mineyama and Saito, 1998; Tsunasawa *et al.*, 1998).

PAP1 can be said to hydrolytically remove L -pGlu from L -pGlu- L -X, where X is any amino acid (except proline), a peptide or an arylamide (e.g. AMC). PAP1 exhibits specificity toward optical isomers. *Pf*/PAP1 had no activity toward D -pGlu- L -Ala and only slight trace of activity toward L -pGlu- D -Ala (the latter possibly due to slight contamination by L -pGlu- L -Ala). Z - D -pGluCK and Z - D -pGluDK, optical isomers of site-specific inhibitors mentioned below were inert toward PAP1.

The residue directly adjacent to pGlu affects the rate of hydrolysis. The order of preference is: Thr > Glu > Met > Ala > Lys > Gly > Ser > Tyr > Arg > Phe > His > Trp > Asp > Leu > Ile >> Val as determined for *Bam*PAP1, which is generally in agreement with the other PAP1s tested. pGlu-Pro was not hydrolysed by PAP1, with one notable exception being *Kc*/PAP1. However, the hydrolysis of pGlu-Pro by *Kc*/PAP1 was significantly slower than pGlu-Ala or even pGlu-Val.

The size of the peptide substrate does not affect specificity. *Pf*/PAP1 removed pGlu from the relatively large bovine fibrinogen without splitting any other peptide bond. Specific cleavage of pGlu from several bioactive peptides including TRH, LHRH, Liliberin, Neurotensin and Bombesin was widely demonstrated, as well as for other large peptides such as immunoglobulin polypeptide chains.

Synthetic analytical substrates pGlu- β -NA, pGlu-AMC and pGlu-His-Pro-AMC were shown to be hydrolysed at satisfactory rates, *Cpo*PAP1 having five times greater affinity toward pGlu-AMC than for TRH.

A recent study by Abe *et al.* (2004) has shown that PAP1 can tolerate some single atom substitutions on the pGlu ring (Figure 1.8). PAP1 cleavage was observed for substrates having the carbon atom at position 4 of the pGlu moiety of pGlu-Ala replaced with sulphur (L -OTCA- L -Ala), oxygen (L -OOCA- L -Ala) and nitrogen (L -OICA- L -Ala). The affinities of these substrates to PAP1 were reduced. The carbon atom at position 4 is located inside the hydrophobic pocket during binding as shown by Ito *et al.* (2001), described in Section 1.3.10.2. Thus the larger size of a sulphur atom or the hydrophilic nature of oxygen or nitrogen compared to carbon might interfere with insertion into the hydrophobic pocket. The tolerance of PAP1 for a nitrogen atom substitution had previously been reported by Capecchi and Loudon (1984), who also demonstrated that a six membered ring was not accepted, highlighting the sensitivity of PAP1 to the ring size of pGlu.

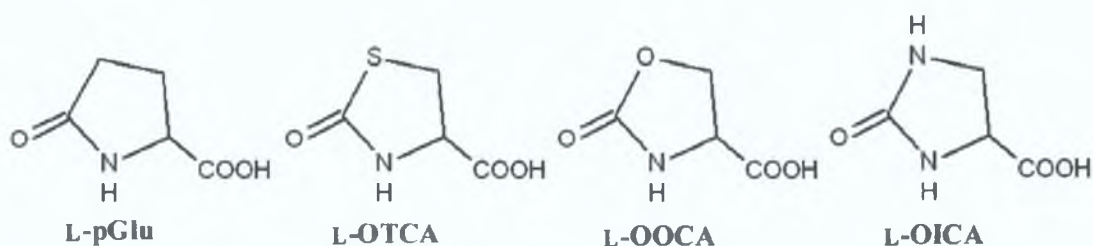


Figure 1.8 Substitution analogues of L -pGlu

Synthetic analogues of L -pGlu (L -5-oxopyrrolidine-2-carboxylic acid) having the carbon atom at position 4 replaced by sulphur (L -5-oxothiazolidine-2-carboxylic acid, L -OTCA), oxygen (L -5-oxooxazolidine-2-carboxylic acid, L -OOCA) and nitrogen (L -5-oxoimidazolidine-2-carboxylic acid, L -OICA). Illustrated using ChemSketch (Section 2.11).

2-pyrrolidone (Figure 1.9) acts as a competitive reversible inhibitor of PAP1 (Armentrout and Doolittle, 1969; Armentrout, 1969; Mudge and Fellows, 1973; Sullivan *et al.*, 1977; Cummins and O'Connor, 1996). Up to 50% inhibition was observed with 10 mM 2-pyrrolidone while complete inhibition was not achieved by increased concentration. Activity could be completely restored upon removal of 2-pyrrolidone by dialysis. This compares with the inhibitory effect free L -pGlu has on the catalysis of other pGlu-substrates (Szewczuk and Mulczyk, 1969; Szewczuk and Kwiatkowska, 1970; Tsuru *et al.*, 1978). A synthetic aldehyde analog of pGlu, 5-oxoprolinal was also found to be a potent inhibitor of PAP1 (Friedman *et al.*, 1985). Competitive inhibition by these compounds is due to their binding to the PAP1 active site instead of the substrate. The similarity in their structures can be seen in Figure 1.9. Other structurally similar molecules such as L -proline and L -hydroxyproline (Figure 1.9) did not inhibit PAP1 activity (Sullivan *et al.*, 1977). A synthetic TRH analogue having pGlu replaced with a sulphonamido analogue (Figure 1.9) was resistant to hydrolysis by PAP1 (Brunetti *et al.*, 2002). Ito *et al.* (2001) have given insight into this specificity by showing that together the nitrogen and carbonyl oxygen of the pGlu moiety are necessary for hydrogen bond formation within the active site of PAP1 during catalysis (Section 1.3.10.2).

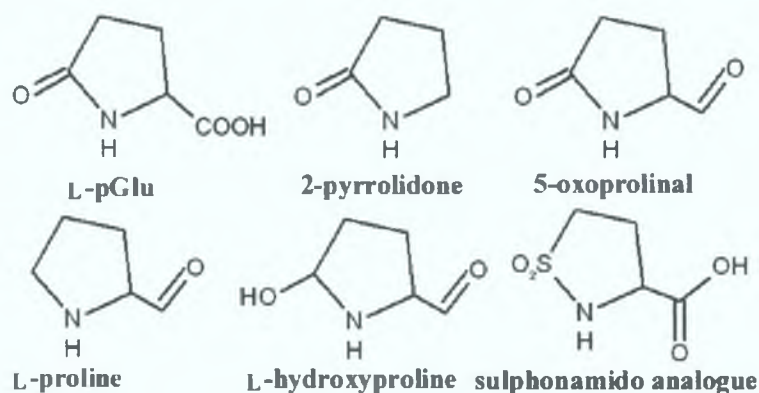


Figure 1.9 Structural analogues of pGlu

Various structural analogues of pGlu. Only those molecules having a nitrogen at position 1 together with a carbonyl oxygen at position 5 can bind into the PAP1 active site. Illustrated using ChemSketch (Section 2.11).

A study using synthetic site-specific inhibitors of PAP1 (Figure 1.10) was carried out (Fujiwara *et al.*, 1981a; 1981b; 1982; Wilk *et al.*, 1985). These compounds cause rapid and complete inhibition of PAP1 activity. That this inhibition was active site-specific was demonstrated by prior treatment of PAP1 with *p*-CMB, a thiol-blocking compound mentioned in Section 1.3.7. Activity could be completely restored by dialysis with 2-ME or DTT containing buffer. Similarly, prior protection of the active site could be achieved with pGlu-Val. The rate of inhibition was significantly higher for pGluCK and pGluDK than for ZpGluCK and ZpGluDK giving further evidence for the importance of the nitrogen and oxygen of the pGlu moiety as mentioned above. Furthermore, the Z-D-pGluCK and Z-D-pGluDK optical isomers were inert toward the enzyme due to the optical specificity of PAP1 as mentioned above.

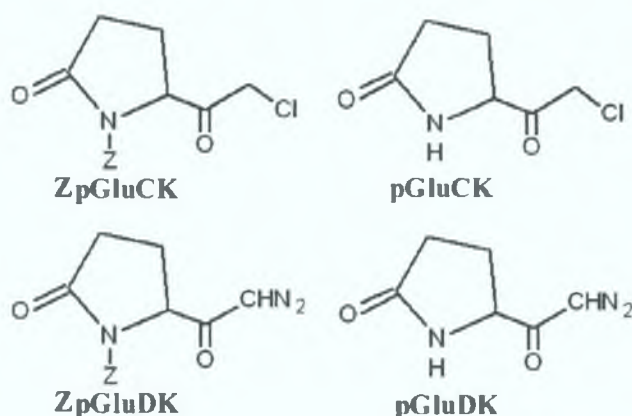


Figure 1.10 Active site-directed inhibitors of PAP1

Synthetic active site-directed inhibitors of PAP1: N^{α} -carbobenzoxymethyl-L-pGlu chloromethyl ketone (pGluCK), N^{α} -carbobenzoxymethyl-L-pGlu chloromethyl ketone (ZpGluCK), L-pGlu diazomethyl ketone (pGluDK) and N^{α} -carbobenzoxymethyl-L-pGlu diazomethyl ketone (ZpGluDK). Illustrated using ChemSketch (Section 2.11).

1.3.9 Characteristics of PAP1

Biochemical studies and characterisation have been carried out on PAP1 enzymes from a range of prokaryotic and eukaryotic organisms. Table 1.6 has a summary of the reported PAP1 enzyme characteristics.

The molecular weight of PAP1 under denaturing and native conditions was determined mainly by SDS-PAGE and gel filtration respectively. The subunit size has mostly been reported as around 24 kDa, which is in good agreement with the sizes predicted from the nucleotide sequences, where available (Tables 1.2 and 1.4). *Sau*PAP1 and *Mbo*PAP1 were found to be slightly larger, attributed to six additional His residues fused to the N-terminal (*Sau*PAP1_{6H} and *Mbo*PAP1_{6H}, Section 1.3.5). *Efa*PAP1 and *Sfa*PAP1 were found to have subunit sizes of around 40 kDa.

The denatured size has been reported for few eukaryotic PAP1 enzymes but considering the high similarity in nucleotide sequence length (Tables 1.2 and 1.4) the assumption is made that their native sizes reflect monomeric enzymes. The larger native size of *Rno*PAP1 and *Gga*PAP1, together with the deduced monomeric size of 22.8 kDa for the former (Table 1.4) suggests that these enzymes occur in oligomeric form. The prokaryotic PAP1s seem to be consistently oligomeric.

More than one native size has been reported for *Bam*PAP1. The 72 kDa size refers to the enzyme purified from *B. amyloliquefaciens* while the size of 51 kDa refers to the recombinant enzyme expressed in *E. coli*. This seems to suggest a difference in oligomeric state between wild type and recombinant enzymes. The tertiary structure of recombinant *Bam*PAP1 has been reported as tetrameric (Section 1.3.10.1).

Table 1.6 Characteristics of various PAP1 enzymes

	PAP1 (Section 1.3.3)	Denatured size (Da)	Native size (Da)	Optimum PH	PI	Optimum Temp. (°C)	K _m (mM)
eukaryotic	<i>Hsa</i> PAP1 ^{r9 r10 r11 r22}	n.d.	24,000	7.0 – 9.5	n.d.	50	0.05 ^b
	<i>Bta</i> PAP1 ^{r17}	24,000	23,700	8.5	n.d.	n.d.	0.015 ^a , 0.021 ^b
	<i>Cpo</i> PAP1 ^{r7}	n.d.	24,000	n.d.	n.d.	n.d.	0.15 ^b
	<i>Ocu</i> PAP1 ^{r1}	n.d.	33,000	7.5	n.d.	n.d.	n.d.
	<i>Col</i> PAP1 ^{r1}	n.d.	33,000	7.5	n.d.	n.d.	0.13 ^a
	<i>Rno</i> PAP1 ^{r5}	n.d.	60,000	8.0 – 8.4	n.d.	n.d.	n.d.
	<i>Gga</i> PAP1 ^{r6}	n.d.	86,000	7.0 – 8.0	5.5	n.d.	0.73 ^a , 0.04 ^b
	<i>Pft</i> PAP1 ^{r15}	23,500	41,000	6.5 – 8.5	4.9	30	0.21 ^a
	<i>Bsu</i> PAP1 ^{r12}	25,200	91,000	6.8 – 7.5	n.d.	n.d.	1.04 ^a
	<i>Spy</i> PAP1 ^{r13}	23,500	85,000	7.0	n.d.	n.d.	1.79 ^a
	<i>Kcl</i> PAP1 ^{r2}	n.d.	74,000	6.0 – 7.5	n.d.	n.d.	n.d.
	<i>Efa</i> PAP1 ^{r18}	40,500	82,000	7.2 – 7.5	4.2	35 – 45	0.55 ^a
	<i>Sfa</i> PAP1 ^{r3}	42,000	n.d.	7.6	n.d.	n.d.	0.86 ^a
prokaryotic	<i>Bam</i> PAP1 ^{r4 r8 r14}	24,000	72,000 or 51,000	7.0 – 8.0	5.4	45	0.13 ^a
	<i>Tli</i> PAP1 ^{r20}	24,000	96,000	7.0 – 8.0	n.d.	70	n.d.
	<i>Pfu</i> PAP1 ^{r19}	22,937	96,300 or 45,643	6.0 – 9.0	n.d.	90	n.d.
	<i>Sau</i> PAP1 _{6H} ^{r16}	30,000	46,000	7.8	n.d.	n.d.	n.d.
	<i>Mbo</i> PAP1 _{6H} ^{r21}	29,000	46,000	n.d.	n.d.	n.d.	n.d.

[n.d. - not determined; a - determined for pGlu-β-NA; b - determined for pGlu-AMC; r1, Szewczuk and Kwiatkowska, 1970; r2, Kwiatkowska et al., 1974; r3, Sullivan et al., 1977; r4, Tsuru et al., 1978; r5, Busby et al., 1982; r6, Tsuru et al., 1982; r7, Browne and O’Cuinn, 1983; r8, Tsuru et al., 1984; r9, Lauffart and Mantle, 1988; r10 Lauffart et al., 1989; r11, Mantle et al., 1991; r12, Gonzales and Awadé, 1992; r13, Awadé et al., 1992; r14, Yoshimoto et al., 1993; r15, Gonzales and Robert-Baudouy, 1994; r16, Patti et al., 1995; r17, Cummins and O’Connor, 1996; r18, Mineyama and Saito, 1998; r19, Tsunasawa et al., 1998; r20, Singleton et al., 2000; r21, Kim et al., 2001; r22, Dando et al. 2003]

The two native sizes reported for *Pfu*PAP1 result from two separate methods for their determination. The 96.3 kDa size was determined by analytical ultracentrifugation while the 45.6 kDa size was determined by ionspray mass spectrometry. This variation in the determination of native size needs to be reflected on regarding all the native size data in this section. The tertiary structure of recombinant *Pfu*PAP1 has been reported as tetrameric (Section 1.3.10.1).

The optimal activity for all PAP1s falls into the pH range 6.0 to 9.5, with a mean of 7.5. The isoelectric point (pI), where reported, is around 5.0. The physiological temperature of 37°C has been widely used as the standard reaction temperature for analysis of eukaryotic PAP1. The optimum temperature for activity for mesophilic prokaryotic PAP1 has been reported ranging from 30 to 45°C. The two thermophilic enzymes *Tli*PAP1 and *Pfu*PAP1 exhibit optimum activity at 70 and 90°C respectively. The contribution of intersubunit disulphide bridges toward high thermostability is discussed in Section 1.3.10.1.

As well as the inhibitory effects of thiol blocking compounds and pGlu analogues mentioned in section 1.3.8, the inhibition of PAP1 by many cations such as Hg^{2+} , Zn^{2+} , Cu^{2+} , Co^{2+} , Ca^{2+} , Mn^{2+} , Mg^{2+} , Ni^{2+} , Ba^{2+} , Sr^{2+} and Cd^{2+} has been demonstrated (Szewczuk and Mulczyk, 1969; Albert and Szewczuk, 1972; Szewczuk and Kwiatkowska, 1974; Kwiatkowska *et al.*, 1974; Prasad and Peterkofsky, 1976; Tsuru *et al.*, 1978; Mantle *et al.*, 1991; Awadé *et al.*, 1992; Bharadwaj *et al.*, 1992; Patti *et al.*, 1995; Cummins and O'Connor, 1996; Mineyama and Saito, 1998; Tsunasawa *et al.*, 1998; Dando *et al.*, 2003).

Two compounds of low anti-microbial value, benarthin and pyrizinostatin, isolated from the genus *Streptomyces*, were found to be inhibitors of *Bta*PAP1 (Aoyagi *et al.*, 1992a; Aoyagi *et al.*, 1992b). Also, the inhibitory effect of an oligosaccharide gum from *Hakea gibbosa* on *Bta*PAP1 activity was reported (Alur *et al.*, 2001).

1.3.10 Structure of PAP1

Recently the three-dimensional structures of four prokaryotic PAP1 enzymes were solved using X-ray crystallography. These four structures; for the mesophilic *Bam*PAP1 (Odagaki *et al.*, 1999) and the thermophilic *Tli*PAP1 (Singleton *et al.*, 1999a; 1999b), *Pfu*PAP1 (Tanaka *et al.*, 2001) and *Pho*PAP1 (Sokabe *et al.*, 2002) have been deposited in the Protein Data Bank (Section 2.11) as structural coordinate files [1AUG], [1A2Z], [1IOF] and [1IU8] respectively. Crystals were obtained using the hanging drop vapour-diffusion method such as described in Section 2.19.7. The crystals diffracted to a resolution of 2.0, 1.73, 2.2 and 1.6 Å for *Bam*PAP1, *Tli*PAP1, *Pfu*PAP1 and *Pho*PAP1 respectively.

1.3.10.1 Quaternary structure of PAP1

The crystal structures of *Bam*PAP1, *Tli*PAP1, *Pfu*PAP1 and *Pho*PAP1 each have a tetrameric arrangement comprising of four identical subunits designated A to D (Figure 1.11). Each monomer of the tetramer makes contact with two other subunit monomers. These interfaces are named by using subunit A as a reference. The tetramers are approximately 80 Å along each side as indicated in Figure 1.11, with the monomers being approximately 50 Å at the longest axis.

The A-D interface of *Bam*PAP1 is illustrated in Figure 1.12. This interaction exists by burying 1427 Å² of hydrophobic surface from the surrounding medium, as well as having several ionic salt bridges, between each monomer. The A-C interaction (Figure

1.13) buries 1355 Å² of hydrophobic surface. It does not feature any ionic salt bridges and a thin layer of water mediates hydrogen bonds between the subunits. Conflicting values for native size of *Bam*PAP1 (Section 1.3.9) suggests that one of the above interactions may be dominant in solution and the tetrameric form of *Bam*PAP1 may therefore be an artefact of crystallisation. If this is the case, the extensive A-D interaction is more likely to be found in solution.

The *Tli*PAP1 tetramer has a central cavity of 6000 Å³, which is partially filled by a core structure comprised of loops from each subunit. This core structure is maintained by hydrophobic interactions involving Phe179, Phe180, Leu181 and Leu182 of each subunit (Figure 1.14). This FLL motif represents a notable variation from other PAP1 sequences (Figure 1.23). This hydrophobic core may contribute towards the thermostability of *Tli*PAP1. The A-B interface has hydrophobic interactions and also salt bridges involving Arg81, Asp88, Asp101 and Arg119 (Figure 1.15). The A-C interface is formed by an extended loop which folds towards the core structure. A disulphide bridge exists between Cys190 of each monomer (Figure 1.16). Other residues along this interface are generally hydrophobic.

The A-C interface of *Pfu*PAP1 consists mostly of hydrophobic interactions and no ionic interactions are present. The A-D interface consists of both hydrophobic interactions and 10 ionic bonds (Figure 1.17).

The A-B interface of the *Pho*PAP1 tetramer has inter-subunit ion bonds and hydrophobic interactions (Figure 1.18). The A-C interface of *Pho*PAP1 features hydrogen bonds that are entirely mediated by a thin layer of water (not shown). Gel filtration suggests that *Pho*PAP1 is a dimer in solution. This dimer is more than likely the A-B variation since this interface is more closely held.

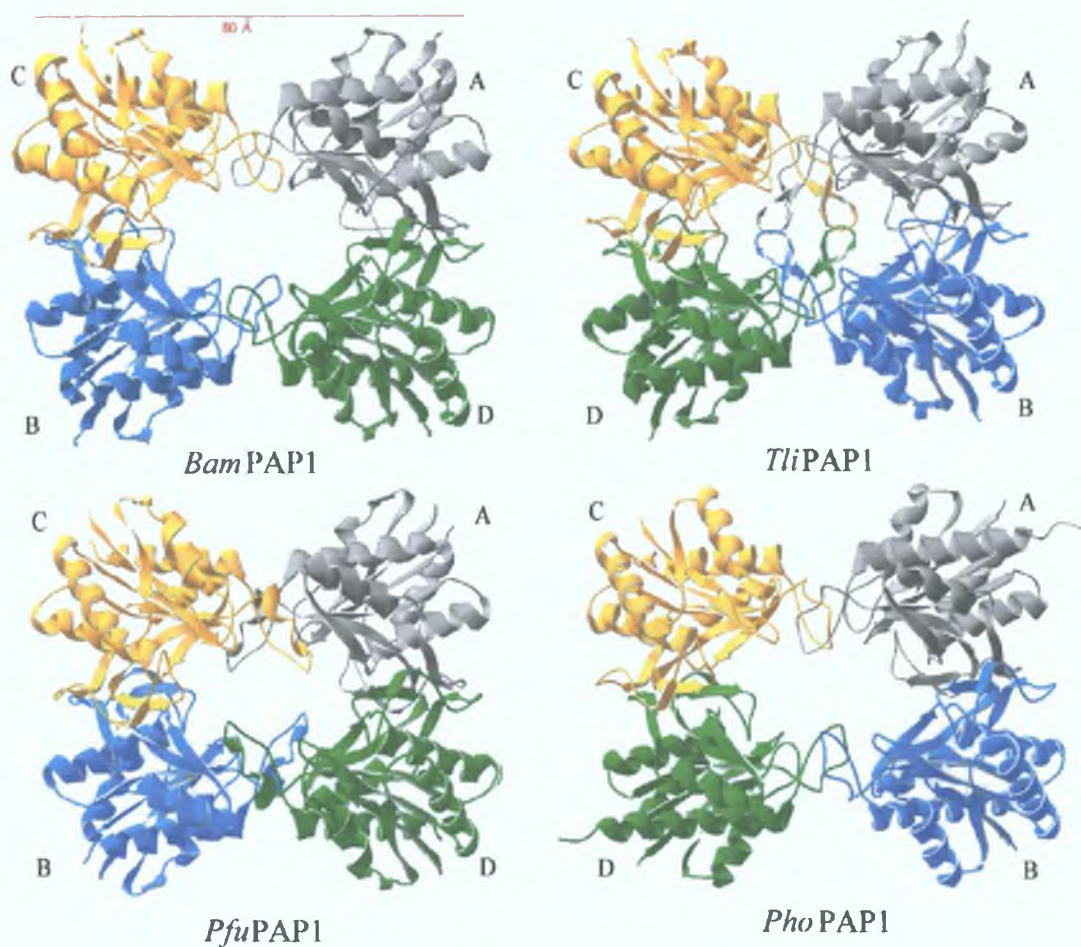


Figure 1.11 Tetrameric arrangement of PAP1.

Ribbon diagrams of the tetrameric crystal structures of *BamPAP1*, *TliPAP1*, *PfuPAP1* and *PhoPAP1*. The monomeric subunits are labelled A to D and individually coloured. To give an idea of size, the distance of 80 Å has been indicated for *BamPAP1*. Generated using DeepView (Section 2.11).

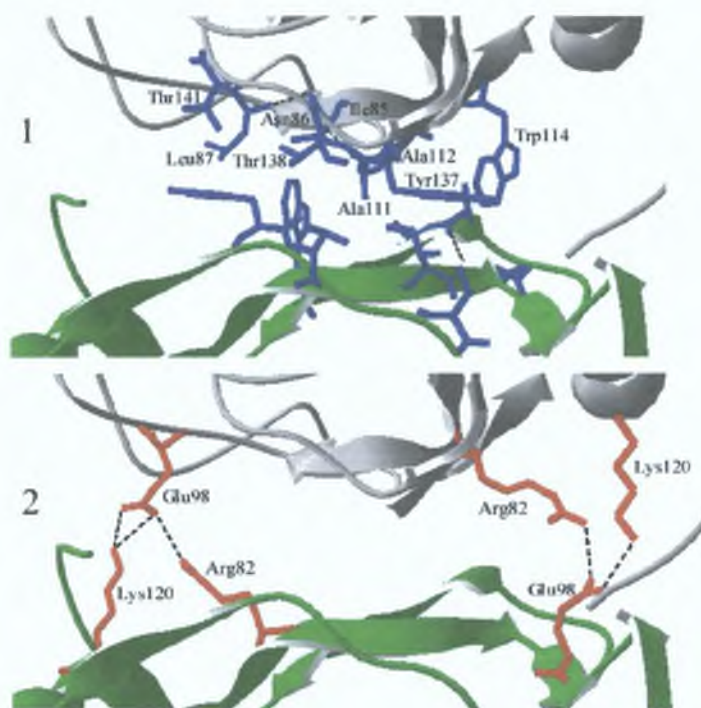


Figure 1.12 The A-D interface of *BamPAP1*.

Ribbon diagram of the A-D interface of *BamPAP1*. The same subunit colour coding as in Figure 1.11 applies. (1): Residues Ile85, Asn86, Leu87, Ala111, Ala112, Trp114, Tyr137, Thr138 and Thr141 of each monomer contributing to hydrophobic interactions are shown. Only the residues from subunit A have been labelled. (2): Residues Arg82, Glu98 and Lys120 of each monomer involved in ionic interactions are shown. For amino acid information see Appendix D. Generated using DeepView (Section 2.11).

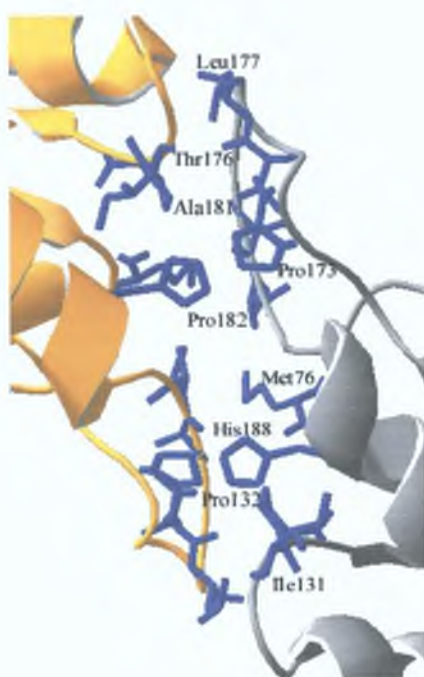


Figure 1.13 The A-C interface of *BamPAP1*

Ribbon diagram of the A-C interface of *BamPAP1*. The same subunit colour coding as in Figure 1.11 applies. Residues Met76, Ile131, Pro132, Pro173, Thr176, Leu177, Ala181, Pro182 and His188 contributing to hydrophobic interactions are shown. Only the residues from subunit A have been labelled. For amino acid information see Appendix D. Generated using DeepView (Section 2.11).

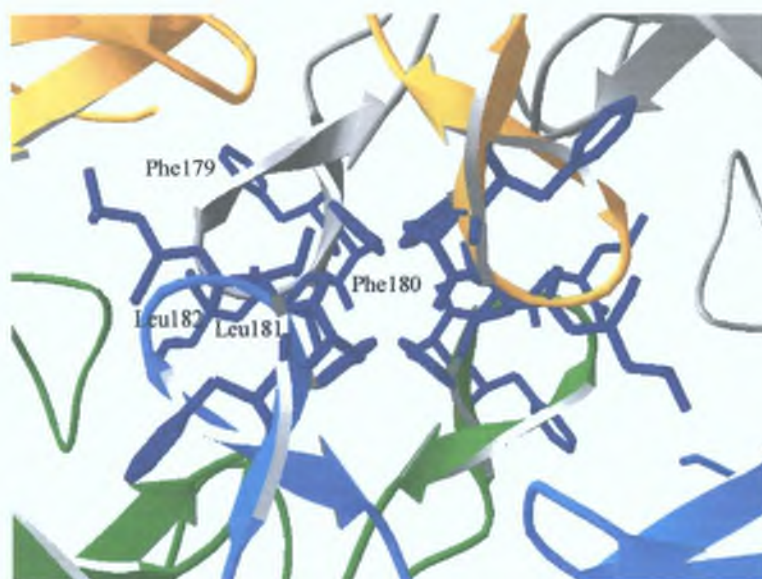


Figure 1.14 Core structure of *TliPAP1*

Ribbon diagram of the core structure of *TliPAP1*. The same subunit colour coding as in Figure 1.11 applies. The residues Phe179, Phe180, Leu181 and Leu182 involved in the hydrophobic core are shown. Only the residues from subunit A have been labelled. For amino acid information see Appendix D. Generated using DeepView (Section 2.11).

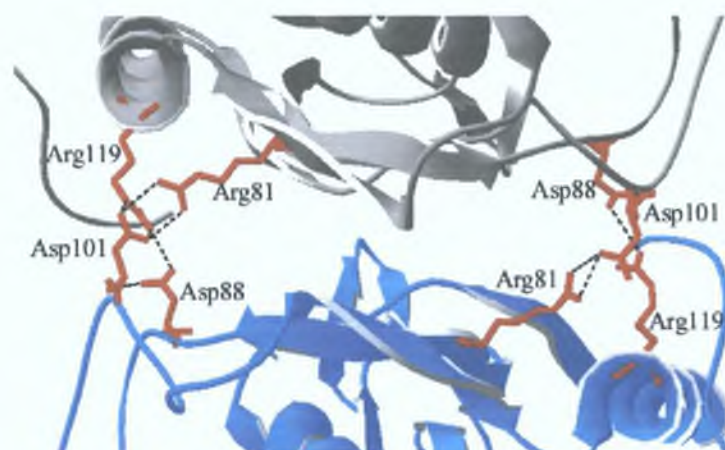


Figure 1.15 The A-B interface of *TliPAP1*

Ribbon diagram of the A-B interface of *TliPAP1*. The same subunit colour coding as in Figure 1.11 applies. Residues involved in ionic interactions are shown. For amino acid information see Appendix D. Generated using DeepView (Section 2.11).

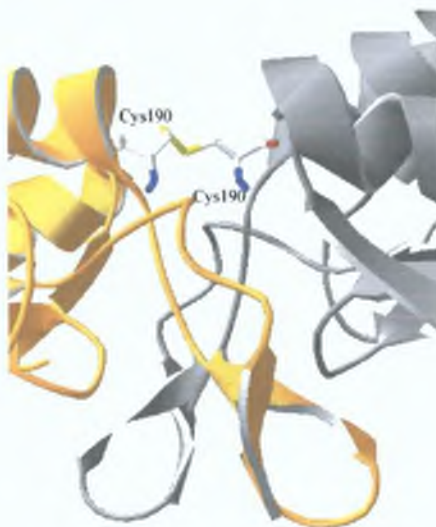


Figure 1.16 The A-C interface of *TliPAP1*

Ribbon diagram of the A-C interface of *TliPAP1*. The same subunit colour coding as in Figure 1.11 applies. The disulphide bridge is shown and coloured by CPK colour scheme (C = white, O = red, N = blue and S = yellow). For amino acid information see Appendix D. Generated using DeepView (Section 2.11).

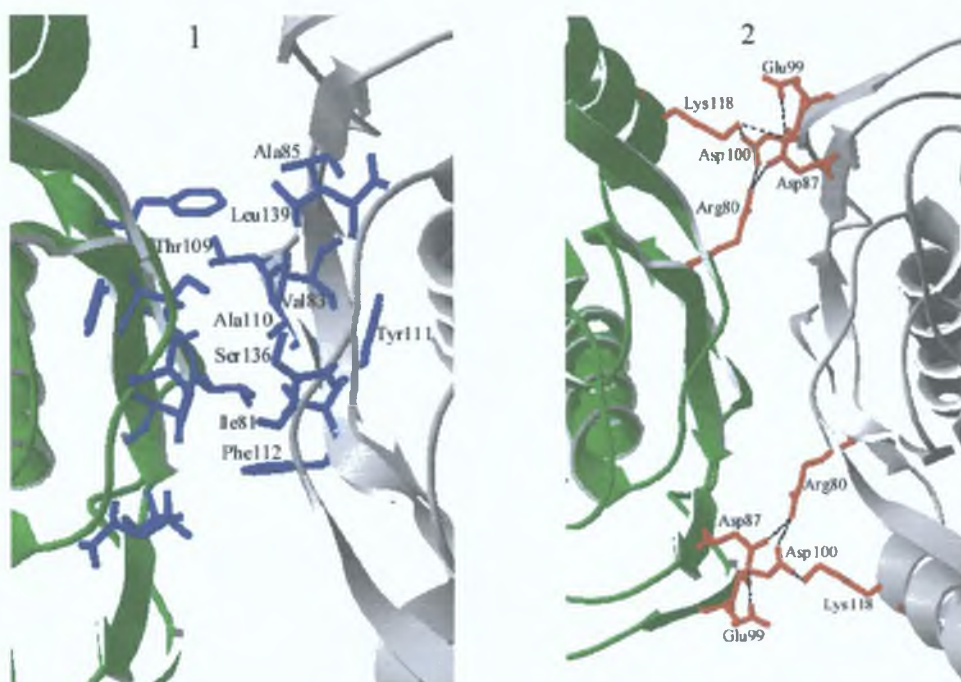


Figure 1.17 The A-D interface of *PfuPAP1*

Ribbon diagram of the A-D interface of *PfuPAP1*. The same subunit colour coding as in Figure 1.11 applies. (1): Residues contributing to hydrophobic interactions are shown. Only the residues from subunit A have been labelled. (2): Residues Arg80, Asp87, Glu99, Asp100 and Lys118 involved in ionic interactions are shown. For amino acid information see Appendix D. Generated using DeepView (Section 2.11).

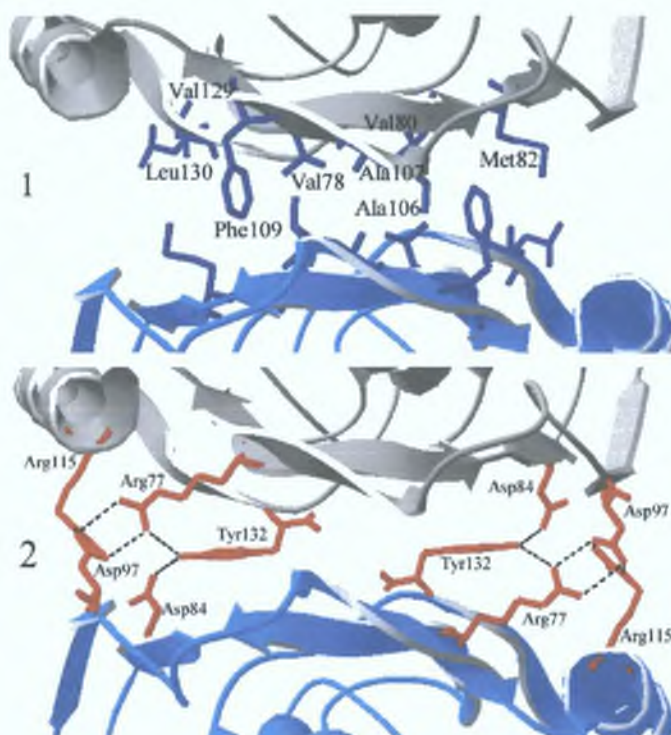


Figure 1.18 The A-B interface of *PhoPAP1*

Ribbon diagram of the A-B interface of *PhoPAP1*. The same subunit colour coding as in Figure 1.11 applies. (1): Residues Val78, Val80, Met82, Ala106, Ala107, Phe109, Val129 and Leu130 contributing to hydrophobic interactions are shown. Only the residues from subunit A have been labelled. (2): Residues Arg77, Asp84, Asp97, Arg115 and Tyr132 involved in ionic interactions are shown. For amino acid information see Appendix D. Generated using DeepView (Section 2.11).

1.3.10.2 Monomeric structure of PAP1

The subunit monomers of *BamPAP1*, *TliPAP1*, *PfuPAP1* and *PhoPAP1* (Figures 1.19 to 1.22 respectively) each fold into single α/β globular domains. The active site is located within each monomer and faces toward the central channel of the tetrameric structure (Figure 1.11). This type of arrangement has previously been described for human β -tryptase (Pereira *et al.*, 1998), which is active only in the tetrameric form.

The active site area of each monomer has a central β -sheet surrounded by α -helices. The catalytic cysteine is located at the N terminus of an α -helix. It has been speculated that the position of an active site nucleophile at the N-terminus of an α -helix can be important for catalysis as the helix dipole polarises the amide bond, enhancing reactivity (Doran and Carey, 1996). The sulfhydryl side-chain of this cysteine forms a hydrogen bond with the imidazolium ring of a histidine located on the β -sheet. The imidazolium ring is orientated/stabilised by a hydrogen bond to a glutamic acid residue, which is the third member of the Cys-His-Glu catalytic triad. A unique feature of *PhoPAP1* is that the side-chain NH of Arg70 is hydrogen bonded to the triad Glu76 backbone oxygen, which in turn bonds to the His163 imidazole ring (Figure 1.22).

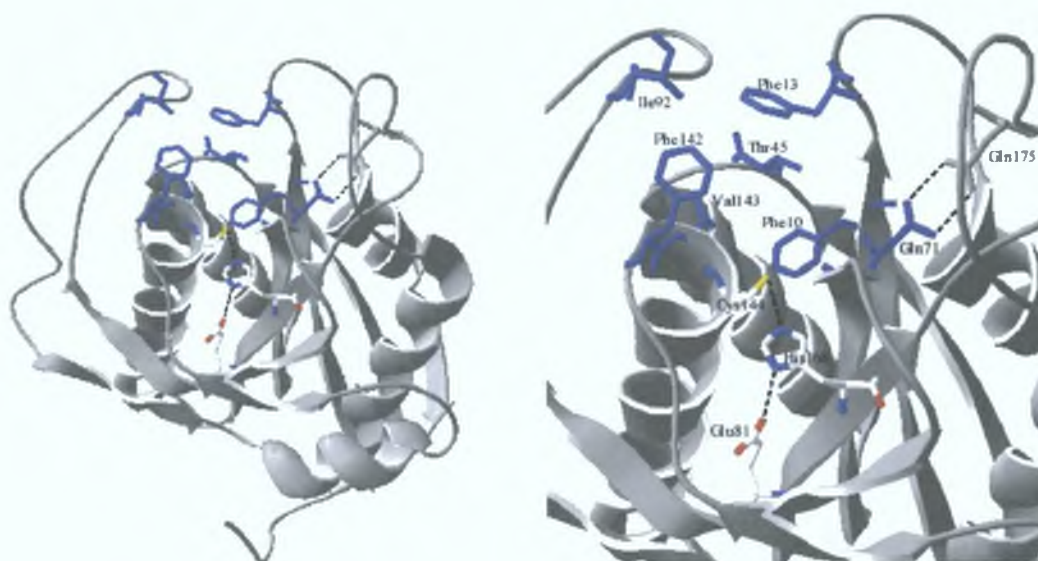


Figure 1.19 Subunit monomer of *BamPAP1* with detail of active site region

Ribbon diagram of the subunit monomer of *BamPAP1*. The entire monomer is shown on the left. On the right is a detail of the active site region. Residues Phe10, Phe13, Thr45, Gln71, Ile92, Phe142 and Val143 contributing to the hydrophobic pocket are coloured blue. The side-chain of Gln71 is oriented by hydrogen bonding to Gln175. The catalytic triad is coloured by CPK colour scheme (C = white, O = red, N = blue and S = yellow). For amino acid information see Appendix D. Generated using DeepView (Section 2.11).



Figure 1.20 Subunit monomer of *TliPAP1* highlighting active site

Ribbon diagram of the subunit monomer of *TliPAP1*. Residues contributing to the hydrophobic pocket are coloured blue while the catalytic triad is coloured by CPK colour scheme (C = white, O = red, N = blue and S = yellow). For amino acid information see Appendix D. Generated using DeepView (Section 2.11).

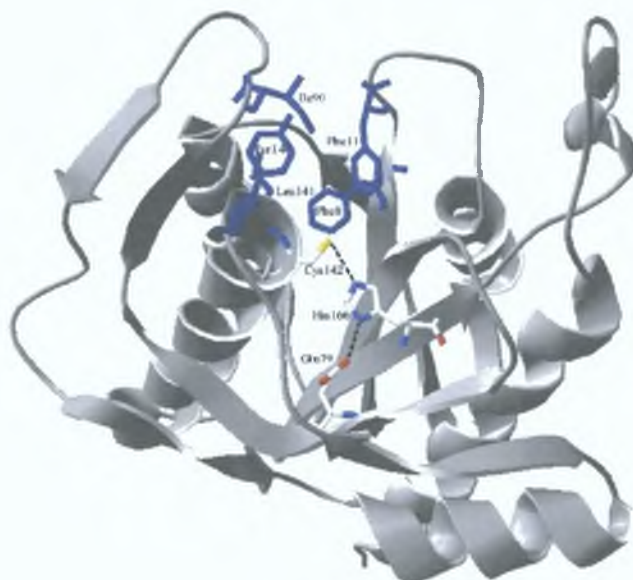


Figure 1.21 Subunit monomer of *Pfu*PAP1 highlighting active site

Ribbon diagram of the subunit monomer of *Pfu*PAP1. Residues contributing to the hydrophobic pocket are coloured blue while the catalytic triad is coloured by CPK colour scheme (C = white, O = red, N = blue and S = yellow). For amino acid information see Appendix D. Generated using DeepView (Section 2.11).



Figure 1.22 Subunit monomer of *Pho*PAP1 highlighting active site

Ribbon diagram of the subunit monomer of *Pho*PAP1. Residues contributing to the hydrophobic pocket are coloured blue while residues involved in catalysis are coloured by CPK colour scheme (C = white, O = red, N = blue and S = yellow). For amino acid information see Appendix D. Generated using DeepView (Section 2.11).

In *Bam*PAP1 a hydrophobic pocket comprising Phe10, Phe13, Thr45, Gln71, Ile92, Phe142 and Val143 is found close to the catalytic triad (Figure 1.19). With the exception of Thr45 and Gln71 these residues are conserved or conservatively substituted in *Tli*PAP1, *Pfu*PAP1 and *Pho*PAP1 (Figure 1.23) and similarly located in the active site region (Figures 1.19 to 1.22). The feasibility of this *Bam*PAP1 binding pocket was investigated by manually placing a pGlu-His model into the pocket (Odagaki *et al.*, 1999) using molecular dynamics simulation software (not shown). In this simulation, two methylene carbon atoms of the pGlu pyrrolidone ring point inward to the hydrophobic pocket. The pyrrolidone ring itself is held between the aromatic ring of Phe13 and the methylene chain of Gln71 while the ring carbonyl points away from the pocket. The pGlu main chain carbonyl, the His imidazole nitrogen and the C-terminal oxygen of the model form hydrogen bonds within the pocket. There does not appear to be any other clearly defined specificity pocket, supporting that PAP1 is not selective for residues in the second position.

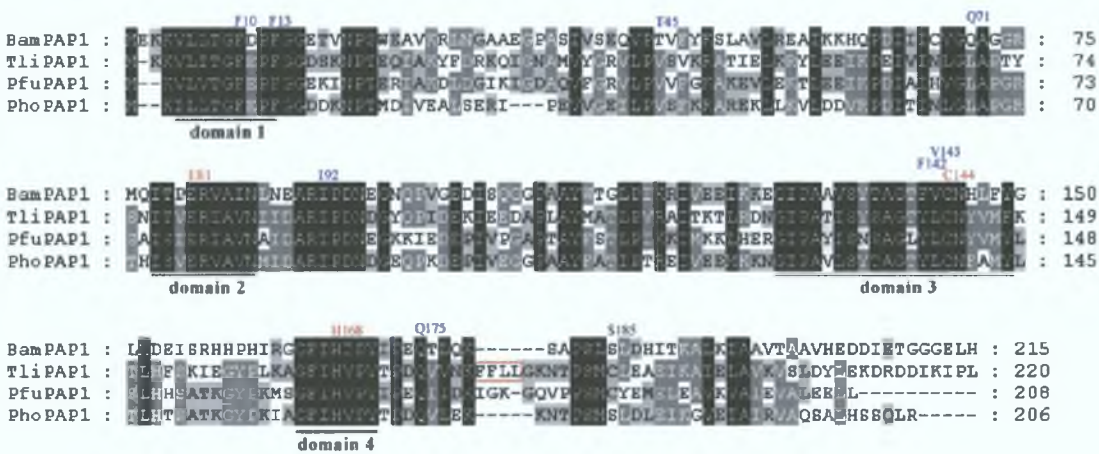


Figure 1.23 Amino acid alignment of *Bam*PAP1, *Tli*PAP1, *Pfu*PAP1 and *Pho*PAP1
Alignment was created using MultAlin with Blosum62-12-2 parameters and edited using GenDoc (Section 2.11). Sequence homology is represented by grey scale shading, with black being the highest homology. The domains introduced in Figure 1.4 are shown. Several functionally relevant residues have been marked in red. The FFL domain specific to *Tli*PAP1 is boxed in red (numbering applies to *Bam*PAP1). For amino acid information see Appendix D.

The Mechanism of substrate recognition was observed (Ito *et al.*, 2001) by X-ray crystallography of the substrate-inhibitor complex of *Bam*PAP1 and aldehyde pGlu analog 5-oxoprolinal (Figure 1.9). The resulting structure data reveals direct interaction of the pyroglutamine with Cys144. Two hydrogen bonds are observed between the enzyme and the pyrrolidone ring carbonyl and nitrogen of 5-oxoprolinal. These

orientate the molecule for nucleophilic attack by Cys144. Van der Waals interaction between the pyrrolidone ring and the hydrophobic pocket contribute also. By comparing the wild type enzyme structure with the substrate-inhibitor complex a particularly large movement of the Phe13 and Gln71 is observed, acting as an induced fit mechanism. While Phe13 is conserved throughout species (Figures 1.4, 1.5, 1.6 and 1.23) Gln71 is substituted with leucine in some prokaryotes and valine in eukaryotes. The benzene rings of Phe10, Phe13 and Phe142 are almost parallel with the 5-oxoprolinal pyrrolidone ring.

As mentioned in Section 1.3.6, no sequence homology is observed with proteins other than PAP1s from different organisms. However, an alignment of protein structure distance matrices (Odagaki *et al.*, 1999) shows close structural identity with *E. coli* purine nucleoside phosphorylase (*EcoPNP*, Figure 1.24). Several structural domains that have very similar orientation are represented by colouration. These domains also have the same order in the primary sequence. The catalytic domains of the two proteins, however, are orientated very differently (not shown). This structural similarity was also reported for *TliPAP1* (Singleton *et al.*, 1999).

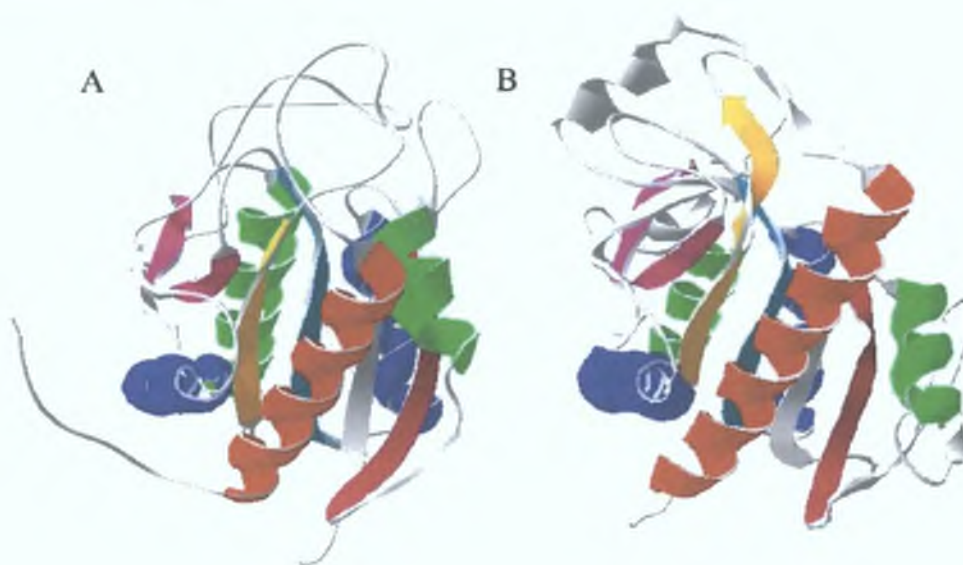


Figure 1.24 Structural comparison of *BamPAP1* with *EcoPNP*

Ribbon diagram comparison of *BamPAP1* (A) and *EcoPNP* (B). Regions of structural similarity are indicated by colouration. Generated using DeepView (Section 2.11).

A similar structural homology is observed with cutinase from *Fusarium solani* (Odagaki *et al.*, 1999), having a catalytic triad Ser120, His188 and Asp175. Ser120 is in a similar position as Cys144 in PAP1. However, the linear order of the catalytic triad in cutinase is Ser-Asp-His as opposed to Glu-Cys-His in PAP1. As discussed by Murzin (1993) the

natural evolution of enzymes can lead to new catalytic functions. A new catalytic active site can form on the existing protein scaffold in which case the positions of the active sites will have no similarity. Alternatively, the existing active site can transform to a new catalytic function.

1.3.11 Mutational analysis of PAP1 structure and function

In addition to the mutational analysis carried out to determine the catalytic triad (Section 1.3.7), several studies have used site-specific mutagenesis to probe the relationship between structure and function of PAP1.

The three *Bam*PAP1 residues Phe10, Phe13 and Phe142 contributing to the hydrophobic pocket (Section 1.3.10.2) were mutated to Tyr and Ala (Ito *et al.*, 2001). *Bam*PAP1_F13mY and *Bam*PAP1_F142mY had kinetic parameters similar to the wild type. The K_m for *Bam*PAP1_F10mY was 3.6-fold higher than wild type. The catalytic efficiency of both *Bam*PAP1_F13mA and *Bam*PAP1_F142mA decreased more than 1000-fold while *Bam*PAP1_F10mA seemed to undergo severe structural change and could not be purified. Phe10 plays an essential role in the hydrophobic pocket of *Bam*PAP1.

The high thermal stability of *Tli*PAP1 (Section 1.3.9) has been attributed to the inter-subunit disulphide bond formed by Cys190 (Figure 1.16), which reduces denaturation entropy (Singleton *et al.*, 1999). By sequence alignment (Figure 1.23) this residue corresponds to Cys188 in *Pfu*PAP1, Ser181 in *Pho*PAP1 and Ser185 in *Bam*PAP1. Cys188 of *Pfu*PAP1 is located at a similar position to Cys190 in *Tli*PAP1 but *Pfu*PAP1 does not exhibit a disulphide bridge in the crystal structure. However, a disulphide bridge in *Pfu*PAP1 involving Cys188 has been shown by differential scanning calorimetry (Ogasahara *et al.*, 2001). The denaturation temperature was 10°C higher in the presence of this disulphide bridge as shown by comparison with a mutant having both Cys142 and Cys188 changed to Ser (*Pfu*PAP1_C142,C188mS). *Pho*PAP1 does not feature a hydrophobic tetramer core as seen in *Tli*PAP1 or a disulphide bridge as seen in *Pfu*PAP1 and *Tli*PAP1. The thermostability of *Pho*PAP1 owes to other interactions and optimal folding conformation (Sokabe *et al.*, 2002).

To investigate the possibility of enhancing the thermal stability of *Bam*PAP1, Ser185 was changed to Cys by site-specific mutagenesis (*Bam*PAP1_C185mS, Kabashima *et al.*, 2001). By analysing *Bam*PAP1_C185mS by SDS-PAGE with and without 2-ME, it was confirmed that a disulphide bond had been formed. The mutation did not affect

catalytic efficiency but thermal stability was increased by 30°C. However, under conditions of 1 mM DTT no difference in thermal stability was determined. Therefore, the enhanced thermal stability was attributed to the disulphide bond.

The thermal stability of *Pfu*PAP1 is very high compared to *Bam*PAP1 and slightly higher than for *Tli*PAP1 (Section 1.3.9). Comparatively, *Pfu*PAP1 has more extensive ion pairing and more compact hydrophobic interactions than *Tli*PAP1 and in turn *Bam*PAP1. By stability profiling it was determined that the high stability of *Pfu*PAP1 is due to increased hydrophobic interaction and ion-pairs as well as an improvement of the overall conformation (Ogasahara *et al.*, 2001). The conformational structure is superior in the combination of positive and negative factors for stabilisation. The denaturation temperature of monomeric *Pfu*PAP1 at pH 2.5 (at which *Bam*PAP1 was completely denatured) was similar to that of tetrameric *Bam*PAP1 at pH 9.0 indicating that the subunit monomer of *Pfu*PAP1 has intrinsic enhanced stability. The comparative guanidine hydrochloride-induced unfolding and refolding of *Pfu*PAP1_C142,C188mS and *Bam*PAP1 was carried out using circular dichroism (CD) measurement at 220 nm (Ogasahara *et al.*, 1998). The rate of unfolding for *Pfu*PAP1_C142,C188mS was seven orders of magnitude slower compared to *Bam*PAP1 while the refolding rates were similar. This slow unfolding rate is a characteristic of the greater stability of *Pfu*PAP1.

1.4 PAP2

The Enzyme Commission list (Section 1.3.2) classifies PAP2 as an omega peptidase (EC 3.4.19.6). In the alternative MEROPS database (Section 1.3.2) PAP2 has been assigned under metallopeptidase clan MA, family M1. Metallopeptidases are hydrolases where the nucleophilic attack is mediated by a water molecule activated by a divalent metal cation (Barrett *et al.*, 199). The metal ion (usually zinc) is held in place by three amino acid ligands.

The metallopeptidase PAP2, originally purified from the synaptosomal membranes of mammalian brain (O'Connor and O'Cuinn, 1984), has been comprehensively reviewed (Cummins and O'Connor, 1998; Barrett *et al.*, 1998). In summary, PAP2 is primarily located in the mammalian central nervous system with significantly smaller levels observed in other tissues. It appears to be associated exclusively with neurons, presumably on postsynaptic membranes. It has an absolute specificity toward TRH (Section 1.2). Some very closely related peptides (tripeptides, tripeptide amides and tetrapeptides commencing with pGlu-His) are also hydrolysed, albeit at a considerably

reduced rate. This high specificity has resulted in postulation that membrane anchored PAP2 is responsible for selectively inactivating neuronally released TRH within the extracellular vicinity of target cells, thus defining it as a “neuropeptide-specific peptidase”. PAP2 was found to have its active site directed toward the extracellular space where TRH would be located when involved in neurotransmission. Inhibition of PAP2 specifically increases recovery of TRH released from brain tissue. PAP2 has a broad pH optimum in the neutral range and is sensitive to metal chelators such as EDTA. It exhibits no sensitivity to sulphhydryl-blocking agents or pGluDK, an active site directed inhibitor of PAP1 (Section 1.3.8). PAP2 has a large relative molecular mass of 230 kDa and was shown to be composed of two identical 116 kDa subunits, a dimeric structure common for membrane anchored ectoenzymes.

The genes for rat and human PAP2 were cloned (Schauder *et al.* 1994; Schomburg *et al.*, 1999). The deduced highly homologous amino acid sequences are consistent with a glycosylated, membrane anchored peptidase. The extracellular domain of PAP2 contains the HEXXH + E motif which is consistent with zinc-dependent metallopeptidases where the three zinc ligands (mentioned above) are the two His residues and the Glu residue C-terminal to the HEXXH motif.

A serum form of PAP2 of liver origin has been reported, which has the same degree of specificity for TRH and identical biochemical characteristics as the membrane-bound form (Cummins and O'Connor, 1998). A recent study (Schmitmeier *et al.*, 2002) has supported the hypothesis that both forms of PAP2 are derived from the same gene, whereby the serum enzyme is generated by proteolytic cleavage of the membrane-bound form in the liver.

1.5 Project aims & objectives

The opening objective of this research project was to clone the gene for human PAP1. The putative human PAP1 sequence (*Hsa-pap1*, Table 1.4), which became available on GenBank as this work began, would provide a suitable starting point for the design of oligonucleotide primers to be used in a PCR-based strategy for the amplification of the gene from human RNA.

Upon cloning of the gene the project would focus on the development of an expression system, capable of producing catalytically active recombinant human PAP1, which had heretofore not been reported. If expression of active recombinant human PAP1 was possible in a prokaryotic system, this would enable the rapid production of large

quantities of the enzyme. Alternatively, a mammalian or yeast-based expression system would be explored.

Following successful expression of the recombinant enzyme, a strategy would be developed to yield highly purified protein. This could be achieved by conventional methodology combining ammonium sulphate fractionation, gel filtration, anion exchange, hydrophobic interaction and thiol-affinity chromatography. An alternative option is to make use of a recombinant fusion protein tagging system, enabling one-step affinity-tag purification. Both of these approaches have been previously reported for other PAP1 enzymes, as described in Section 1.3.5.

The availability of purified recombinant human PAP1 would allow a series of detailed molecular, biochemical and kinetic studies to be carried out.

Finally, using the recombinant PAP1 expression system, a series of site-specific mutagenesis experiments would be carried out, giving valuable insight into the catalytic and structural properties of human PAP1 and opening the way for the design of inhibitors and pharmacological agents. Such studies would significantly add to the growing amount of published data on PAP1 and represent the first detailed study of the human form of this unique enzyme.

2.0 Materials & Methods

2.1 Bacterial strains, primers and plasmids

The bacterial strains, primers and plasmids used in this work are listed in Tables 2.1, 2.2 and 2.3 respectively.

Table 2.1 Bacterial strains

Strain	Genotype	Features/Uses	Source
<i>Escherichia coli</i>			
DH5α	F ⁻ <i>endA1 recA1 relA1 gyrA96 supE44 thi-1 hsdR17</i> (r _K ⁻ , m _K ⁺) Φ80 <i>lacZ</i> ΔM15 Δ(<i>lacZYA-argF</i>)U169 <i>deoR phoA λ</i> ⁻	High transformation efficiency	Bethesda Research Labs
INVαF'	F ⁻ <i>endA1 recA1 relA1 gyrA96 supE44 thi-1 hsdR17</i> (r _K ⁻ , m _K ⁺) Φ80 <i>lacZ</i> ΔM15 Δ(<i>lacZYA-argF</i>)U169	High transformation efficiency	Invitrogen
XL-10 Gold	Δ(<i>mcrA</i>)183 Δ(<i>mcrCB-hsdSMR-mrr</i>)173 <i>endA1 recA1 relA1 gyrA96 supE44 thi-1 lac Hte</i> [F ['] <i>proAB lacI</i> ^q ΔM15 ::Tn10(<i>tet</i> ^R) <i>Amy (cam</i> ^R)]	High transformation efficiency Antibiotic resistance Expression host	Stratagene
BL21	F ⁻ <i>dcm ompT hsdS</i> _B (r _B ⁻ , m _B ⁻) <i>gal</i>	Protease deficient Negative control for Rosetta	Novagen
Rosetta	F ⁻ <i>dcm ompT hsdS</i> _B (r _B ⁻ , m _B ⁻) <i>gal</i> pRARE[<i>argU argW ileX glyT leuW proL (cam</i> ^R)]	BL21 with pRARE plasmid tRNAs for rare codons	Novagen
Nova Blue	F ['] [<i>proAB lacI</i> ^q ΔM15 ::Tn10(<i>Tet</i> ^R)] <i>endA1 recA1 relA1 gyrA96 supE44 thi-1 hsdR17</i> (r _{K12} ⁻ ,m _{K12} ⁺) <i>lac</i>	Antibiotic resistance Expression host Negative control for Rosetta Blue	Novagen
Rosetta Blue	F ['] [<i>proAB lacI</i> ^q ΔM15 ::Tn10(<i>Tet</i> ^R)] <i>endA1 recA1 relA1 gyrA96 supE44 thi-1 hsdR17</i> (r _{K12} ⁻ ,m _{K12} ⁺) <i>lac</i> pRARE[<i>argU argW ileX glyT leuW proL (cam</i> ^R)]	NovaBlue with pRARE plasmid tRNAs for rare codons Antibiotic resistance	Novagen

Table 2.2 Primers (obtained from MWG-Biotech AG)

Name	Sequence (5' – 3')	T _m (°C)
Cloning/Analysis		
PAPHsA	AACAGAAGCAGGTCCGAGG	58.8
PAPHsB	CAGGATGAGGTCTTAGGAGA	57.3
PAPHsF	ATGGAGCAGCCGAGGAAGGCG	65.7
PAPHsG	GAGCGTCCCTCAGTGTGTTGTGGC	66
PAPHsH	GCCATGGAGCAGCCGAGGAAG	65.7
PAPHsI	<u>GGATCC</u> GTGTTTGTGGCAATAGTTGATTTTGCC	61.9
PAPHsK	CACACAGGAAACAGAATTCATTAAAG	58.5
PAPHsL	CGCGGATCCT TAGTGTTTGTGGCAATAGTTGATTTTG	58.1
PAPHsM	ATGGAGCAGCCGAGGAAGGCG	65.7
PAPHsN	GTGTTTGTGGCAATAGTTGATTTTGCCC	63.7
PAPHsR	CGCGGATCCT TACTCTGACTGCTCCAGGAGGTCC	65.8
PAPHsS	CCGCCAAAACAGAAAGCTTAGTGATGG	64.8
PAPHsT	CGCGGATCCT CTGACTGCTCCAGGAGGTCC	65.9
Amp-for	GAGTATTCAACATTTCCGTGTCCG	61
Amp-rev	CCAATGCTTAATCAGTGAGGCAC	60.7
pKK-testA	GTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCG	77.9
Non-binding tags are in bold type. Restriction sites are underlined.		
T _m takes into account only those bases that bind.		
T _m = [69.3 + 0.41(%GC)] – 650/length		

Mutagenesis		
PAPmR5K6-f	CCATGGAGCAGCCGCGTAAAGCGGTGGTAGTGAC	78.2
PAPmR5K6-r	GTCACTACCACCGCTTTACGCGGCTGCTCCATGG	78.2
PAPmR5a-f	CCATGGAGCAGCCGCGCAAGGCGGTGGTAGTGAC	83.5
PAPmR5a-r	GTCACTACCACCGCCTTGCGCGGCTGCTCCATGG	83.5
PAPmR5b-f	CCATGGAGCAGCCGCGTAAAGGCGGTGGTAGTGAC	82.3
PAPmR5b-r	GTCACTACCACCGCCTTACGCGGCTGCTCCATGG	82.3
PAPmF13W-f	GTAGTGACGGGATG <u>GGGCC</u> TTTGGGGAAC	78.4
	<i>Apal</i>	
PAPmF13W-r	GTTCCCCAAAAGGGCCCATCCCGTCACTAC	78.4
	<i>Apal</i>	
PAPmF13Y-f	GTGACGGGATATGGCCCTTTTGGCGAACACACCG	79.9
	<i>BglI</i>	
PAPmF13Y-r	CGGTGTGTTCCGCCAAAAGGGCCATATCCCGTCAC	79.9
	<i>BglI</i>	
PAPmF13L-f	GTAGTGACGGGATTGGGCCCTTTTGGGGAAC	80.3
	<i>Apal</i>	
PAPmF13L-r	GTTCCCCAAAAGGGCCCAATCCCGTCACTAC	80.3
	<i>Apal</i>	
PAPmF13L_F16L-f	GACGGGATTGGGCCCTTTGGGGGAACACAC	78.3
	<i>Apal</i>	
PAPmF13L_F16L-r	GTGTGTTCCCCCAAAGGGCCCAATCCCGTC	78.3
	<i>Apal</i>	
PAPmF16Y-f	GGATTTGGCCCTTATGGCGAACACACCGTGAACG	78.7
	<i>BglI</i>	
PAPmF16Y-r	CGTTCACGGTGTGTTCCGCATAAGGGCCAAATCC	78.7
	<i>BglI</i>	
PAPmF16L-f	GTGACGGGATTTGGCCCTCTGGGCGAACACACCGTGAAC	81.7
	<i>BglI</i>	
PAPmF16L-r	GTTACACGGTGTGTTCCGCCAGAGGGCCAAATCCCGTCAC	81.7
	<i>BglI</i>	
PAPmD97E-f	GGGCTGGAGAACTGCCGCTTTTGCCCCGGATCCCAGTGC	86.4
	<i>BamH1</i>	
PAPmD97E-r	GCACTGGGATCCGGGGCAAAGCGGCAGTTCTCCAGCCC	86.4
	<i>BamH1</i>	
PAPmN98V-f	GGGCTGGACGTCTGCCGCTTTTGCCCCGGATCCCAGTGC	84.9
	<i>BamH1</i>	
PAPmN98V-r	GCACTGGGATCCGGGGCAAAGCGGCAGACGTCCAGCCC	84.9
	<i>BamH1</i>	
PAPmN98Q-f	GGGCTGGACCAGTGCCGCTTTTGCCCCGGATCCCAGTGC	84.9
	<i>BamH1</i>	
PAPmN98Q-r	GCACTGGGATCCGGGGCAAAGCGGCAGTGGTCCAGCCC	84.9
	<i>BamH1</i>	
PAPmSID115AYF-f	GCGTGGAGGACGGGCCTGAGGCCTATTTCTCCATCATCGACATGGATGC	77.7
	<i>StuI</i>	
PAPmSID115AYF-r	GCATCCATGTCGATGATGGAGAAATAGGCCTCAGGCCCGTCCTCCACGC	77.7
	<i>StuI</i>	
PAPmD117+GGG-f	GGCCTGAAAGCATTGACGGCGGC <u>GGATCC</u> ATCATCGACATGGATG	70.2
	<i>BamH1</i>	
PAPmD117+GGG-r	CATCCATGTCGATGATGGATCCGCGCCGCTCAATGCTTTCAGGCC	70.2
	<i>BamH1</i>	
PAPmI120G-f	GCCTGAAAGCATTGACTCGATCGGCGACATGGATGCTGTGTGC	81.7
	<i>PvuI</i>	
PAPmI120G-r	GCACACAGCATCCATGTCCGCGATCGAGTCAATGCTTTCAGGC	81.7
	<i>PvuI</i>	
PAPmY147F-f	GCAGGATGCCGGCAGATTTCTCTGCGACTTTACCTAC	82.7
PAPmY147F-r	GTAGGTAAAGTCGAGAGAAATCTGCCGGCATCCTGC	82.7
PAPmC149W-f	GATGCCGGCAGATACCTCTGGGACTTTACCTACTACAC	80.1
PAPmC149W-r	GTGTAGTAGGTAAAGTCCCAGAGGTATCTGCCGGCATC	80.1
PAPmH168D-f	CGGT <u>CGATCG</u> GCCTTCGTGACGTGCC	79.5
	<i>PvuI</i>	

PAPmH168D-r	GGGCACGTCGACGAAGGCCGATCGACCG	79.5
	<i>Pvu</i> I	
PAPmQ180E-f	CGTACAACGCGGACGAGCTCGGCAGGGCACTG	82.3
	<i>Sac</i> I	
PAPmQ180E-r	CAGTGCCCTGCCGAGCTCGTCCGCGTTGTACG	82.3
	<i>Sac</i> I	
Mutation-inducing mismatches are in bold type.		
Diagnostic restriction sites are underlined.		
$T_m = [81.5 + 0.41(\%GC)] - 675/\text{length} - \% \text{ mismatch}$		
Sequencing		
M13rev(-29)	CAGGAAACAGCTATGACC	(pCR2.1 forward primer, MWG-Biotech)
M13uni(-21)	TGTAAAACGACGGCCAGT	(pCR2.1 reverse primer, MWG-Biotech)
PQEfor	GTATCACGAGGCCCTTTCGTCT	(pQE-60 forward primer, MWG-Biotech)
PQErev	CATTACTGGATCTATCAACAGGAG	(pQE-60 reverse primer, MWG-Biotech)
PTrcHisrev	CTTCTGCGTTCTGATTTAATCTG	(pPC225 forward primer, MWG-Biotech)
M13rev(-49)	GAGCGGATAACAATTTACACAGG	(pPC225 reverse primer, MWG-Biotech)
pKK-for	CGGCTCGTATAATGTGTGG	(pPC225 forward primer, Qiagen)
pKK-rev	GAAAATCTTCTCTCATCCGCC	(pPC225 reverse primer, Qiagen)

Table 2.3 Plasmids

Plasmid	Description	Source
Vectors		
pCR2.1	TA cloning vector: <i>Plac</i> , <i>amp</i> ^R , <i>kan</i> ^R , <i>lacZ</i> _α , <i>ColE1</i> origin. Figure 2.1	Invitrogen
pQE-60	Expression vector: T5 promoter/ <i>lac</i> operon, <i>amp</i> ^R , 6xHis sequence at 3' end of MCS, <i>ColE1</i> origin. Figure 2.2	Qiagen
pPC223	Modified derivative of pKK223-3 expression vector from Amersham Pharmacia: <i>Ptac</i> , <i>amp</i> ^R , <i>ColE1</i> origin. Elimination of BamHI site external to MCS	Clarke (2000)
pPC225	Modified derivative of pPC223. Insertion of 1800 bp fragment in BamHI of MCS for ease of cloning. Figure 2.3	Clarke (2000)
pRARE	Plasmid carrying genes for tRNAs that recognise codons considered rare in <i>E. coli</i> : <i>ileX</i> (AUA), <i>argU</i> (AGG/AGA), <i>argW</i> (AGG), <i>glyT</i> (GGA), <i>leuW</i> (CUA) and <i>proL</i> (CCC). <i>cam</i> ^R , p15a origin. Figure 2.4	Novagen
pUC18	Cloning vector: <i>Plac</i> , <i>amp</i> ^R , <i>lacZ</i> _α , <i>ColE1</i> origin.	Amersham Pharmacia
Constructs		
pRV1	pCR2.1 containing PAPI ORF	This Work
pRV3	pQE-60 containing PAPI ORF with 3' 6xHis sequence fusion, for expression and subsequently purification. (Erroneous deletion within promoter sequence).	This Work
pRV8	As pRV3 (without erroneous deletion)	This Work
pRV5	pPC225 containing PAPI ORF with 3' 6xHis sequence fusion, sub-cloned from pRV3, for expression and subsequently purification	This Work
pRV5_Δ6H	Modified derivative of pRV5, deletion of the 3' 6xHis sequence	This Work
pRV5 mutant derivatives		
pRV5_R5 ^S K6 ^S	Silent mutations: R5 (AGG to CGT), K6 (AAG to AAA)	This Work
pRV5_R5 ^S a	Silent mutation: R5 (AGG to CGC)	This Work
pRV5_R5 ^S b	Silent mutation: R5 (AGG to CGT)	This Work
pRV5_F13mW	Mutation: F13 to W (TTT to TGG)	This Work
pRV5_F13mY	Mutation: F13 to Y (TTT to TAT)	This Work
	Silent mutation: G17 (GGG to GGC)	
pRV5_F13mY.P175mS	Mutation: F13 to Y (TTT to TAT), P175 to S (CCG to TCG)	This Work
	Silent mutation: G17 (GGG to GGC)	

pRV5_F13mL	Mutation: F13 to L (TTT to TTG)	This Work
pRV5_F13mL.F16mL	Mutation: F13 to L (TTT to TTG), F16 to L (TTT to TTG)	This Work
pRV5_F16mY	Mutation: F16 to Y (TTT to TAT)	This Work
	Silent mutation: G17 (GGG to GGC)	
pRV5_F16mL	Mutation: F16 to L (TTT to CTG)	This Work
	Silent mutation: G17 (GGG to GGC)	
pRV5_D97mE	Mutation: D97 to E (GAC to GAG)	This Work
	Silent mutation: G104 (GGC to GGA)	
pRV5_D97mE.N98C99mM	Mutation: D97 to E (GAC to GAA), N98 and C99 to M (AAC TGC to ATG)	This Work
	Silent mutation: G104 (GGC to GGA)	
pRV5_N98mV	Mutation: N98 to V (AAC to GTC)	This Work
	Silent mutation: G104 (GGC to GGA)	
pRV5_N98mQ	Mutation: N98 to Q (AAC to CAG)	This Work
	Silent mutation: G104 (GGC to GGA)	
pRV5_SID115mAYF	Mutation: S115 to A (AGC to GCC), I116 to Y (ATT to TAT), D117 to F (GAC to TTC)	This Work
	Silent mutation: E114 (GAA to GAG)	
pRV5_D117+GGG	Mutation: Addition of GGG (GGC GGC GGA) after D117 (GAC)	This Work
pRV5_I120mG	Mutation: I120 to G (ATC to GGC)	This Work
	Silent mutation: S118 (TCC to TCG)	
pRV5_Y147mF	Mutation: Y147 to F (TAT to TTT)	This Work
pRV5_C149mW	Mutation: C149 to W (TGC to TGG)	This Work
	Silent Mutation: Y147 (TAT to TAC)	
pRV5_H168mD	Mutation: H168 to D (CAC to GAC)	This Work
	Silent Mutation: S164 (TCA to TCG)	
pRV5_Q180mE	Mutation: Q180 to E (CAG to GAG)	This Work
	Silent mutation: L181 (CTG to CTC)	
pRV5_Q180mE.N177My	Mutation: Q180 to E (CAG to GAG), N177 to Y (AAC to TAC)	This Work
	Silent mutation: L181 (CTG to CTC)	
pRV5_Δ3'	Mutation: Deletion of G201 to H209	This Work
pRV5_Δ3'Δ6H	Mutation: Deletion of G201 to H209 Deletion of the 3' 6xHis sequence	This Work

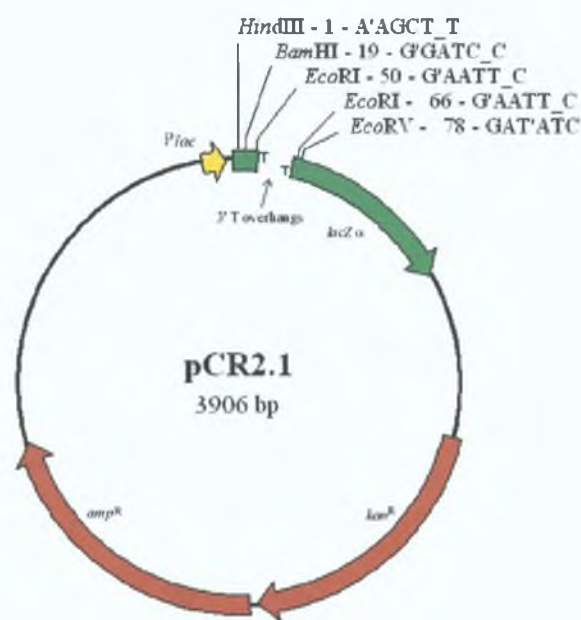


Figure 2.1 pCR2.1 vector
 The 3906 bp TA cloning vector pCR2.1 (Table 2.3). The multiple cloning site (MCS) enzymes are indicated. The TA cloning site is situated within the *LacZα* ORF (green), which is under the control of the *P_{lac}* promoter (yellow). Ampicillin and Kanamycin resistance genes (*amp^R* & *kan^R*) are shown in red. Generated using pDRAW32 (Section 2.11).

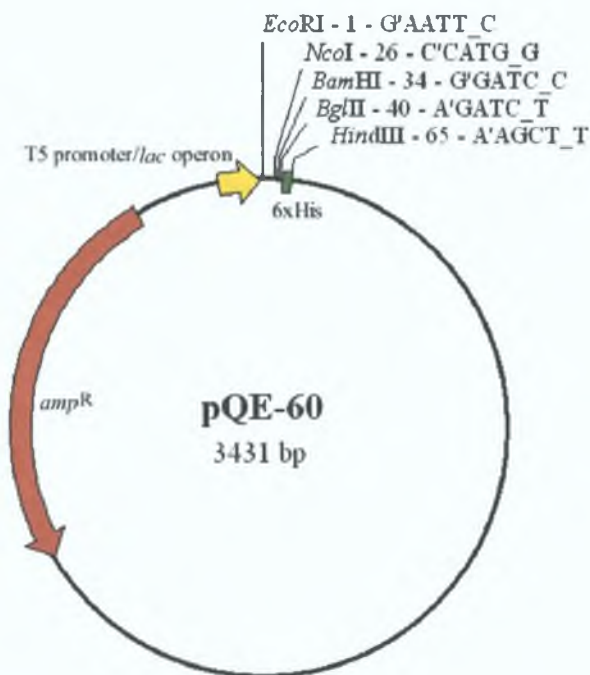


Figure 2.2 pQE-60 vector

The 3431 bp cloning/expression vector pQE-60 (Table 2.3). The multiple cloning site (MCS) enzymes are indicated. The 6xHis coding sequence (green) is situated at the 3' end of the MCS, which in turn is situated downstream of the T5 promoter/lac operon (yellow). Ampicillin resistance gene (*amp^R*) is shown in red. Generated using pDRAW32 (Section 2.11).

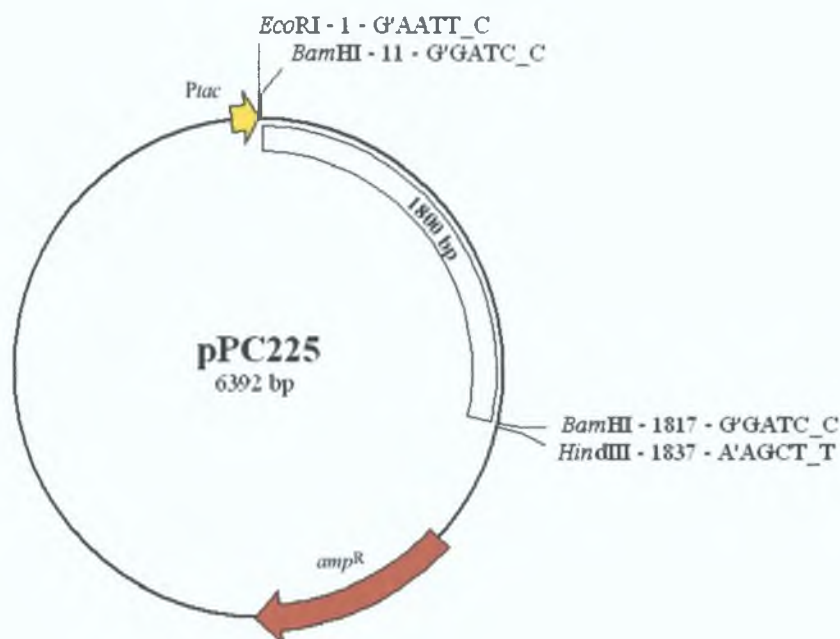


Figure 2.3 pPC225 vector

The 6392 bp cloning/expression vector pPC225 (Table 2.3). The multiple cloning site (MCS) enzymes are indicated. A 1800 bp sequence (white box) is situated within the MCS, which in turn is situated downstream of the *P_{lac}* promoter (yellow). Ampicillin resistance gene (*amp^R*) is shown in red. Generated using pDRAW32 (Section 2.11).

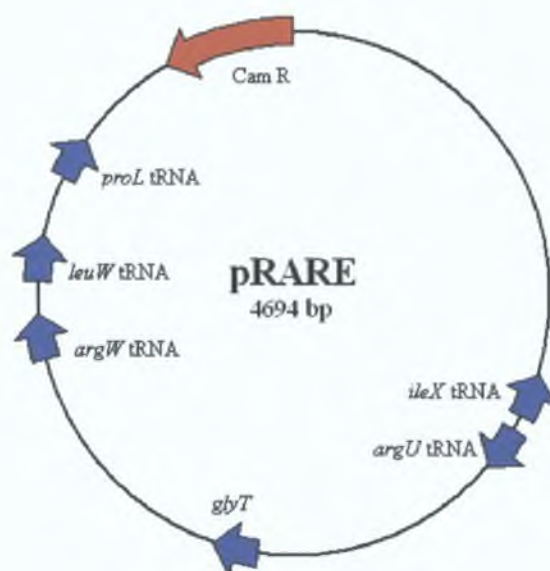


Figure 2.4 pRARE vector

The 4694 bp pRARE plasmid (Table 2.3) codes for 6 tRNAs that recognise codons considered rare in *E. coli*: *ileX* (AUA), *argU* (AGG/AGA), *argW* (AGG), *glyT* (GGA), *leuW* (CUA) and *proL* (CCC). Chloramphenicol resistance gene (*cam^R*) is shown in red. Generated using pDRAW32 (Section 2.11).

2.2 Media, solutions and buffers

All chemicals and reagents were obtained from Sigma-Aldrich unless otherwise stated. All chemicals were Analar grade. Microbiological media were obtained from Oxoid. Sterilisation was achieved by autoclaving at 121°C and 15 lb/in² for 20 min, unless otherwise stated.

Luria Bertani broth (LB)

Tryptone	10 g/L
NaCl	10 g/L
Yeast Extract	5 g/L

Adjusted to pH 7.0 with NaOH. Sterilised by autoclaving. For solid, LB 15 g/L Technical Agar No.3 (Oxoid) was included.

SOB broth

Tryptone	20 g/L
NaCl	500 mg/L
Yeast Extract	5 g/L
KCl	2.5 mM
pH	7.0

After autoclaving, MgCl_2 and MgSO_4 were added to 10 mM from sterile 1 M stock solutions.

TE buffer

Tris-HCl	10 mM
Na_2 -EDTA	1 mM
pH	8.0

TAE buffer (50X)

Tris	242 g/L
Glacial Acetic Acid	57.1 ml/L
EDTA	100 ml/L (of 0.5 M stock)
pH	8.0

STET buffer

Sucrose	8% (w/v)
Triton X-100	5% (v/v)
Tris-HCl	50 mM
Na_2 -EDTA	50 mM
pH	8.0

Solution 1 of 1-2-3 method (Section 2.4.1.2)

Glucose	50 mM
Na_2 -EDTA	10 mM (from 0.5 M stock)
Tris-HCl	25 mM (from 1 M stock)

Solution 2 of 1-2-3 method (Section 2.4.1.2)

NaOH	200 mM
SDS	1% (w/v)

Solution 3 of 1-2-3 method (Section 2.4.1.2)

Potassium acetate	3 M
pH	4.8

To 60 ml of 5 M potassium acetate, 11.5 ml of glacial acetic acid and 28.5 ml of dH₂O was added. The resulting solution was 3 M with respect to potassium and 5 M with respect to acetate.

TB buffer

Pipes	10 mM
CaCl ₂	15 mM
KCl	250 mM
pH	6.7

The pH was adjusted with KOH and then MnCl₂ was added to 55 mM. The solution was filter sterilised through a 0.22 µm membrane and stored at 4°C.

RF1 buffer

RbCl	100 mM
CaCl ₂	10 mM
Potassium acetate	30 mM
Glycerol	15% (v/v)
pH (with HCl)	5.8

After the pH had been adjusted MnCl₂ was added to 50 mM. The solution was filter sterilised through a 0.22 µm membrane and stored at 4°C.

RF2 buffer

RbCl	10 mM
MOPS	10 mM
CaCl ₂	75 mM
Glycerol	15% (v/v)
pH	6.8

The solution was filter sterilised through a 0.22 µm membrane and stored at 4°C.

Transfer buffer

Tris-HCl	25 mM
Glycine	150 mM
Methanol	10% (v/v)

TBS buffer

Tris-HCl	10 mM
NaCl	150 mM
pH	7.5

TBS-Tween/Triton buffer

Tris-HCl	20 mM
NaCl	500 mM
Tween 20	0.05% (v/v)
Triton X-100	0.2% (v/v)
pH	7.5

Gel loading dye (6X)

Bromophenol Blue	0.25%
Xylene Cyanol	0.25%
Ficoll (Type 400)	15%

Bromophenol Blue and/or Xylene Cyanol were used as appropriate. On a 1% agarose gel, bromophenol blue and xylene cyanol migrate approximately with the 300 bp and 4000 bp fragments respectively.

Ethidium bromide stain

A 10 mg/ml stock solution in dH₂O was stored at 4°C in the dark. For the staining of agarose gels, 100 µl of the stock solution was mixed into 1 L of dH₂O. The staining solution was kept in a plastic tray and covered to protect against light. Used ethidium bromide stain was collected and filtered through a deactivating filter (Schleicher & Schuell).

2.3 Antibiotics

- Ampicillin was prepared in dH₂O at a concentration of 100 mg/ml and stored at -20°C. The working concentration for *E. coli* was 100 µg/ml.
- Chloramphenicol was prepared in ethanol at a concentration of 100 mg/ml and stored at -20°C. The working concentration for *E. coli* was 25 µg/ml.
- Tetracycline was prepared in 50 % ethanol at a concentration of 10 mg/ml and stored at -20°C. The working concentration for *E. coli* was 10 µg/ml.

- Kanamycin was prepared in dH₂O at a concentration of 100 mg/ml and stored at –20°C. The working concentration for *E. coli* was 30 µg/ml.

2.4 Isolation and purification of DNA and RNA

2.4.1 Isolation of plasmid DNA

Three procedures for the isolation of plasmid DNA were variably employed. The Rapid Boiling Method (Section 2.4.1.1) was used for convenient plasmid isolation from large numbers of samples, mostly for the purpose of screening. The 1-2-3 Method (Section 2.4.1.2), yielding higher quality DNA than the previous method, was used routinely. The Genelute Plasmid Miniprep Kit (Sigma, Section 2.4.1.3) was used to prepare consistently pure and supercoiled plasmid DNA, mostly for the purpose of DNA sequencing.

2.4.1.1 Rapid boiling method

This method is adapted from the procedure outlined by Holmes and Quigley (1981). 1.5 ml of a bacterial culture in a microfuge tube was centrifuged at 13,000 rpm for 5 min to collect the cells. The supernatant was discarded and the cell pellet re-suspended in 350 µl of STET buffer. Alternatively, bacterial growth was taken off an LB agar culture plate with a sterile loop and re-suspended in 350 µl of STET buffer. 20 µl of a freshly made 10 mg/ml lysozyme solution was added and the microfuge tube was incubated at room temperature for 10 min. The microfuge tube was placed in a boiling water bath for 60 seconds and then centrifuged at 13,000 rpm for 10 min. The supernatant was removed to a new microfuge tube. An equal volume of isopropanol was added and mixed by inversion. The tube was left at room temperature for 5 min and then centrifuged at 13,000 rpm for 20 min to pellet the plasmid DNA. The pellet was washed with 70% ethanol and then dried briefly in a SpeedVac (Savant) vacuum centrifuge. The plasmid DNA was resuspended in 50 µl of TE buffer and 1 µl of Ribonuclease A (Section 2.8) was added to digest co-purified RNA. Plasmid DNA was stored at –20°C.

2.4.1.2 1-2-3 method

This method is adapted from the procedure described by Birnboim and Doly (1979). 1.5 ml of a bacterial culture in a microfuge tube was centrifuged at 13,000 rpm for 5 min to collect the cells. The supernatant was discarded and the cell pellet re-suspended in 200 µl of Solution 1. Alternatively, bacterial growth was taken off an LB agar culture plate

with a sterile loop and re-suspended in 200 µl of Solution 1. The re-suspension was left for 5 min at room temperature. 200 µl of Solution 2 was added, the tube was mixed by inversion and placed on ice for 5 min. 200 µl of Solution 3 was added, the tube was mixed by inversion and placed on ice for 10 min. A clot of chromosomal DNA was collected by centrifugation at 13,000 rpm for 10 min. The supernatant was removed to a new microfuge tube with 400 µl of phenol chloroform isoamylalcohol (25:24:1) and mixed by brief vortexing. Upon centrifugation at 13,000 rpm for 5 min the mixture is divided into an upper aqueous and lower organic layer. The aqueous layer was removed to a new microfuge tube with an equal volume of isopropanol and mixed by inversion. The tube was left at room temperature for 5 min and then centrifuged at 13,000 rpm for 20 min to pellet the plasmid DNA. The pellet was washed with 70% ethanol and then dried briefly in a SpeedVac (Savant) vacuum centrifuge. The plasmid DNA was resuspended in 50 µl of TE buffer and 1 µl of Ribonuclease A (Section 2.8) was added to digest co-purified RNA. Plasmid DNA was stored at -20°C.

2.4.1.3 GenElute plasmid miniprep kit

The kit was used according to the manufacturer's instructions: 1.5 ml of a bacterial culture in a microfuge tube was centrifuged at 13,000 rpm for 5 min to collect the cells. The supernatant was discarded and the cell pellet was completely re-suspended in 200 µl of re-suspension solution. 200 µl of lysis solution was mixed in by inversion to lyse the cells. 350 µl neutralisation/binding buffer was added and mixed by inversion to precipitate cell debris, lipids, proteins and chromosomal DNA. The precipitate was collected by centrifugation at 13,000 rpm for 10 min. The supernatant was transferred to a spin column in a microfuge tube and centrifuged at 13,000 rpm for 1 min to bind the plasmid DNA. The flow through was discarded and 750 µl of washing solution was added followed by further centrifugation at 13,000 rpm for 1 min. The flow through was discarded and the column was centrifuged at 13,000 rpm for 1 min to dry the spin column. The spin column was transferred to a fresh microfuge tube and 100 µl TE buffer was added. The DNA was eluted by centrifugation at 13,000 rpm for 1 min.

2.4.2 Isolation of DNA from agarose gels

Both the GeneClean method (Section 2.4.2.1) and the PerfectPrep Kit (Eppendorf, Section 2.4.2.2) were used to purify DNA from agarose gels, the latter being more consistent and reliable.

2.4.2.1 GeneClean

This method was described by Mather *et al.* (1993). The DNA band to be isolated was excised from the agarose gel using a scalpel. The gel slice was weighed and placed in a microfuge tube. Three gel slice volumes of saturated sodium iodide solution were added. The tube was incubated at 65°C for 10 minutes to completely dissolve the agarose in the chaotropic solution. 2 µl (typically for up to 100 ng DNA) of silica 325 mesh glass beads were added and mixed by vortexing. The tube was kept at -20°C for 30 minutes. The beads were collected by centrifugation at 13,000 rpm for 5 seconds. The supernatant was discarded and the beads were washed three times with 300 µl 70% ethanol. The beads were re-suspended in 15-30 µl TE buffer and incubated at 55°C for 15 minutes to elute the DNA. The beads were removed by centrifugation at 13,000 rpm for 5 seconds and the supernatant was placed in a fresh microfuge tube.

2.4.2.2 PerfectPrep kit

The kit was used according to the manufacturer's instructions: The desired DNA band was excised from the agarose gel using a scalpel. The gel slice was weighed and placed in a microfuge tube. The tube was incubated at 65°C for 10 min to completely dissolve the agarose in the chaotropic solution. One gel slice volume of isopropanol was added and mixed by inversion. The solution was transferred to a spin column in a microfuge tube and centrifuged at 13,000 for 1 min to bind the plasmid DNA. The flow through was discarded and 750 µl of washing solution was added followed by further centrifugation at 13,000 rpm for 1 min. The flow through was discarded and the column was centrifuged at 13,000 rpm for 1 min to dry the spin column. The spin column was transferred to a fresh microfuge tube and 30 µl TE buffer was added. The DNA was eluted by centrifugation at 13,000 rpm for 1 min.

2.4.3 Purification and concentration of DNA samples

The sample containing the DNA to be precipitated was brought to 400 µl with dH₂O. 400 µl of phenol chloroform isoamylalcohol (25:24:1) was added and mixed by brief vortexing. Upon centrifugation at 13,000 rpm for 5 min the mixture is divided into an upper aqueous and lower organic layer. The aqueous layer was removed to a new microfuge tube with an equal volume of chloroform and mixed by brief vortexing. The tube was centrifuged at 13,000 rpm for 5 min and the aqueous layer was transferred to a

new microfuge tube. A 1/10 volume of 3 M sodium acetate was added followed by an equal volume of isopropanol and mixed by inversion. The tube was left at room temperature for 60 min and then centrifuged at 13,000 rpm for 20 min to pellet the DNA. The pellet was washed with 70% ethanol and then dried briefly in a SpeedVac (Savant) vacuum centrifuge. The DNA was resuspended in 20-50 µl of TE buffer.

2.4.4 Isolation of RNA

RNA was isolated using Trizol Reagent (Invitrogen), a mono-phasic solution of phenol and guanidine isothiocyanate. This is an improvement of the method developed by Chomczynski and Sacchi (1987). 100 mg of tissue was homogenised in 1 ml of Trizol reagent using a glass-Teflon homogeniser (treated with RNase AWAY, Molecular Bio-Products, *inc.*) and transferred to a microfuge tube. Alternatively, a pellet of cultured cells was lysed in 1 ml of Trizol reagent by repeated pipetting. The sample was incubated at room temperature for 5 min to allow complete dissociation of nucleoprotein complexes. 200 µl of chloroform was added, mixed by brief vortexing and incubated at room temperature for 3 min. The phases were separated by centrifugation at 13,000 rpm for 15 min. The upper aqueous layer was transferred to a fresh microfuge tube. 500 µl of isopropanol was added and mixed by inversion. The sample was incubated at room temperature for 10 min to precipitate the RNA and then centrifuged at 13,000 rpm for 10 min. The supernatant was discarded and the RNA pellet was washed with 1 ml of 70% ethanol. The RNA pellet was air-dried and dissolved in 30 µl RNase-free water. Any possible DNA contamination was eliminated by treatment with Deoxyribonuclease I (Section 2.8).

2.4.5 Quantification of DNA and RNA

Nucleic Acid concentration was quantified by UV spectrometry. A dilution of the sample (typically 1/50) in dH₂O was measured for absorbance at 260 nm. A reading of 1.0 corresponds to 50 µg/ml of DNA or 40 µg/ml of RNA. The approximate quantity of DNA in individual bands on agarose gels (Section 2.5) could be judged by comparing to bands of known amounts of linear DNA on a reference gel (Figure 2.5).

2.5 Agarose gel electrophoresis

DNA was analysed by electrophoresing through agarose gels in a BioRad horizontal gel apparatus. Agarose was added to TAE buffer to the required concentration (typically

0.7-2%) and dissolved by boiling. The agarose solution was poured into plastic trays and allowed to set with a plastic comb fitted to create sample wells. TAE buffer was used as the running buffer. Loading dye was mixed with the DNA samples to facilitate loading and to give indication of migration distance during electrophoresis. When RNA samples were being analysed 0.1% (v/v) DEPC was included in the TAE buffer. Gels were run at 140 volts for 20-40 min depending on size of gel. Gels were stained for 15 min by immersion in an ethidium bromide staining solution. Gels were visualised using a UV transilluminator coupled with an image analyser to capture the image to a PC. On every gel 0.5 μ g of 1 Kb Plus DNA Ladder (Invitrogen, Figure 2.5) was run as a molecular size marker.

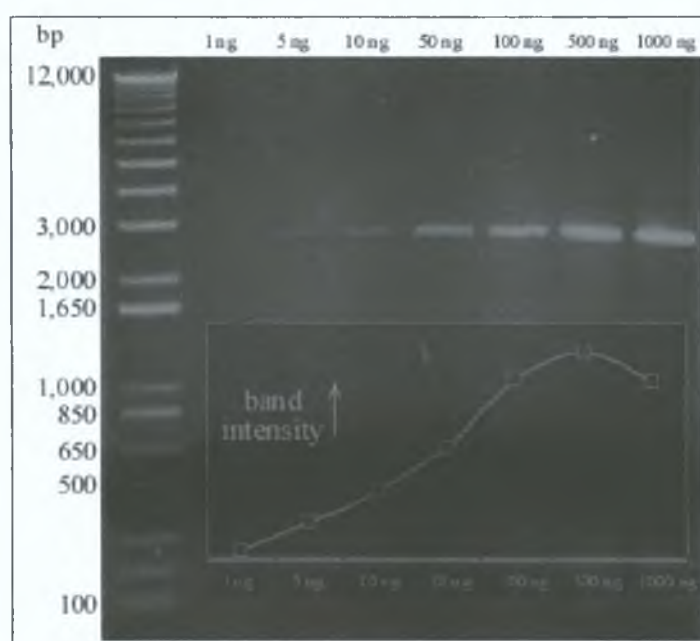


Figure 2.5 Reference DNA gel

1.0% Agarose Gel. Lane 1, 0.5 μ g DNA Ladder (band sizes are indicated); Lanes 2-8, Increasing amounts of linear 2686 bp pUC18 DNA (amounts are indicated); Inset graphically shows increase in band intensity as determined by densitometry (Section 2.16.5).

2.6 Competent cells

The two methods used to prepare competent cells were the rubidium chloride method (section 2.6.1) and the TB method (section 2.6.2). The former method was favoured for its reliability.

2.6.1 Rubidium chloride method

This is an adaptation of the method outlined by Hanahan (1985). Sterile conditions were used throughout. 10 ml of LB broth containing the relevant antibiotics was inoculated

with a single colony of the desired bacterial strain from a plate stock and cultured overnight at 37°C. A 1 L flask with 200 ml of LB broth was inoculated with 2 ml of the overnight culture and incubated at 37°C shaking at 225 rpm. When the culture had reached an OD₆₀₀ of ~0.5 (early-mid exponential phase) the flask was cooled in ice water. All subsequent transactions took place at 4°C. The culture was transferred to a sterile centrifuge bottle. The cells were collected by centrifugation at 3,000 rpm for 5 min (using a Beckman JA-14 rotor). The supernatant was decanted and the cells gently re-suspended in 60 ml of chilled RF1 buffer. The suspension was left on ice for 90 min. The cells were again collected by centrifugation at 3,000 rpm for 5 min. The supernatant was decanted and the cells gently re-suspended in 8 ml of chilled RF2 buffer. Aliquots of 800 µl were prepared in sterile 1.5 ml microfuge tubes and flash frozen using -70°C ethanol. The competent cells were stored at -70°C. Cells were routinely used within a few weeks.

2.6.2 TB method

This method was developed by Inoue *et al.* (1990). Sterile conditions were used throughout. 10 ml of LB broth containing the relevant antibiotics was inoculated with a single colony of the desired bacterial strain from a plate stock and cultured overnight at 37°C. A 1 L flask with 200 ml of SOB broth was inoculated with 2 ml of the overnight culture and incubated at 37°C shaking at 225 rpm. When the culture had reached an OD₆₀₀ of 0.4, as optimised by Inoue *et al.* (1990), the flask was cooled in ice water. All subsequent transactions took place at 4°C. The culture was transferred to a sterile centrifuge bottle. The cells were collected by centrifugation at 3,000 rpm for 5 min (using a Beckman JA-14 rotor). The supernatant was decanted and the cells gently re-suspended in 80 ml of chilled TB buffer. The suspension was left on ice for 10 min. The cells were again collected by centrifugation at 3,000 rpm for 5 min. The supernatant was decanted and the cells gently re-suspended in 15 ml of chilled TB buffer. DMSO was added drop-wise to 7% (v/v). The suspension was left on ice for 10 min. Aliquots of 800 µl were prepared in sterile 1.5 ml microfuge tubes and flash frozen using -70°C ethanol. The competent cells were stored at -70°C. Cells were routinely used within a few weeks.

2.6.3 Transformation of competent cells

An aliquot of competent cells was thawed on ice. 200 μl of the cell suspension was mixed gently with 1-50 μl of plasmid DNA in a sterile 1.5 ml microfuge tube. The mixture was left on ice for 30 min. The cells were heat-shocked at 42°C for 30 seconds and placed back on ice for 2 min. 800 μl of LB broth was added to the cells followed by incubation at 37°C for 60 min. 100 μl of the transformation suspension was spread on an LB agar plate containing the relevant antibiotics and incubated at 37°C overnight.

2.6.4 Determining cell efficiency

Competent cell efficiency is defined in terms of the number of colony forming units obtained per μg of transformed plasmid DNA. A 10 ng/ μl stock of pUC18 plasmid DNA was diluted to 1 ng/ μl , 100 pg/ μl and 10 pg/ μl . 1 μl of each dilution was transformed as described above. The cell efficiency was calculated from the number of colonies obtained, taking into account the dilution factor and the fraction of culture transferred to the spread plate.

2.7 Bacterial storage

Bacterial strains were stored as 40% glycerol stocks. 750 μl of an overnight culture was mixed with 750 μl sterile 80% glycerol in a microfuge tube. If the bacterial strains contained plasmids, the selective antibiotic was included in the culture. Duplicate stocks were stored at -20°C and -70°C. Working stocks streaked on LB agar plates, containing antibiotics where appropriate, were stored at 4°C. Bacterial samples containing plasmids to be sent for commercial DNA sequencing were prepared as stab cultures. 1 ml of LB agar containing the selective antibiotic was added to a microfuge tube and allowed to set. The tube was inoculated by stabbing using a sterile pin with bacterial growth taken from an LB broth culture.

2.8 Enzymes

All restriction endonucleases and T4 DNA ligase were obtained from Invitrogen Life Technologies or New England Biolabs. Deoxyribonuclease I (Ribonuclease free), Ribonuclease A (Deoxyribonuclease free), RED*Taq* and RED*AccuTaq* DNA polymerases were obtained from Sigman-Aldrich. *PfuTurbo* DNA polymerase was obtained from Stratagene. AMV reverse Transcriptase was obtained from Promega.

Enzymes were used with their relevant buffers according to the manufacturers instructions.

2.8.1 Reverse transcription

1 µg of RNA was combined with 0.5 µg oligo(dT)₁₅ primer (alternatively 1 µl of 25 µM specific reverse primer or 1 µl of 50 µM random nonamers) and the volume made to 5 µl with dH₂O. The mixture was incubated at 70°C for 10 min and then placed on ice for 5 min. To the mixture were added dNTP's to 500 µM each, 1 unit AMV reverse transcriptase and 2 µl specific 10x enzyme buffer. The volume was brought to 20 µl with dH₂O. The first strand was synthesised at 42°C for 60 min followed by inactivation of the transcriptase at 95°C for 2 min. 2 µl of the reaction was used as template for PCR.

2.8.2 Polymerase chain reaction

PCR reactions (Mullis and Faloona, 1987) were carried out using a Hybaid PCR Express Thermocycler. The standard PCR reaction volume was 50 µl containing 1 µl template (10-100 ng), 1-2.5 mM MgCl₂, 0.5 µM of each primer, 200 µM of each dNTP, 1 unit RED*Taq* or RED*AccuTaq* DNA polymerase and 5 µl specific 10x enzyme buffer.

The standard PCR program was:

Stage 1	Step 1: 95°C for 10 min
Stage 2 (30 cycles)	Step 1: 95°C for 1 min
	Step 2: T _{ann} for 30 sec
	(T _{ann} was routinely 5°C below the T _m of the primers)
	Step 3: 72°C for 1 min per Kb to be synthesised
Stage 3	Step 1: 72°C for 10 min

2.9 Gene manipulation

A variety of plasmid-based gene cloning approaches such as described by Maniatis *et al.* (1982) were used, employing PCR techniques (Section 2.8.2), restriction endonucleases and DNA ligase (Section 2.8).

2.9.1 TA cloning of PCR products

PCR products were cloned using the TA cloning vector pCR2.1 (Invitrogen, Table 2.3, Figure 2.1). The cloning method using this vector exploits the fact that thermostable

polymerases such as *Taq* DNA polymerase leave 3' A-overhangs due to their lack of 3'-5' exonuclease activity. The pCR2.1 vector is provided with a location in the *lacZa* gene open with 3' T overhangs (Figure 2.1). PCR products with 3' A-overhangs can be directly ligated into this cloning site. The ligation was transformed into *E. coli* cells (routinely DH5 α or INV α F' strains were used, Table 2.1) as described in Section 2.6.3. 100 μ l of the transformation reaction was plated on an LB agar plate containing ampicillin and kanamycin. The agar plate was previously overlaid with 60 μ l of 40 mg/ml X-gal to test for α -complementation of β -galactosidase. X-gal stock was prepared in DMF and stored in the dark at -20°C. The *lacZa* gene encodes the α -peptide of β -galactosidase, which cleaves the X-gal substrate yielding a blue product. This allows for positive selection of transformants harbouring plasmids with PCR product inserts. Routinely a small proportion of colonies are blue. It is assumed that these arise as a result of re-circularisation of the pCR2.1 vector.

2.9.2 Site-specific mutagenesis

Point mutations were introduced into open reading frames on plasmid constructs by PCR amplification using complementary primers carrying the desired mutation. 50 μ l reactions were set up, containing 10-100 ng of plasmid template isolated from *dam*⁺ *E. coli* (e.g. DH5 α , INV α F' or XL-10 Gold strains, Table 2.1), 0.5 μ M of each primer, 200 μ M of each dNTP, 2.5 units of *PfuTurbo* DNA polymerase and 5 μ l specific 10x enzyme buffer. The particular PCR program used was:

Stage 1	Step 1: 95°C for 1 min
Stage 2 (18 cycles)	Step 1: 95°C for 50 sec
	Step 2: 60°C for 50 sec
	Step 3: 68°C for 2 min per Kb of template plasmid
Stage 3	Step 1: 68°C for 7 min

The template DNA was eliminated by digestion with 20 units of *Dpn* I restriction endonuclease. *Dpn* I is biased toward a methylated recognition sequence. It selectively digests the template DNA from *dam*⁺ *E. coli* strains mentioned above and not the newly synthesised DNA. Subsequently the samples were transformed into *E. coli* XL10-Gold. Transformants were screened by restriction analysis. Where possible, a diagnostic restriction site had been created by co-introducing a silent mutation. Potential mutants were verified by DNA sequencing.

2.10 DNA sequencing

Recombinant clones and potential mutants were verified by DNA sequencing. Commercial sequencing services were provided by MWG Biotech AG (also Lark Technologies Inc. and Qiagen Ltd.). Suitable sequencing primers (Table 2.2) for standard vectors were provided as part of the service. Samples were variably sent as dried plasmid DNA or as stab cultures.

2.11 Bio-Informatics

Nucleotide and Amino Acid sequences were analysed using a variety of web-based tools. The BLAST programs (Altschul *et al.*, 1997) at the National Centre for Biotechnology Information (<http://www.ncbi.nlm.nih.gov>) were used to identify homologous sequences deposited in GenBank (Benson *et al.*, 1996). Protein sequences, structure files and tools to calculate protein molecular weights and isoelectric points were obtained from the Swiss-Prot database (Bairoch and Apweiler, 1996) at (<http://us.expasy.org>) and the Protein Data Bank (PDB, Berman *et al.*, 2000) at (<http://www.rcsb.org/pdb>). Alignments of DNA and Protein sequences were performed using the MultAlin program (Corpet, 1988) available at (<http://prodes.toulouse.inra.fr/multalin/multalin.html>) and edited using the GeneDoc program (Nicholas *et al.*, 1997) available for download at (www.psc.edu/biomed/genedoc). Plasmid maps were constructed using the pDRAW32 program available for download at (<http://www.acaclone.com>). Chemical structures were drawn using the ACD/Labs ChemSketch program available for download at (<http://www.acdlabs.com>). DNA sequences were analysed for restriction enzyme sites using the Webcutter 2.0 tool at (<http://rna.lundberg.gu.se/cutter2>). Tertiary protein structure was predicted by the automated SWISS-MODEL server (Schwede *et al.*, 2003, <http://swissmodel.expasy.org>), subsequently analysed and visualised using the DeepView software (Guex and Peitsch, 1997) available for download at (<http://ca.expasy.org/spdbv>). The secondary structure of mRNA was predicted using Mfold (Zuker, 2003; <http://mfold.burnet.edu.au>).

2.12 Protein expression

2.12.1 Standard expression culture

100 ml of LB broth was inoculated with 1 ml of a stationary phase culture of *E. coli* that had been transformed with an expression plasmid. Selective antibiotics were included in

the LB broth. The culture was incubated at 37°C, shaking at 220 rpm, until an optical absorbance (A_{600}) of 0.3-0.5 was reached. IPTG was added (from 10 mM stock) to a final concentration of 50 μ M to induce expression (if required). The culture was allowed to incubate for a defined period, sampling 1 ml routinely at hourly intervals. Both these samples and the remaining culture were centrifuged at 5,000 rpm for 5 min (using a Beckman JA-14 rotor) to pellet the cells. The supernatant was discarded and the pellets were stored at -20°C.

2.12.2 Preparation of cleared lysate

A cell pellet from a 100 ml expression culture was washed in potassium phosphate buffer, pH 8.0 followed by centrifugation at 4,000 rpm for 5 min (using a Beckman JA-20 rotor). The supernatant was discarded and the cells were re-suspended in 10 ml Potassium phosphate buffer, pH 8.0. In the case of the 1 ml samples, 100 μ l of potassium phosphate buffer, pH 8.0 was used. The cells were disrupted on ice with a 3 mm micro-tip sonicator (Sonics & Materials Inc.) using 2.5 sec, 40 kHz pulses for 30 sec. The cell debris was removed by centrifugation at 4,000 rpm for 20 min at 4°C (using a Beckman JA-20 rotor). The cleared lysate was transferred to a fresh universal container and stored at 4°C.

2.13 Protein purification

Immobilised Metal Affinity Chromatography (IMAC) was used to purify recombinant human PAP1 having a C-terminal 6xHis tag (Table 2.3).

2.13.1 Standard IMAC procedure

1 ml of nickel-nitrilotriacetic acid resin (Ni-NTA, Qiagen) was gently mixed with 10 ml of cleared lysate for 60 min at 4°C. The mixture was poured into a 0.7 x 15 cm column, allowing the resin to settle. The column was washed three times with 10 ml potassium phosphate buffer, pH 8.0 containing 20 mM imidazole and then eluted with 5 ml potassium phosphate buffer, pH 8.0 containing 200 mM imidazole. The elute was dialysed overnight against 1 L potassium phosphate buffer, pH 8.0. Samples taken throughout the procedure were analysed by SDS-PAGE (Section 2.16). Protein concentration was determined by the standard BCA or Coomassie Plus assay (Section 2.14) and PAP1 activity was determined by the fluorimetric assay (Section 2.15). The purified PAP1 sample was stored at 4°C or -20°C with 40% glycerol.

2.13.2 Recharging of Ni-NTA resin

This procedure was routinely used before re-using the Ni-NTA Resin. The resin was poured into a column and washed with 2 column volumes (2cv) of distilled water followed by 2cv 50% ethanol. The resin was stripped with 3cv 100 mM EDTA, pH 8.0. The resin was washed with 2cv 500 mM NaCl followed by 2cv distilled water. The resin was re-charged with 2cv 100 mM NiSO₄. The resin was washed with 2cv distilled water, transferred to a plastic container and stored at 4°C in 20% ethanol.

2.13.3 Preparation of dialysis tubing

The required amount of tubing was placed in a 1 L glass beaker and rinsed thoroughly with distilled water. The beaker was filled with distilled water and ~1 g of EDTA was added. The beaker was brought to boil and boiled for 2 min. The water was allowed to cool and then poured off. The beaker was re-filled with fresh distilled water and boiled again for 2 min. The water was allowed to cool and poured off. The tubing was thoroughly rinsed with distilled water. The tubing was stored at 4°C in distilled water.

2.14 Protein concentration

2.14.1 Quantitative determination by BCA assay

The Bicinchoninic acid (BCA) assay described by Smith *et al.* (1985) was used for the colorimetric detection and quantification of total protein in the range of 20-2,000 µg/ml. All samples were dialysed against dH₂O and diluted appropriately to achieve a concentration within range of the assay. 100 µl of sample was added in triplicate to 2 ml of BCA reagent (Sigma) and incubated at 37°C for 30 min. Absorbances were read at 562 nm on a spectrometer blanked with dH₂O. Bovine serum albumin (BSA) was used as the reference protein. BSA standards (0-2 mg/ml) were prepared in dH₂O and assayed in triplicate to yield a standard curve (Appendix A). Protein concentration of samples was determined from this standard curve.

2.14.2 Quantitative determination by Coomassie assay

The Coomassie Plus assay (Bradford, 1976) was used for the colorimetric detection and quantification of total protein in the range of 100-1,500 µg/ml. All samples were dialysed against dH₂O and diluted appropriately to achieve a concentration within range of the assay. 50 µl of sample was added in triplicate to 1 ml of Coomassie Plus reagent

(Pierce) and mixed gently. Absorbances were read at 595 nm on a spectrometer blanked with dH₂O. Bovine serum albumin (BSA) was used as the reference protein. BSA standards (0-2 mg/ml) were prepared in dH₂O and assayed in triplicate to yield a standard curve (Appendix A). Protein concentration of samples was determined from this standard curve.

2.14.3 Qualitative determination by Coomassie assay

A 96-Well plate version of the Coomassie assay (Section 2.14.2) with a working range of 0-25 µg/ml was used for post column chromatography fractions. 150 µl of sample was added to 150 µl of Coomassie Plus reagent and read at 595 nm using a Tecan Spectra Plate Reader. A BSA standard curve (Appendix A) was prepared under the same conditions.

2.15 Fluorescence spectrometry

2.15.1 AMC standard curve

A 100 µM stock of 7-Amino-4-Methyl-Coumarin (AMC) was prepared in 50 mM potassium phosphate buffer, pH 8.0 containing 4% (v/v) methanol. This stock solution was stored in the dark at 4°C for up to one month. AMC standards were prepared by dilution of the stock solution with potassium phosphate buffer, pH 8.0 to the range of 0-20 µM. 500 µl of each standard was incubated in triplicate at 37°C for 15 min followed by the addition of 1 ml of 1.5 M acetic acid. The fluorescence intensities of the standards were read using a Perkin Elmer LS50 Fluorescence Spectrophotometer at excitation and emission wavelengths of 370nm and 440nm respectively. Excitation slit width was 10nm while emission slit width was 2.5nm. This AMC standard curve is shown in Appendix A.

2.15.2 Quantitative fluorimetric PAP1 assay

Quantitative pGlu-AMC degrading PAP1 activity was determined according to the method by Fujiwara and Tsuru (1978), as modified by Browne and O'Cuinn (1983). PAP1 sample was suitably diluted in potassium phosphate buffer, pH 8.0. 100 µl was incubated in triplicate at 37°C for 15 min with 400 µl of pGlu-AMC substrate (final concentration 250 µM) in 50 mM potassium phosphate, pH 8.0 containing at final concentration 10 mM DTT, 2 mM EDTA and 5% (v/v) DMSO. Blanks were set up by substituting 100 µl 50 mM potassium phosphate buffer, pH 8.0 for PAP1 sample. The

reaction was terminated by the addition of 1 ml 1.5 M acetic acid. Liberated AMC was detected using a Perkin-Elmer LS-50 fluorescence spectrophotometer with excitation and emission wavelengths of 370 and 440 nm, respectively. The fluorescence readings were converted to nanomoles of AMC released per minute by calculations shown in Appendix C, using a standard curve of free AMC (Appendix A), prepared as described in Section 2.15.1 under identical assay conditions. Units of PAP1 activity were defined as nanomoles of AMC released per minute at 37°C (unit = nmoles.min⁻¹).

To monitor the effect of temperature on PAP1 activity, the standard assay was carried out at a range of temperatures from 4 to 70°C. PAP1 samples were pre-incubated for 5 min before the addition of the substrate solution.

The stability of PAP1 at various temperatures was determined by incubating samples at a range of 4 to 70 °C. Aliquots were removed at appropriate time intervals for up to 120 min and assayed under standard conditions at 37°C. The initial activity was compared to the residual activity at each temperature.

2.15.3 Qualitative 96-well plate fluorimetric PAP1 assay

A qualitative version of the above assay, for large scale screening of samples was carried out on a 96-well plate. 10 µl of sample was incubated in triplicate with 40 µl of the substrate solution used in Section 2.15.2 at 37°C for 15 min. The reaction was terminated by the addition of 100 µl 1.5 M acetic acid. Liberated AMC was detected fluorometrically using a Perkin Elmer LS-50B plate reader attachment with excitation and emission wavelengths of 370 and 440 nm, respectively.

2.15.4 Fluorimetric colony assay

Screening for transformants with PAP1 activity was carried out by a fluorimetric plate assay similar to that described by Mulczyk and Szewczuk (1970). Selective agar plates with colonies from an overnight transformation were flooded with 250 µM pGlu-AMC substrate in 50 mM potassium phosphate, pH 8.0 with 10 mM DTT, 2 mM EDTA and 5% (v/v) DMSO. After 30 min incubation at 37°C the plates were viewed under UV light. A digital camera was used to capture the images to a PC.

2.16 SDS-PAGE

Protein samples were analysed by sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE), based on the method outlined by Laemmli (1970).

2.16.1 Preparation of SDS gels

10% and 15% resolving and 4% stacking polyacrylamide gels were prepared as per Table 2.4. Gels were cast using an ATTO vertical mini electrophoresis system. Upon the addition of the TEMED to the resolving gel it was poured immediately and overlaid with 50% ethanol. After polymerisation, the overlay was removed. Upon the addition of the TEMED to the stacking gel it was poured immediately and allowed to polymerise around a comb placed into the top of the gel liquid to form loading wells.

Table 2.4 Preparation of SDS-PAGE gels

Solution	10% Resolving Gel	15% Resolving Gel	4% Stacking Gel
1.5 M Tris-HCl, pH 8.8 Resolving Gel buffer	1.625 ml	1.625 ml	-
0.5 M Tris-HCl, pH 6.8 Stacking Gel buffer	-	-	0.625 ml
dH ₂ O	2.64 ml	1.56 ml	1.538 ml
Acrylamide/Bis-acrylamide 30%/0.8% (w/v)	2.17 ml	3.25 ml	0.335 ml
10% (w/v) Ammonium Persulphate	32.5 µl	32.5 µl	12.5 µl
20% (w/v) SDS	32.5 µl	32.5 µl	12.5 µl
TEMED	3.25 µl	3.25 µl	2.5 µl

2.16.2 Sample preparation

20 µl of sample was added to 5µl solubilisation buffer (5X) consisting of 50% (v/v) glycerol, 2% (w/v) SDS, 5% (v/v) 2-mercaptoethanol, 0.1% (w/v) bromophenol blue and 62.5 mM Tris-HCl, pH 6.8. Samples were boiled for 3 minutes and stored on ice until application.

2.16.3 Sample application

Routinely 25 µl of each prepared sample was applied to the SDS PAGE gel. Standard loading was achieved using a dilution approach, together with densitometry readings (Section 2.16.5) for constitutively expressed standard protein bands.

10 µl relative molecular weight protein marker (M_r) solution (SigmaMarker, Sigma, Figure 2.6) was also applied to the gel consisting of Rabbit Muscle Myosin (205 kDa), *E. coli* β-Galactosidase (116 kDa), Rabbit Muscle Phosphorylase b (97 kDa), Rabbit Muscle Fructose-6-phosphate Kinase (84 kDa), Bovine Serum Albumin (66 kDa), Bovine Liver Glutamic Dehydrogenase (55 kDa), Chicken Egg Ovalbumin (45 kDa), Rabbit Muscle Glyceraldehyde-3-phosphate Dehydrogenase (36 kDa), Bovine Erythrocyte Carbonic Anhydrase (29 kDa), Bovine Pancreas Trypsinogen (24 kDa), Soybean Trypsin Inhibitor (20 kDa), Bovine Milk α-Lactalbumin (14.2 kDa) and

Bovine Lung Aprotinin (6.5 kDa). Gels were run at 25 mA for 1-2 hrs at room temperature using Running Buffer containing 25 mM Tris-HCl, 192 mM glycine and 0.1% (w/v) SDS.

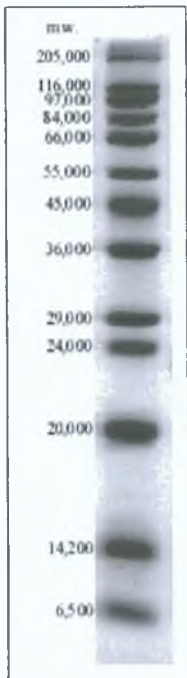


Figure 2.6 Relative molecular weight protein marker (M_r)
15% polyacrylamide gel

2.16.4 Gel staining

Polyacrylamide gels removed from the electrophoresis chamber and washed with dH₂O for 5 min. Routinely, gels were stained for 60 min in a 0.25% (w/v) Coomassie blue solution containing 10% (v/v) Acetic Acid/45% (v/v) methanol and then de-stained overnight using 10% (v/v) Acetic Acid/45% (v/v) methanol. Subsequent soaking in dH₂O enhanced the protein bands further. If this did not sufficiently stain the gel, the more intense silver staining method (Blum *et al.*, 1987) was used as outlined in table 2.5.

Table 2.5 Silver staining of SDS-PAGE gels

Step	Duration	Reagent
Fix	60 min	50% ethanol, 12% acetic acid, 0.05% formaldehyde (37% stock)
Wash	3 x 20 min	50% ethanol
Pre-Treat	1 min	200 µl of a 5% Na ₂ S ₂ O ₃ x H ₂ O stock solution in 100 ml dH ₂ O
Rinse	2 x 20 sec	dH ₂ O
Impregnate	20 min	0.1g AgNO ₃ , 70 µl formaldehyde in 100 ml dH ₂ O
Rinse	2 x 20 sec	dH ₂ O
Development	10 min (max.)	3g Na ₂ CO ₃ , 50 µl formaldehyde, 4 µl Na ₂ S ₂ O ₃ x H ₂ O stock solution in 100 ml dH ₂ O
Stop	5 min	0.1 M EDTA

2.16.5 Gel analysis

Gels were placed between two transparent sheets and scanned using a flatbed scanner, allowing the image to be saved to a PC. Analysis and densitometry of gel images was performed using the ImageJ program available for download at (<http://rsb.info.nih.gov/ij>). Densitometry values for specific sample bands were determined, taking into account relative loadings (Section 2.16.3) and expressed as percentages of either total protein or as relative ratios to defined bands used as standards (i.e. that have been shown to be constitutively expressed).

2.17 UV Zymography

2.17.1 Native PAGE

A 15% native polyacrylamide gel was prepared as in section 2.16.1 with the omission of SDS and set up at 4°C with running buffer as in section 2.16.3 with the omission of SDS. 20 µl of sample was added to 5 µl non-denaturing solubilisation buffer consisting of 50% (v/v) glycerol, 0.0125% (w/v) bromophenol blue and 62.5 mM Tris-HCl, pH 6.8. The samples were applied directly to the gel. The gel was run at constant 100 V at 4°C for up to 2.5 hrs.

2.17.2 UV zymogram development

After native PAGE the gel was gently rinsed with ice cold dH₂O and then incubated at 37°C in 20 ml of 5 µM pGlu-AMC in 50 mM potassium phosphate, pH 8.0 containing at final concentration 10 mM DTT, 2 mM EDTA and 5% (v/v) DMSO. After 10-15 min the gel was visualised using a UV transilluminator coupled with an image analyser to capture the image to a PC.

2.18 Western blotting

An SDS-PAGE gel was run as described in section 2.16. Coloured molecular weight markers (CM_r, Sigma ColorBurst, Figure 2.7) were included on the gel. 4 pieces of 3 mm filter paper (Whatman) and 1 piece of nitrocellulose membrane (Schleicher and Schuell) were cut to the size of the gel. The filter paper sheets and the membrane were soaked in Transfer Buffer for 10 min. Avoiding air bubbles, 2 sheets of filter paper were placed on the cathode of a horizontal semi-dry electro-blotter (ATTO), followed by the membrane, the gel, 2 sheets of filter paper and finally the anode. The protein transfer

was allowed to take place at a constant 20 V for 8 min. To verify protein transfer the membrane was incubated in 0.1% Ponceau S (Sigma), 5.0% Acetic Acid staining solution with gentle agitation for 2 min followed by de-staining in dH₂O. The membrane was washed twice for 10 min with TBS buffer. The membrane was then blocked in 30 ml 5% (w/v) Milk Powder in TBS buffer for 60 min followed by incubation in 20 ml of Anti-His-HRP conjugate Antibody solution (prepared in blocking buffer) for 60 min. The membrane was washed 4 times for 10 min with TBS-Tween/Triton buffer. Finally the membrane was incubated briefly with the HRP substrate DAB (3,3'-Diaminobenzidine Tetrahydrochloride) until the signal was clearly visible. Using a flatbed scanner an image of the blot was saved to a PC.



Figure 2.7 Coloured molecular weight marker (CM_r)
10% polyacrylamide gel

2.19 PAP1 characterisation

A range of biochemical and kinetic properties of recombinant PAP1 were determined (Sections 2.19.1 to 2.19.7) including native size, pH optimum, Michaelis constant, turnover number, maximal velocity, inhibition constants, effect of inhibitors, isoelectric point and crystallisation conditions.

2.19.1 Size exclusion chromatography

The native molecular mass of recombinant PAP1 under native conditions was determined by size exclusion chromatography. A 2.5 x 48 cm Sephadex G-100 gel-

filtration column (Sigma-Aldrich) was equilibrated with potassium phosphate buffer, pH 8.0 containing 100 mM NaCl at 4°C. The void volume of the column was determined by eluting 250 µl blue dextran (8 mg/ml) at a flowrate of 0.3 ml/min. 1 ml fractions were collected and optical density read at 620 nm. To calibrate the column, standard molecular mass markers were applied separately to the column and eluted at a flowrate of 0.3 ml/min. The markers applied were: 100 µl BSA (66 kDa, 2 mg/ml), 300 µl Carbonic Anhydrase (29 kDa, 1 mg/ml), and 100 µl lysozyme (14.2 kDa, 2 mg/ml). 1 ml fractions were assayed for protein concentration using the Coomassie method described in Section 2.14.3. 50 µl of purified PAP1 (250 µg/ml) was applied to the column. 1 ml fractions were assayed for activity using the 96-well plate method described in Section 2.15.3. A linear plot of log molecular mass versus V_e/V_o was constructed. The relative molecular mass of PAP1 was estimated using this plot.

2.19.2 Determination of pH optimum

The pH activity profile of PAP1 was determined by carrying out the standard activity assay as described in Section 2.15.2 at pH range 6.0-10.5. This range was established using the following buffers: 50 mM potassium phosphate for pH range 6.0-8.0, 50 mM Tris-HCl for pH range 7.5-9.5 and 50 mM NaOH/glycine for pH range 9.5-10.5. PAP1 samples were pre-incubated in the above buffers for 10 min at 37°C prior to addition of the substrate solution, also prepared in the respective buffers.

2.19.3 Determination of Michaelis constant

The Michaelis constant (K_m) was determined using a range of concentrations of pGlu-AMC (10-500 µM), prepared in 50mM potassium phosphate, pH 8.0 containing at final concentration 10 mM DTT, 2 mM EDTA and 5% (v/v) DMSO. Purified PAP1 activity was assayed in triplicate with each concentration as in Section 2.15.2. K_m , V_{max} (maximal velocity) and K_{cat} (turnover number) values of PAP1 for the substrate pGlu-AMC were obtained by fitting the data to Lineweaver-Burk, Eadie-Hofstee and Hanes-Woolf kinetic models (Appendix C).

2.19.4 Determination of dissociation constant

The effect of selected synthetic peptides (TRH, pGlu-His-Gly, pGlu-Ala and pGlu-Val) on the kinetic interaction between PAP1 and the substrate pGlu-AMC was determined. pGlu-AMC substrate concentrations were prepared (10-500 µM) in 50 mM potassium

phosphate, pH 8.0 containing at final concentration 10 mM DTT, 2 mM EDTA, 5% (v/v) DMSO and 250 μ M peptide. PAPI activity was assayed in triplicate as outlined in Section 2.15.2 in the presence of the above peptides. The data obtained was applied to the Lineweaver-Burk, Eadie-Hofstee and Hanes-Woolf kinetic models, whereby the dissociation, or inhibition constant (K_i), and the type of inhibition observed were determined as outlined in Appendix C.

2.19.5 Inhibition

A selection of compounds (Table 2.6) were investigated for their effect on the pGlu-AMC degrading activity of PAPI. If the effect was inhibitory, the IC_{50} value was determined, the inhibitor concentration resulting in the PAPI activity being reduced to 50%. pGlu-AMC (250 μ M) substrate solutions were prepared in 50 mM potassium phosphate, pH 8.0 containing, at final concentration, 10 mM DTT, 2 mM EDTA, 5% (v/v) DMSO and the compounds summarised in Table 2.6. PAPI samples were pre-incubated with the same relevant compound concentration for 10 min at 37°C. PAPI activity was assayed in triplicate using these substrate solutions as outlined in Section 2.15.2.

Table 2.6 Compounds tested for inhibition

Inhibitor	Assay conc. range (mM)
L-pGlu	0-20
2-pyrrolidone	0-20
Iodoacetate	0-1.0
1,10-phenanthroline	0-3.5
4,7-phenanthroline	0-5.0
diisopropylfluorophosphate (DFP)	0-5.0
Phenylmethylsulphonyl fluoride (PMSF)	0-5.5

2.19.6 Isoelectric focusing

The Isoelectric point of purified PAPI was determined by Isoelectric focusing (IEF). A 1 mm Novex pre-cast IEF gel (Invitrogen) containing a pH 3-10 gradient was used. PAPI samples were dialysed against dH₂O and 5 μ l mixed with 5 μ l of Novex IEF sample buffer. 5 μ l (50 μ g) of standard IEF markers (SERVA) were loaded alongside the samples (Figure 2.8). The gel was run at 100 V for 60 min, then 200 V for 60 min and finally 500 V for 30 min. The gel was stained and visualised as described in Sections 2.16.4 and 2.16.5.

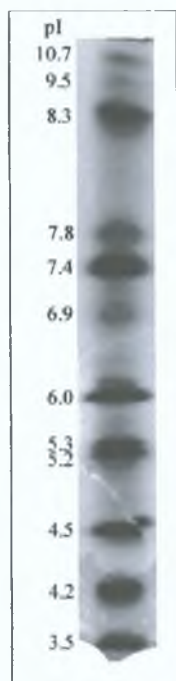


Figure 2.8 IEF Markers

2.19.7 Determination of crystallisation conditions

Preliminary screening was carried out to determine conditions under which recombinant PAP1 could form crystals. The sparse matrix screening strategy (Jancarik and Kim, 1991; Cudney *et al.*, 1994) was employed using Crystal Screen 1 and 2 reagent kits (Hampton Research). The kits consist of a panel of reagents combining various salts, buffers, precipitants and solvents. The composition of the Screen 1 and Screen 2 reagents are shown in Tables 2.7 and 2.8 respectively.

Purified PAP1 was dialysed extensively into dH₂O and then concentrated to 1.5-12 mg/ml using a SpeedVac (Savant) vacuum centrifuge. The screening was carried out by the hanging drop vapour diffusion method. 1 ml of each Screen reagent was aliquoted into a separate reservoir of a 24-Well sample plate. Vaseline was applied around the upper edge of each reservoir. 2 µl of reagent was taken from each reservoir and placed onto a glass cover slide. 2 µl of PAP1 sample was mixed with the drop of reagent. The slide was inverted and pressed gently over the reservoir, creating a seal with the Vaseline and resulting in the drop hanging over the reservoir. dH₂O from the dialysis reservoir was used as negative control. The drops were monitored daily using a microscope (10-100x magnification). Observations were recorded and images transferred to a PC using a digital camera attached to the microscope.

Table 2.7 Crystal screen 1

Reagent	Salt	Buffer	Precipitant
1	0.02 M Calcium chloride dehydrate	0.1 M Sodium acetate trihydrate pH 4.6	30 % v/v 2-methyl-2,4-pentanediol
2	None	None	0.4 M Potassium sodium tartrate tetrahydrate
3	None	None	0.4 M Ammonium dihydrogen phosphate
4	None	0.1 M Tris hydrochloride pH 8.5	2 M Ammonium sulfate
5	0.2 M tri-Sodium citrate dehydrate	0.1 M Sodium HEPES pH 7.5	30% v/v 2-methyl-2,4-pentanediol
6	0.2 M magnesium chloride hexahydrate	0.1 M Tris hydrochloride pH 8.5	30% w/v polyethylene glycol 4000
7	None	0.1 M sodium cacodylate pH 6.5	1.4 M sodium acetate trihydrate
8	0.2 M tri-sodium citrate dehydrate	0.1 M sodium cacodylate pH 6.5	30% v/v iso-propanol
9	0.2 M ammonium acetate	0.1 M tri-sodium citrate dihydrate pH 5.6	30% w/v polyethylene glycol 4000
10	0.2 M ammonium acetate	0.1 M Sodium acetate trihydrate pH 4.6	30% w/v polyethylene glycol 4000
11	None	0.1 M tri-sodium citrate dihydrate pH 5.6	1 M Ammonium dihydrogen phosphate
12	0.2 M magnesium chloride hexahydrate	0.1 M Sodium HEPES pH 7.5	30% v/v iso-propanol
13	0.2 M tri-Sodium citrate dehydrate	0.1 M Tris hydrochloride pH 8.5	30% v/v polyethylene glycol 400
14	0.2 M Calcium chloride dehydrate	0.1 M Sodium HEPES pH 7.5	30% v/v polyethylene glycol 400
15	0.2 M ammonium sulphate	0.1 M sodium cacodylate pH 6.5	30% w/v polyethylene glycol 8000
16	None	0.1 M Sodium HEPES pH 7.5	1.5 M lithium sulfate monohydrate
17	0.2 M lithium sulfate monohydrate	0.1 M Tris hydrochloride pH 8.5	30% w/v polyethylene glycol 4000
18	0.2 M magnesium acetate tetrahydrate	0.1 M sodium cacodylate pH 6.5	20% w/v polyethylene glycol 8000
19	0.2 M ammonium acetate	0.1 M Tris hydrochloride pH 8.5	30% v/v iso-propanol
20	0.2 M ammonium sulphate	0.1 M Sodium acetate trihydrate pH 4.6	25% w/v polyethylene glycol 4000
21	0.2 M magnesium acetate tetrahydrate	0.1 M sodium cacodylate pH 6.5	30% v/v 2-methyl-2,4-pentanediol
22	0.2 M sodium acetate trihydrate	0.1 M Tris hydrochloride pH 8.5	30% w/v polyethylene glycol 4000
23	0.2 M magnesium chloride hexahydrate	0.1 M Sodium HEPES pH 7.5	30% v/v polyethylene glycol 400
24	0.2 M Calcium chloride dehydrate	0.1 M Sodium acetate trihydrate pH 4.6	20% v/v iso-propanol
25	None	0.1 M imidazole pH 6.5	1 M Sodium acetate trihydrate
26	0.2 M ammonium acetate	0.1 M tri-sodium citrate dihydrate pH 5.6	30% v/v 2-methyl-2,4-pentanediol
27	0.2 M tri-Sodium citrate dehydrate	0.1 M Sodium HEPES pH 7.5	20% v/v iso-propanol
28	0.2 M sodium acetate trihydrate	0.1 M sodium cacodylate pH 6.5	30% w/v polyethylene glycol 8000
29	None	0.1 M Sodium HEPES pH7.5	0.8 M Potassium sodium tartrate tetrahydrate
30	0.2 M ammonium sulphate	None	30% w/v polyethylene glycol 8000
31	0.2 M ammonium sulphate	None	30% w/v polyethylene glycol 4000
32	None	None	2 M Ammonium sulfate
33	None	None	4 M Sodium formate
34	None	0.1 M sodium acetate trihydrate pH 4.6	2 M Sodium formate
35	None	0.1 M Sodium HEPES pH 7.5	0.8 M Sodium dihydrogen phosphate 0.8 M potassium dihydrogen phosphate
36	None	0.1 M Tris hydrochloride pH 8.5	8% w/v polyethylene glycol 8000
37	None	0.1 M Sodium acetate trihydrate pH 4.6	8% w/v polyethylene glycol 4000
38	None	0.1 M Sodium HEPES pH 7.5	1.4 M tri-sodium citrate dihydrate
39	None	0.1 M Sodium HEPES pH 7.5	2% v/v polyethylene glycol 400 2 M ammonium sulfate
40	None	0.1 M tri-sodium citrate dihydrate pH 5.6	20% v/v iso-propanol 20% w/v polyethylene glycol 4000
41	None	0.1 M Sodium HEPES pH 7.5	10% v/v iso-propanol 20% w/v polyethylene glycol 4000
42	0.05 M mono-Potassium dihydrogen Phosphate	None	20% w/v polyethylene Glycol 8000
43	None	None	30% w/v polyethylene glycol 1500
44	None	None	0.2 M magnesium formate
45	0.2 M Zinc acetate dehydrate	0.1 M sodium cacodylate pH 6.5	18% w/v polyethylene glycol 8000
46	0.2 M calcium acetate hydrate	0.1 M sodium cacodylate pH 6.5	18% w/v polyethylene glycol 8000
47	None	0.1 M Sodium acetate trihydrate pH 4.6	2 M Ammonium sulfate
48	None	0.1 M Tris hydrochloride pH 8.5	2 M Ammonium dihydrogen phosphate
49	1 M Lithium sulfate monohydrate	None	2% w/v polyethylene glycol 8000
50	0.5 M lithium sulfate monohydrate	None	15% w/v polyethylene glycol 8000

Table 2.8 Crystal screen 2

Reagent	Salt	Buffer	Precipitant
1	2.0 M sodium chloride	None	10% w/v PEG 6000
2	0.01 M Hexadecyltrimethylammonium Bromide	None	0.5 M sodium chloride 0.01 M magnesium chloride hexahydrate
3	None	None	25% v/v ethylene glycol
4	None	None	35% v/v dioxane
5	2.0 M ammonium sulfate	None	5% v/v iso-propanol
6	None	None	1.0 M imidazole pH 7.0
7	None	None	10% w/v polyethylene glycol 1000 10% w/v polyethylene glycol 8000
8	1.5 M sodium chloride	None	10% v/v ethanol
9	None	0.1 M sodium acetate trihydrate pH 4.6	2.0 M sodium chloride
10	0.2 M sodium chloride	0.1 M sodium acetate trihydrate pH 4.6	30 % v/v MPD
11	0.01 M cobaltous chloride hexahydrate	0.1 M sodium acetate trihydrate pH 4.6	1.0 M 1,6 hexanediol
12	0.1 M cadmium chloride dehydrate	0.1 M sodium acetate trihydrate pH 4.6	30% v/v polyethylene glycol 400
13	0.2 M ammonium sulfate	0.1 M sodium acetate trihydrate pH 4.6	30% w/v polyethylene glycol monomethyl ether 2000
14	0.2 M potassium sodium tartrate tetrahydrate	0.1 M tri-sodium citrate dihydrate pH 5.6	2.0 M ammonium sulfate
15	0.5 M ammonium sulfate	0.1 M tri-sodium citrate dihydrate pH 5.6	1.0 M lithium sulfate monohydrate
16	0.5 M sodium chloride	0.1 M tri-sodium citrate dihydrate pH 5.6	2% w/v ethylene imine polymer
17	None	0.1 M tri-sodium citrate dihydrate pH 5.6	35% v/v tert-butanol
18	0.01 M ferric chloride hexahydrate	0.1 M tri-sodium citrate dihydrate pH 5.6	10% v/v jeffamine M-600
19	None	0.1 M tri-sodium citrate dihydrate pH 5.6	2.5 M 1,6 hexanediol
20	None	0.1 M MES pH 6.5	1.6 M magnesium sulfate heptahydrate
21	0.1 M sodium dihydrogen phosphate 0.1 M potassium dihydrogen phosphate	0.1 M MES pH 6.5	2.0 M sodium chloride
22	None	0.1 M MES pH 6.5	12% w/v polyethylene glycol 20,000
23	1.6 M ammonium sulfate	0.1 M MES pH 6.5	10% v/v dioxane
24	0.05 M cesium chloride	0.1 M MES pH 6.5	30% v/v jeffamine M-600
25	0.01 M cobaltous chloride hexahydrate	0.1 M MES pH 6.5	1.8 M ammonium sulfate
26	0.2 M ammonium sulfate	0.1 M MES pH 6.5	30% w/v polyethylene glycol monomethyl ether 5000
27	0.01 M zinc sulfate heptahydrate	0.1 M MES pH 6.5	25% v/v polyethylene glycol monomethyl ether 550
28	None	None	1.6 M tri-sodium citrate dihydrate
29	0.5 M ammonium sulfate	0.1 M HEPES pH 7.5	30% v/v MPD
30	None	0.1 M HEPES pH 7.5	10% w/v polyethylene glycol 6000 5% v/v MPD
31	None	0.1 M HEPES pH 7.5	20% v/v jeffamine M-600
32	0.1 M sodium chloride	0.1 M HEPES pH 7.5	1.6 M ammonium sulfate
33	None	0.1 M HEPES pH 7.5	2.0 M ammonium formate
34	0.05 M cadmium sulfate hydrate	0.1 M HEPES pH 7.5	1.0 M sodium acetate
35	None	0.1 M HEPES pH 7.5	70% v/v MPD
36	None	0.1 M HEPES pH 7.5	4.3 M sodium chloride
37	None	0.1 M HEPES pH 7.5	10% w/v polyethylene glycol 8000 8% v/v ethylene glycol
38	None	0.1 M HEPES pH 7.5	20% w/v polyethylene glycol 10000
39	0.2 M magnesium chloride hexahydrate	0.1 M tris pH 8.5	3.4 M 1,6 hexanediol
40	None	0.1 M tris pH 8.5	25% v/v tert-butanol
41	0.01 M nickel (II) chloride hexahydrate	0.1 M tris pH 8.5	1.0 M lithium sulfate monohydrate
42	1.5 M ammonium sulfate	0.1 M tris pH 8.5	12% v/v glycerol anhydrous
43	0.2 M ammonium dihydrogen phosphate	0.1 M tris pH 8.5	50% v/v MPD
44	None	0.1 M tris pH 8.5	20% v/v ethanol
45	0.01 M nickel (II) chloride hexahydrate	0.1 M tris pH 8.5	20% w/v polyethylene glycol monomethyl ether 2000
46	0.1 M sodium chloride	0.1 M bicine pH 9.0	20% w/v polyethylene glycol monomethyl ether 550
47	None	0.1 M bicine pH 9.0	2.0 M magnesium chloride hexahydrate
48	2% v/v Dioxane	0.1 M bicine pH 9.0	10% w/v polyethylene glycol 20000

3.0 Cloning, expression and purification of recombinant human PAP1

3.1 Overview

This chapter describes cloning of the human PAP1 gene from cDNA and the subsequent development of a recombinant expression system, capable of producing catalytically active recombinant human PAP1 at high levels, which could be isolated to a high purity. Various biochemical, kinetic (Chapter 4) and functional studies (Chapter 5) were then carried out.

3.2 Cloning of human PAP1 gene

Cell pellets ($\sim 1 \times 10^7$ cells) of two human cancer cell lines, SW620 and Hs 578T (American Type Culture Collection No. CCL-227 and HTB-126 respectively), were obtained from workers in the laboratory of Dr. Susan McDonnell (DCU, Ireland). SW620 is derived from a metastatic lymph node colorectal adenocarcinoma. Hs 578T was derived from a carcinoma of the breast. The SW620 cells (suspended in 1 ml of 50 mM potassium phosphate buffer, pH8.0) were found to be positive for PAP1 activity (2.8 units/ml) as determined by the method described in Section 2.15.2. In contrast, the Hs 578T cells registered a negligible 0.08 units/ml by equivalent analysis. Total RNA was isolated from a fresh SW620 cell pellet as described in Section 2.4.4. Figure 3.1A shows analysis of this RNA by agarose gel electrophoresis (Section 2.5). Sharp, clear 28S and 18S ribosomal RNA (rRNA) bands were an indication that the RNA preparation had not been degraded to any significant extent.

The sequence for putative human PAP1 mRNA (AJ278828, Table 1.4, Figure 3.2) was obtained from GenBank (Benson *et al.*, 1996). This putative sequence was confirmed during the progression of this work (Dando *et al.*, 2003). The 707 bp sequence contains a 630 bp open reading frame (ORF) for human PAP1 (*Hsa-pap1*). Primers PAPHsA and PAPHsB (Table 2.2) were designed external to this ORF to amplify the putative PAP1 gene (Figure 3.2) as a 698 bp fragment. cDNA was generated from the SW620 RNA using the specific reverse primer PAPHsB as described in Section 2.8.1. The cDNA was used as template for PCR (Section 2.8.2) using primers PAPHsA and PAPHsB. The resulting RT-PCR product, analysed by agarose gel electrophoresis (Section 2.5), is shown in Figure 3.1B. A band corresponding to the expected size of 698 bp was obtained. Some non-specific bands were also present. Several attempts were made to eliminate these bands by increasing the annealing temperature during the PCR reaction (data not shown). However, these attempts were unsuccessful.

The RT-PCR products shown in Figure 3.1B were cloned into the TA cloning vector pCR2.1 (Table 2.3, Figure 2.1) as described in Section 2.9.1. Several white colonies were purified and plasmid DNA was isolated (Section 2.4.1). Screening for a TA clone, having the desired ~700 bp recombinant *Hsa-pap1* insert (*rHsa-pap1*), was carried out by enzymatic restriction analysis with *EcoR*I (Section 2.8). *EcoR*I specifically cuts the TA insert out of the pCR2.1 vector. A candidate clone was identified and further analysed by PCR using PAPHsF and PAPHsG (Figure 3.2). A band corresponding to the expected 638 bp *rHsa-pap1* product was obtained, as shown in Figure 3.1C.

A restriction digest of this clone (pRV1) analysed by agarose gel electrophoresis is shown in Figure 3.3. Restriction with *EcoR*I produces a band corresponding to the expected 713 bp fragment containing the *rHsa-pap1* sequence. The *EcoR*IV restriction resulted in a band corresponding to 492 bp, suggesting that the *rHsa-pap1* sequence had inserted in the opposite orientation to the *lacZ* α ORF of pCR2.1 (a band corresponding to 245 bp would have indicated the same orientation as *lacZ* α). The *Nco*I restriction similarly supported this opposite orientation. Single restriction with *Hind*III and *Bam*HI produced bands corresponding to the expected 4603 bp linearised plasmid. The successful cloning of *rHsa-pap1* was confirmed by DNA sequencing (Section 2.10). The sequencing data is given in Appendix B. A map of pRV1 is shown in Figure 3.4.

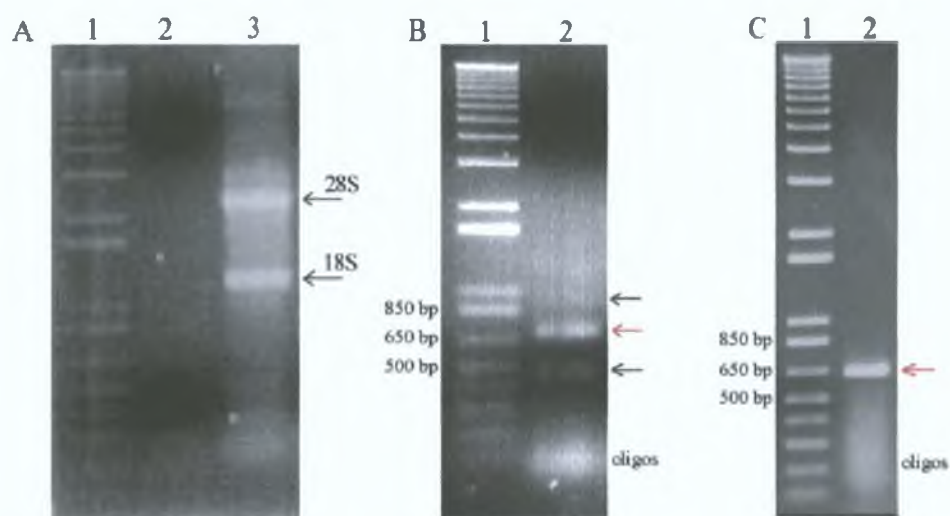


Figure 3.1 Cloning of human PAP1 gene

1% agarose gels. (A): Lane 1, DNA Ladder (not quantifying for RNA); Lane 2, RNA isolated from SW620 cells (Section 2.4.4) treated with Ribonuclease A (Section 2.8); Lane 3, RNA isolated from SW620 cells. Prominent 18S and 28S rRNA bands are indicated (B): Lane 1, DNA Ladder (sizes as in Figure 2.5); Lane 2, product of RT-PCR (Section 2.8.1 & 2.8.2) on SW620 RNA using primers PAPHsA and PAPHsB (Table 2.2, Figure 3.2). The band corresponding to the expected 698 bp PAP1 product is indicated by red arrow, while non-specific bands are indicated by black arrows. (C): Lane 1, DNA Ladder; Lane 2, product of PCR (Section 2.8.2) on pRV1 using primers PAPHsF and PAPHsG (Table 2.2). The band corresponding to the expected 638 bp PAP1 product is indicated by red arrow.

AJ278828 : AGTCGCAACAGAGCAGGTCCGAGGCACAGCCCGATCCCGCCATGGAGCAGCCGAGGAAGGCGTGGTAGTGACGGGATTGGCCCTTTTGGGGAACACAC : 101
PAPHsA PAPHsF

AJ278828 : CGTGAACGCCAGTGGATTGCAGTTCAGGAGCTAGAAAAGCTAGGCCTTGGCGACAGCGTGGACCTGCATGTGTACGAGATCCGGTTGAGTACCAACAG : 202

AJ278828 : TCCAGAGACTCATCCCCGCCCTGTGGGAGAAGCACACTCCACAGCTGGTGGTGCATGTGGGGGTGTGAGGCATGGCGACCACAGTCACACTGGAGAAATGT : 303

AJ278828 : GGACACAACAAGGGCTACAAGGGGCTGGACAACAGCCGCTTTTGGCCCGGCTCCAGTGTGCGTGGAGGACGGGCTGAAAGCATGACTCCATCATCGA : 404

AJ278828 : CATGGATGCTGTGTGCAAGCGAGTCACCACGTTGGGCTGGATGTGTGGTGACCATCTCGCAGGATGCCGGCAGATATCTCTGGGACTTTACCTACTACA : 505
EcoRV

AJ278828 : CCTCTTTGTACCAGAGTCACGGTCGATCAGCCTTCGTCCACGTGCCCCACTGGGGAAGCCGTACAACGCGGACCAGCTGGCAGGGGCACTGAGAGCCATC : 606

AJ278828 : ATTGAGGAGATGTGGACCTCCTGGAGCAGTCAGAGGGCAAATCAACTATTGCCACAACACTGAGGGACGCTCAGGTCTCTAAGACCTCATCCTGCTG : 707
PAPHsG PAPHsB

Figure 3.2 Human PAP1 putative mRNA sequence
Nucleotide sequence AJ278828: human putative PAP1 mRNA. ORF is highlighted in blue. Binding locations of primers PAPHsA, PAPHsB, PAPHsF and PAPHsG (Table 2.2) are indicated. Restriction sites are marked in red. Illustrated using GenDoc (Section 2.11).

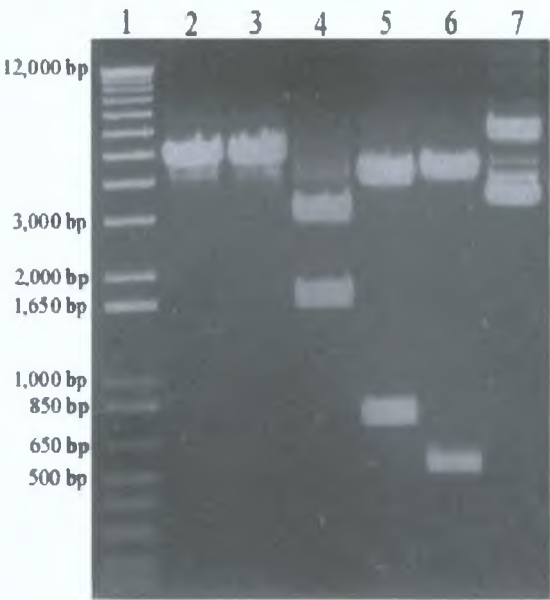


Figure 3.3 Verification digest of pRV1
Restriction digest of pRV1 clone analysed on 1% agarose gel (Section 2.5): Lane 1, DNA Ladder (sizes as in Figure 2.5); Lane 2, *Hind*III; Lane 3, *Bam*HI; Lane 4, *Nco*I; Lane 5, *Eco*R1; Lane 6, *Eco*RV; Lane 7, uncut (prominent covalently closed and open circular bands are visible as well as a fainter linear band).

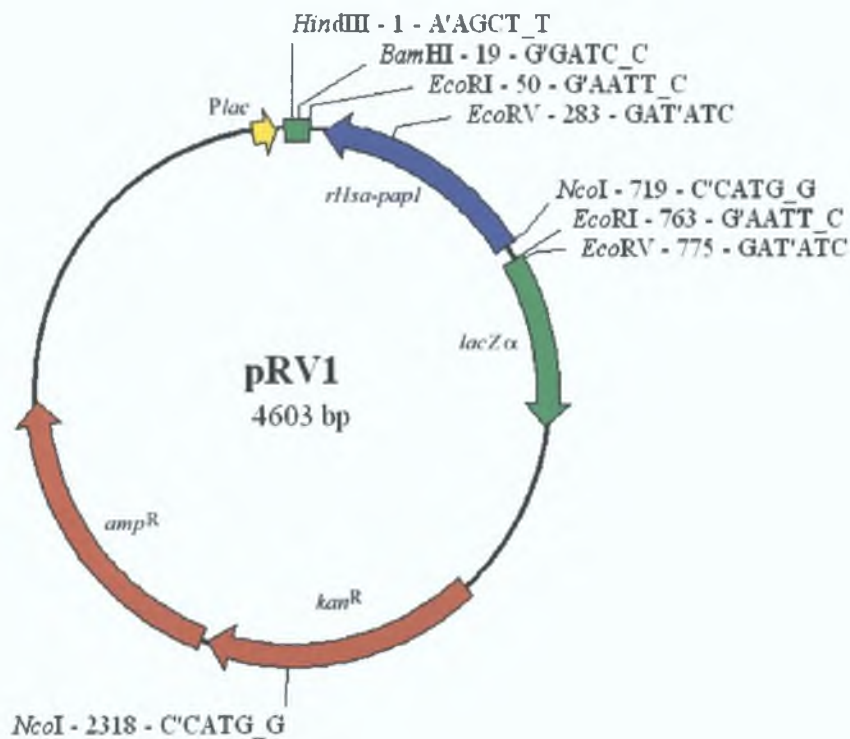


Figure 3.4 pRV1 plasmid map

The *rHsa-pap1* gene fragment (blue) is inserted in opposite orientation relative to the *LacZα* ORF (green), which is under the control of the *Plac* promoter (yellow). Ampicillin and kanamycin resistance genes (*amp^R* & *kan^R*) are shown in red. Some restriction sites are indicated with position and sequence. See Figure 2.1 for map of parent vector pCR2.1. Illustrated using pDRAW32 (Section 2.11).

3.3 Choice of expression host for recombinant human PAP1

The Gram-negative bacterium *E. coli* is one of the most commonly used host of recombinant heterologous protein expression (Baneyx, 1999). It grows rapidly at high density on inexpensive substrates. It has well-characterised genetics and many variable strains are available, as well as a broad range of compatible cloning and expression vectors. The major drawbacks of *E. coli* as an expression system, especially for recombinant eukaryotic proteins, include the inability to perform many of the post-translational modifications necessary for eukaryotic proteins, in particular glycosylation; and a limited ability to facilitate extensive disulphide bond formation (Makrides, 1996).

Studies to date on the native mammalian PAP1 enzymes indicate that the active forms are not glycosylated. A close sequence identity, as well as very similar biochemical and catalytic properties between eukaryotic and prokaryotic PAP1 enzymes, has been reported in Chapter 1. Active recombinant forms of several prokaryotic PAP1 enzymes have been successfully expressed in *E. coli* (Section 1.3.5), which has no endogenous PAP1 activity (Section 1.3.3). Thus, although there was some probability that

catalytically active recombinant human PAP1 (rHsaPAP1) would not be expressed in *E. coli*, it was decided to investigate this possibility.

3.4 Construction of *E. coli* expression system for recombinant human PAP1

In order to create a potential system for the expression of rHsaPAP1 in *E. coli*, the *rHsa-pap1* sequence (Section 3.2) was sub-cloned into *E. coli*-compatible expression plasmids. The plasmids pQE-60 and pPC225 were used, as presented in Sections 3.4.1 and 3.4.2. The resulting constructs had introduced 6 consecutive His codons (6xHis) as a 3' extension to the *rHsa-pap1* ORF. This His-tag would benefit protein purification (Section 3.7), assuming it did not interfere with catalytic activity. A construct lacking the 6xHis sequence was also created to provide the means of expressing non His-tagged rHsaPAP1 (Section 3.4.3).

3.4.1 Sub-cloning of human PAP1 gene into *E. coli* expression vector pQE-60

The vector pQE-60 (Table 2.3, Figure 2.2) features an optimised hybrid promoter-operator element (Bujard *et al.*, 1987) consisting of the phage T5 transcriptional promoter and a *lac* operator sequence, which increases *lac* protein (LacI) binding, enabling repression of this strong promoter. Associated with this operator region is a synthetic ribosome binding site (RBS) designed for high translation rates. The β -lactamase gene conferring resistance to ampicillin (*amp*^R) is present on the vector. Most significantly, the vector has six consecutive His codons (6xHis) followed by a stop codon, at the 3' end of the multiple cloning site (MCS). This enables translation of a fusion protein having a C-terminal His-tag, which can be exploited during protein purification (Section 3.7).

The *rHsa-pap1* sequence was sub-cloned from pRV1 into pQE-60 by the strategy outlined in Figure 3.5. The pQE-60 MCS conveniently has an *Nco*I site as the 5' insertion point. As can be seen in Figure 3.2 the *rHsa-pap1* sequence already features this site within the mRNA. *rHsa-pap1* was amplified by PCR with primers PAPHsH and PAPHsI (Table 2.2) and ligated into vector pCR2.1 as described in section 2.9.1. PAPHsH was designed to include the *Nco*I site at the 5' end of *rHsa-pap1* (Figure 3.2) and PAPHsI to tag a *Bam*HI site at the 3' end, replacing the stop codon. *rHsa-pap1* was excised from this intermediate clone by an *Nco*I/*Bam*HI restriction and ligated into pQE-60, which had been opened by an *Nco*I/*Bgl*II restriction. *Bam*HI and *Bgl*II sites

have compatible ends which, when fused, eliminate the restriction site. The cloning of *rHsa-pap1* into pQE-60 thus results in a 3' 6xHis fusion to the ORF.

Figure 3.6 shows a verification digest of this construct, pRV3. The *Nco1/HindIII* restriction resulted in a band corresponding to the expected 654 bp fragment containing the *rHsa-pap1*-6xHis fusion. *EcoR1*, *Nco1*, *HindIII* and *EcoRV* single restrictions produced bands corresponding to the expected 4046 bp linearised plasmid.

DNA sequencing of pRV3 (Section 2.10; Appendix B) revealed a 32 bp deletion (Figure 3.7), upstream of the *EcoR1* site within the vector sequence, affecting critical elements of the pQE-60 T5 promoter/*lac* operator. Other workers in the laboratory subsequently confirmed the presence of identical deletions in their pQE-60 constructs, indicating this deletion was carried forward from the original laboratory plasmid stock. Consequently, a new sample of pQE-60 was obtained from Qiagen and the cloning procedure in Figure 3.5 was repeated, yielding the construct pRV8, which was confirmed by DNA sequencing (Appendix B).

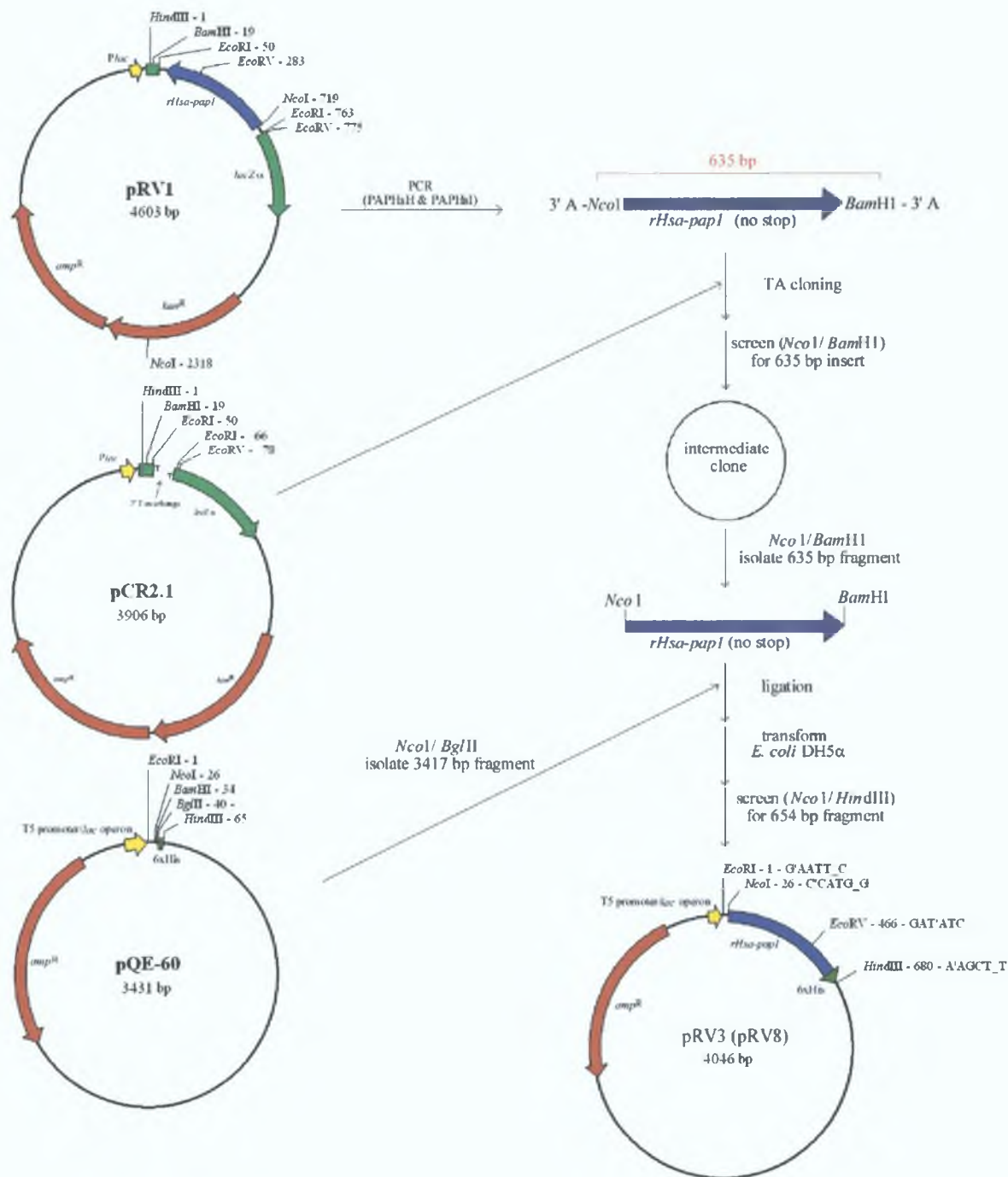


Figure 3.5 Cloning strategy for pRV3 (pRV8)

Outline of sub-cloning strategy from pRV1 to generate pRV3 (pRV8). PCR (Section 2.8.2) on pRV1 using primers PAPHsH and PAPHsI (Table 2.2) generates product containing *rHsa-pap1* without stop codon and having *NcoI* and *BamHI* site tags, which is ligated into pCR2.1. *rHsa-pap1* is excised from intermediate clone by *NcoI*/*BamHI* restriction and ligated into pQE-60 (Table 2.3, Figure 2.2), which has been opened with *NcoI*/*BglII* restriction. Resulting construct, pRV3, has 6xHis sequence cloned to 3' end of *rHsa-pap1* followed by stop codon. Restriction sites are indicated with positions (and sequence for final construct). Illustrated using pDRAW32 (Section 2.11).

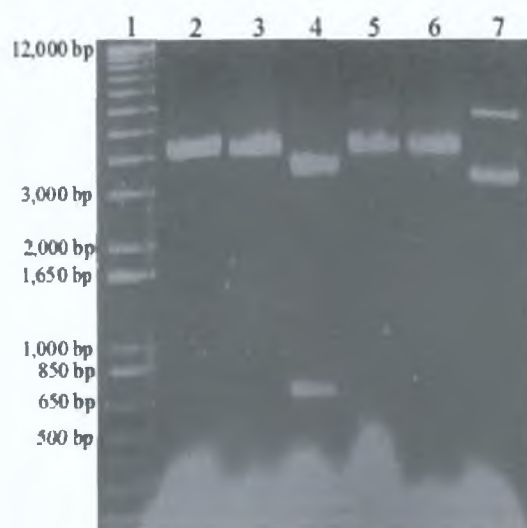


Figure 3.6 Verification digest of pRV3

Restriction digest of pRV3 construct analysed on 1% agarose gel (Section 2.5): Lane 1, DNA Ladder (sizes as in Figure 2.5); Lane 2, *EcoRI*; Lane 3, *NcoI*; Lane 4, *NcoI/HindIII*; Lane 5, *HindIII*; Lane 6, *EcoRV*; Lane 7, uncut (prominent covalently closed and open circular bands are visible). Due to lack of RNase treatment, significant RNA levels are visible at lower end of gel.

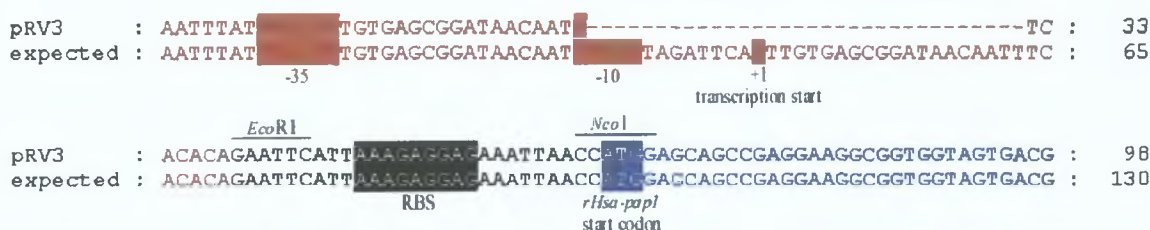


Figure 3.7 Deletion within pRV3

Alignment of the actual pRV3 sequence (Appendix B) with the expected sequence. Promoter region is highlighted in red with -35 box, -10 box and transcription start site indicated. Ribosome binding site (RBS) is highlighted in black. *rHsa-papI* ORF is highlighted in blue. *EcoRI* and *NcoI* restriction sites are marked. The 32 bp deletion encompasses the transcription start site and the -10 box. Illustrated using GenDoc (Section 2.11).

3.4.2 Sub-cloning of human PAP1 gene into *E. coli* expression vector pPC225

While awaiting the arrival of another sample of pQE-60 vector for the construction of pRV8 (Section 3.4.1), *rHsa-papI* was sub-cloned from pRV3 into the vector pPC225 (Table 2.3, Figure 2.3) as outlined in Figure 3.8. pPC225 features the *tac* transcriptional promoter (*P_{tac}*). *P_{tac}* is a hybrid promoter consisting of the -35 region from the *trp* promoter fused to the -10 region (Pribnow box), operator and RBS (Shine-Dalgarno sequence) from *Plac*. DeBoer *et al.* (1983) reported on the high efficiency of *P_{tac}* in expressing foreign genes in *E. coli* as opposed to the parental promoters. *P_{tac}* can be repressed by the LacI protein. pPC225 contains amp^R and another useful feature, noted during cloning, was that the 1800 bp fragment within the MCS (Figure 2.3) allowed for easy differentiation of linear from uncut vector.

An *EcoRI/HindIII* restriction enabled the entire *rHsa-pap1*-6xHis fusion to be transferred from pRV3 to pPC225, which had been opened with *EcoRI/HindIII*. Figure 3.9 shows a restriction digest of this construct, pRV5. The *EcoRI/HindIII* restriction produced a band corresponding to the expected 680 bp fragment containing the *rHsa-pap1*-6xHis fusion. *EcoRV*, *NcoI*, *EcoRI* and *HindIII* single restrictions produced bands corresponding to the expected 5235 bp linearised plasmid. pRV5 was subsequently verified by DNA sequencing (Appendix B). Figure 3.10 details the *rHsa-pap1* region of pRV5 showing the promoter elements, translated sequence, restriction sites and the 6xHis tag.

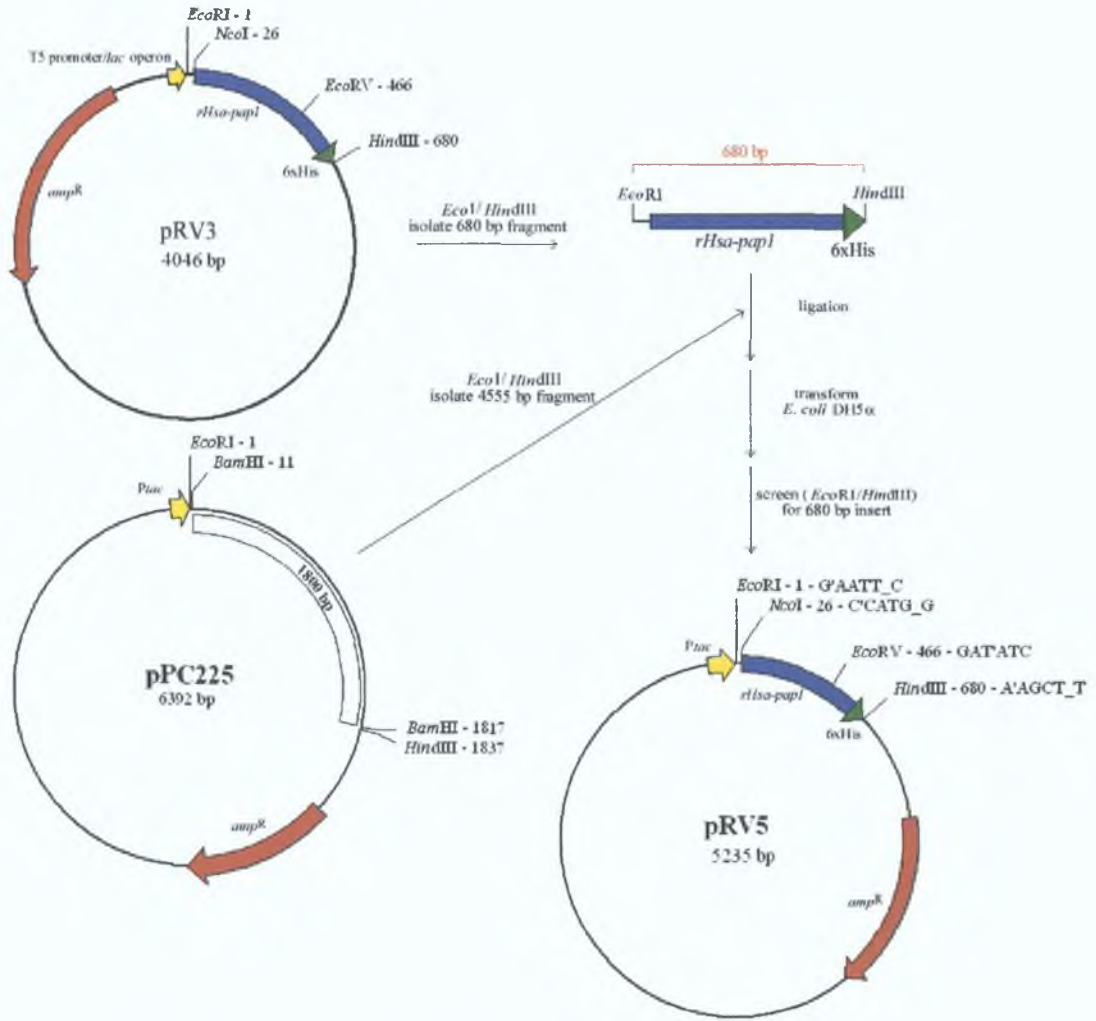


Figure 3.8 Cloning strategy for pRV5

Outline of sub-cloning strategy from pRV3 to generate pRV5. The *rHsa-pap1*-6xHis fusion is excised from pRV3 by restriction with *EcoRI* and *HindIII*. This 680 bp fragment is ligated into pPC225, which has been opened with *EcoRI* and *HindIII*. Resulting construct, pRV5, has *rHsa-pap1*-6xHis fusion under control of *Ptac* promoter. Restriction sites are indicated with positions (and sequences for final construct). Illustrated using pDRAW32 (Section 2.11).

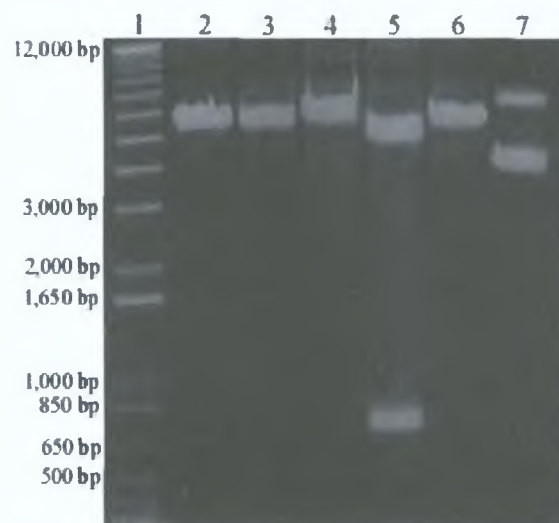


Figure 3.9 Verification digest of pRV5

Restriction digest of pRV5 construct analysed on 1% agarose gel (Section 2.5): Lane 1, DNA Ladder (sizes as in Figure 2.5); Lane 2, *EcoRV*; Lane 3, *NcoI*; Lane 4, *EcoRI*; Lane 5, *EcoRI/HindIII*; Lane 6, *HindIII*; Lane 7, uncut (prominent covalently closed and open circular bands are visible as well as a fainter linear band).

pRV5 : TCTGAAATGAGCTG **TGAC** ATTAATCATCGGCTCG **ATAA** GTGTGG **ATTGTGAGCGGATAACAATTTACAC** **ATG** AACA : 5235 bp
-35 -10 +1 transcription start RBS

pRV5 : **EcoRI** **GAATTCATT** **AAAGAGGA** **AAATTAACCA** **AT** **GAGCAGCCGAGGAAGGCGGTGGTAGTGACGGGATTTGGCCCTTTTGGGGAACA** : 83 bp
RBS *rHsa-pap1* start codon

PAPHsK →

T V N A S W I A V Q E L E K L **G L** G D S V D L H V Y E : 46 AA
pRV5 : CACCGTGAACGCCAGTTGGATTGCAGTTCAGGAGCTAGAAAAGCTAGGCCCTTGGCGACAGCGTGGACCTGCATGTGTACGAGA : 166 bp

I P V E Y Q T V Q R L I P A L W E K H S P Q L V V H V G : 74 AA
pRV5 : TTCCGGTTGAGTACCAAACAGTCCAGAGACTCATCCCGCCCTGTGGGAGAAGCACAGTCCACAGCTGGTGGTGCATGTGGGG : 249 bp

V S G M A T T V T L E K C G H N K G Y K G L D N C R F C : 102 AA
pRV5 : GTGT CAGGCATGGCGACCACAGT CACACTGGAGAAATGTGGACACAACAAGGGCTACAAGGGGCTGGACAAC T GCCGCTTTTG : 332 bp

P G S Q C C V E D G P E S I D S I I D M D A V C K R V : 129 AA
pRV5 : CCCCGGCTCCCAGTGCTGCGTGGAGGACGGGCTGAAAGCATTGACTCCATCATCGACATGGATGCTGTGTGCAAGCGAGTCA : 415 bp

T T L G L D V S V T I S Q D A G R Y L C D F T Y Y T S L : 157 AA
pRV5 : CCACGTTGGGCTGGATGTGTGCGTGACCATCTCGCAGGATGCCGGCAGATATCTCTGCGACTTTACCTACTACACCTCTTTG : 498 bp
EcoRV

Y Q S H G R S A F V H V P P L G K P Y N A D Q L G R A L : 185 AA
pRV5 : TACCAGAGTCACGGTCGATCAGCCTTCGTCCACGTGCCCCCACTGGGGAAGCCGTACAACGCGGACCAGCTGGGCAGGGCACT : 581 bp

R A I I E E M L D L L E Q S E G K I N Y C H K H G S H : 212 AA
pRV5 : GAGAGCCATCATTTGAGGAGATGTTGGACCTCCTGGAGCAGTCAGAGGGCAAAATCAACTATTGCCACAAACACGGATCT **AT** : 664 bp
linker

PAPHsL → TAA→ *BamHI* -clamp

H H H H H **HindIII** : 217 AA
pRV5 : **ACCATGACGATGATTA** **GCTTGGCTGTTTGGCGGATGAGAGAAGATTTTCAGCCTGATACAGATTAAATCAGAACGCAGAGAAG** : 747 bp
6xHis stop codon

Figure 3.10 pRV5 sequence detail

The *rHsa-pap1* (blue) and *Ptac* promoter (red) regions of pRV5 are shown in detail. The translation of *rHsa-pap1* is shown in blue above the ORF. The 6xHis tag sequence is shown in green, joined to *rHsa-pap1* by a Gly-Ser linker. Two ribosome binding sites (RBS, black) are shown. The first RBS upstream from the *rHsa-pap1* start codon was subcloned from pRV3 (Figure 3.5) while the second RBS was present on pPC225 associated with the *Ptac* promoter. Restriction enzymes and primers PAPHsK & PAPHsL (Table 2.2) are marked in red. For amino acid information see Appendix D. Illustrated using GenDoc (Section 2.11).

3.4.3 Construction of human PAP1 expression plasmid without His-tag

A strategy to sub-clone *rHsa-pap1* from pRV5 is shown in Figure 3.11. PCR was performed on pRV5 using primers PAPHsK and PAPHsL (Table 2.2, Figure 3.10). This resulted in a product encompassing the *EcoRI* site upstream from *rHsa-pap1* and the 6xHis being replaced with a stop codon followed by a *BamHI* site with restriction clamp. The use of this clamp allowed the PCR product to be restricted directly without the need of an intermediate clone such as was used in a previous strategy (Figure 3.5).

The PCR product was restricted by *Eco*R1/*Bam*H1 and ligated into pPC225, which had been opened by *Eco*R1/*Bam*H1.

Figure 3.12 shows a restriction digest of the resulting construct, pRV5_Δ6H. The *Eco*R1/*Bam*H1 restriction produced a band corresponding to the expected 658 bp fragment containing the *rHsa-pap1* sequence. *Eco*R1, *Bam*H1, *Eco*RV and *Hind*III single restrictions produced bands corresponding to the expected 5233 bp linearised plasmid. pRV5_Δ6H was verified by DNA sequencing (Appendix B). The pRV5_Δ6H construct is identical to pRV5 except that the 6xHis sequence has been deleted.

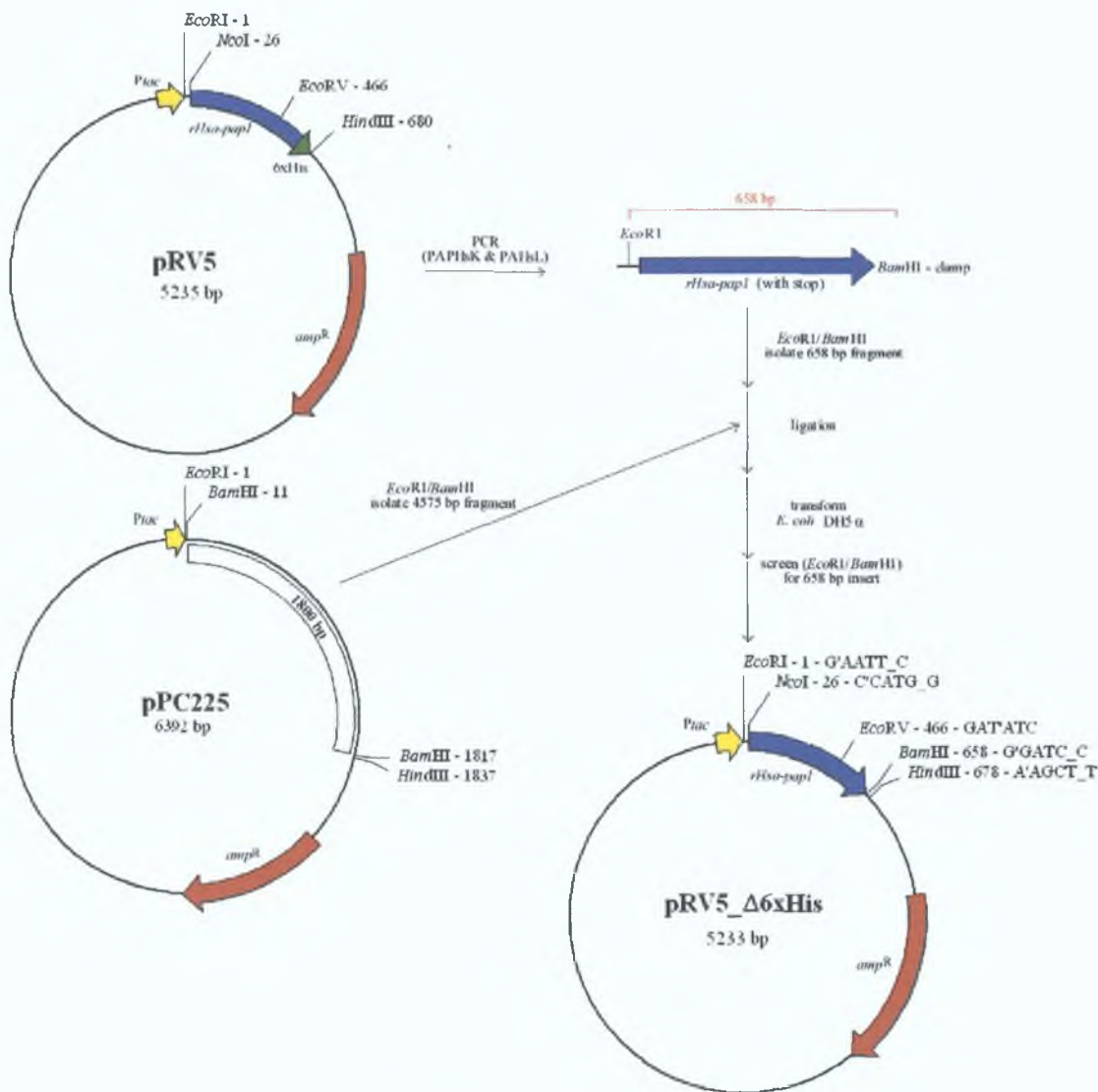


Figure 3.11 Cloning strategy for pRV5_Δ6H

Outline of sub-cloning strategy from pRV5 to generate pRV5_Δ6H. PCR (Section 2.8.2) on pRV5 using primers PAPHsK and PAPHsL (Table 2.2) generates product with *Eco*R1 site upstream from *rHsa-pap1* and having the 6xHis fusion replaced by stop codon followed by *Bam*H1 tag with clamp. Fragment is restricted with *Eco*R1 and *Bam*H1 and then ligated into pPC225, which has been opened with *Eco*R1 and *Bam*H1. pRV5_Δ6H has *rHsa-pap1* (without 6xHis) under control of *Ptac* promoter. Restriction sites are indicated with positions (and sequences for final construct). Illustrated using pDRAW32 (Section 2.11).

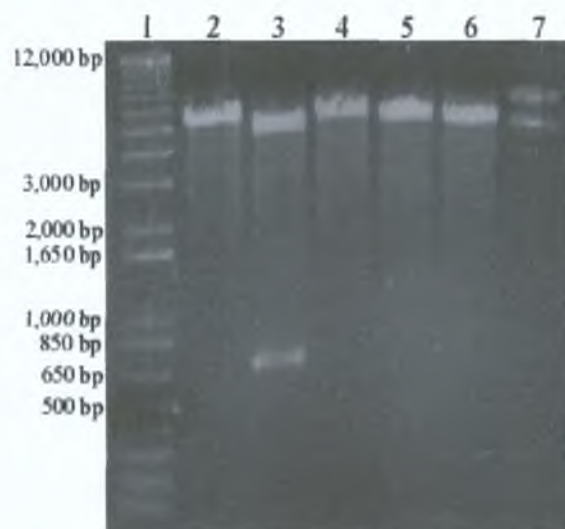


Figure 3.12 Verification digest of pRV5_Δ6H

Restriction digest of pRV5_Δ6H construct analysed on 1% agarose gel (Section 2.5). Lane 1, DNA Ladder (sizes as in Figure 2.5); Lane 2, *EcoRI*; Lane 3, *EcoRI/BamHI*; Lane 4, *BamHI*; Lane 5, *EcoRV*; Lane 6, *HindIII*; Lane 7, uncut (covalently closed and open circular bands are visible).

3.5 Preliminary characterisation of recombinant human PAP1

The effort of constructing *E. coli* expression systems for recombinant human PAP1 (Section 3.4) was justified by the successful expression of catalytically active native (*rHsaPAP1*) and His-tagged (*rHsaPAP1*_{6H}) enzyme.

The *rHsaPAP1* expression plasmid pRV5_Δ6H (Section 3.4.3), as well as the *rHsaPAP1*_{6H} expression plasmids pRV5 (Section 3.4.2) and pRV8 (Section 3.4.1), were transformed into *E. coli* XL10-Gold (Table 2.1) by the method described in Section 2.6.3. The plasmid vector pPC223 (Table 2.3), which serves as a negative control to pRV5 and pRV5_Δ6H, was also transformed.

Expression cultures were prepared by the methodology described in Section 2.12. Samples were taken 4 hrs after induction with 50 μM isopropyl-β-D-thiogalactopyranoside (IPTG, Section 3.8). Cleared lysate samples obtained from *rHsaPAP1* and *rHsaPAP1*_{6H} expression cultures (using pRV5_Δ6H and pRV5, respectively), as well as a negative control (using pPC223) were analysed for PAP1 activity as outlined in Section 2.15.2. The *rHsaPAP1* sample measured an activity of 2894.7 units/mg, while the *rHsaPAP1*_{6H} sample measured 2057.7 units/mg. The negative control measured 0.01 units/mg. Figure 3.13A shows SDS-PAGE analysis (Section 2.16) of these samples. Prominent protein bands are evident, running between the 20 and 24 kDa markers. The molecular weights of *rHsaPAP1* and *rHsaPAP1*_{6H} are

23,138 (Table 1.4) and 24,105 respectively, as deduced from their amino acid sequences. A subtle size difference of about 1 kDa is observed between *rHsaPAP1* and *rHsaPAP1_{6H}*, which is attributed to the His-tag and linker amino acid residues (Figure 3.10). Comparative kinetic analysis of the two forms of recombinant human PAP1 within cleared lysate is presented in Section 4.2.

Samples from *rHsaPAP1_{6H}* expression cultures using pRV5 and pRV8 were analysed by SDS-PAGE (Figure 3.13B). Both expression systems gave comparable levels of recombinant protein. This recombinant protein expression was not observed from expression cultures with pRV3 (data not shown), attributed to the deletion within the promoter sequence (Figure 3.7).

Zymogram analysis (Section 2.17) of a *rHsaPAP1_{6H}* cleared lysate sample provided a correlation between the prominent protein band and PAP1 activity (Figure 3.14). Further zymography, of purified *rHsaPAP1_{6H}*, can be seen in Figure 3.34. Confirmation that the His-tag is fused to the protein of expected size for human PAP1 was obtained by western blotting (Section 2.18) with anti-His antibody (Figure 3.15). Further western blot verification, including pPC223 and pRV5_Δ6H as negative controls, can be seen in Figure 5.19.

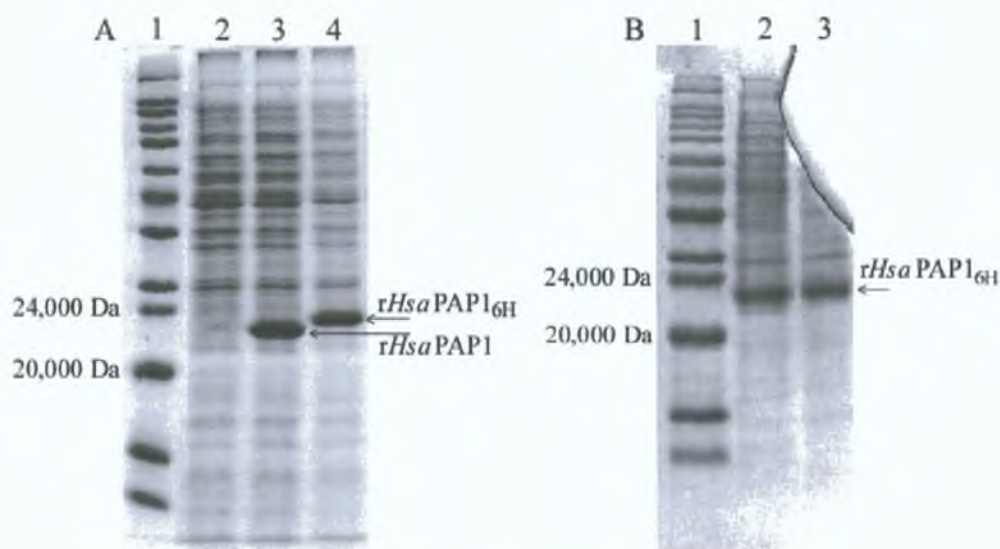


Figure 3.13 Expression of *rHsaPAP1* and *rHsaPAP1_{6H}* in *E. coli*

Analysis of cleared lysate samples by 15% SDS-PAGE (Section 2.16). (A): Lane 1, *M_r* (sizes as in Figure 2.6); Lane 2, *E. coli* XL10-Gold cultured with pPC223 (Table 2.3); Lane 3, *rHsaPAP1* (expressed from pRV5_Δ6H in *E. coli* XL10-Gold as outlined in Section 2.12); Lane 4, *rHsaPAP1_{6H}* (expressed from pRV5 in *E. coli* XL10-Gold). (B): Lane 1, *M_r* (sizes as in Figure 2.6); Lane 2, *rHsaPAP1_{6H}* (expressed from pRV5 in *E. coli* XL10-Gold); Lane 3, *rHsaPAP1_{6H}* (expressed from pRV8 in *E. coli* XL10-Gold).



Figure 3.14 Zymogram analysis of $rHsaPAP1_{6H}$

Zymogram analysis (Section 2.17) of cleared lysate sample from *E. coli* XL10-Gold culture expressing (Section 2.12) $rHsaPAP1_{6H}$ from pRV5 (Figure 3.8). Lane 1, 1/10 dilution; Lane 2, 1/100 dilution; Lane 3, 1/1000 dilution. The fluorescence corresponds to the prominent protein bands visible after coomassie staining (data not shown).

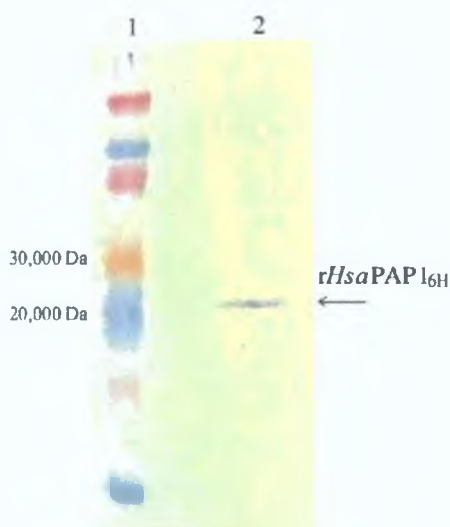


Figure 3.15 Western blot analysis of $rHsaPAP1_{6H}$

Western blot (Section 2.18) analysis, using anti-His antibodies, of cleared lysate sample from *E. coli* XL10-Gold culture expressing (Section 2.12) $rHsaPAP1_{6H}$ from pRV5 (Figure 3.8). Lane 1, CM_r (sizes given in Figure 2.7); Lane 2 cleared lysate.

3.6 Optimisation of recombinant human PAP1 expression in *E. coli*

The merits of different *E. coli* genotypes for the expression of $rHsaPAP1_{6H}$ were investigated. The expression plasmid pRV5 (Section 3.4.2, Figure 3.8) was transformed into various *E. coli* strains, namely DH5 α , INV α F', BL21, Nova Blue, XL10-Gold, Rosetta and RosettaBlue (Table 2.1), by the method described in Section 2.6.3.

The very similar DH5 α and INV α F' strains are used routinely in this work for the cloning and propagation of plasmid constructs. BL21 is a strain deficient in the OmpT

protease, purportedly allowing higher recovery of heterologous recombinant proteins. The Nova Blue and XL10-Gold strains contain F factors, which carry the *lacI*^q allele. The *lacI*^q allele is a promoter mutation that expresses the LacI repressor protein at high levels, resulting in strong repression of the *Ptac* promoter (Section 3.4.2) unless the inducer IPTG is added, which binds and inactivates LacI and thus induces transcription from *Ptac*.

Genes in both prokaryotes and eukaryotes exhibit a non-random usage of synonymous codons (A codon table is given in Appendix D). This means there is a bias toward one or two codons in most degenerate codon families. The frequency of codon usage is usually reflected by the abundance of the associated tRNAs in different cell backgrounds. This implies that recombinant heterologous genes enriched with codons rarely used in *E. coli* may experience poor translation efficiency (Makrides, 1996). Figure 3.16 shows the nucleotide sequence coding for rHsaPAP1_{6H}, as it is found on the pRV5 construct (Figure 3.8). Codons of optimal and minimal usage in *E. coli* have been indicated, according to international DNA sequence databases (Nakamura *et al.*, 2000). *E. coli* Rosetta and RosettaBlue are derivatives of the strains BL21 and NovaBlue, respectively, carrying the pRARE plasmid (Table 2.3, Figure 2.4). pRARE (Novy *et al.*, 2001) encodes genes for six tRNAs that recognise codons considered rare in *E. coli*. The use of pRARE during the expression of heterologous proteins is one strategy to overcome a possible codon bias of *E. coli*. Those codons, which are compensated for by the pRARE system, have been marked in Figure 3.16. It should be noted this system does not provide for all the rare codons.



Figure 3.16 *E. coli* codon bias relative to *rHsa-pap1*
 Nucleotide sequence coding for *rHsaPAP1_{6H}*, as found on pRV5 (Figure 3.8). Amino acids corresponding to each codon are specified. Codons of optimal usage in *E. coli* are highlighted green. Codons considered rare in *E. coli* are highlighted red. Those codons whose corresponding tRNAs are encoded on the pRARE plasmid (Table 2.3, Figure 2.4) are marked by red stars. For amino acid information see Appendix D. Illustrated using GenDoc (Section 2.11).

Expression of *rHsaPAP1_{6H}* was carried out by the methodology outlined in Section 2.12. The growth curves of these expression cultures, measured by optical absorbance at 600 nm (A_{600}), are shown in Figures 3.17 and 3.22. Growth rates, during exponential phase, are expressed as the increase in A_{600} per hour. The strains had comparable growth rates, although some took longer to reach the exponential growth phase. One reason for this may be slightly different inoculation concentrations. However, repeated experiments with consistent results (data not shown), conclude that some *E. coli* strains, especially DH5 α , exhibit a longer lag phase.

The cleared lysate samples obtained from these expression cultures were analysed by SDS-PAGE (Section 2.16), shown in Figure 3.18 (DH5 α and INVaF⁺), Figure 3.19 (BL21 and Rosetta), Figure 3.20 (Nova Blue and RosettaBlue) and Figure 3.23 (XL10-Gold). Also, Figure 3.21 shows samples taken from expression cultures in Nova Blue, RosettaBlue and XL10-Gold (induced with 50 μ M IPTG), compared to non-induced cultures. *rHsaPAP1_{6H}* activity was determined for each sample as outlined in Section 2.15.2. Using densitometry (Section 2.16.5) the percentage of total protein attributed to

the *rHsaPAP1_{6H}* band was approximated. In Figure 3.21 a discrepancy is observed between the ratios of densitometry and activity values comparing induced to non-induced samples. It is likely that this discrepancy is the result of underestimation in the densitometry readings obtained from the SDS gel. Irrespective of which set of data is taken, *E. coli* XL10-Gold was considered to give the optimal expression level.

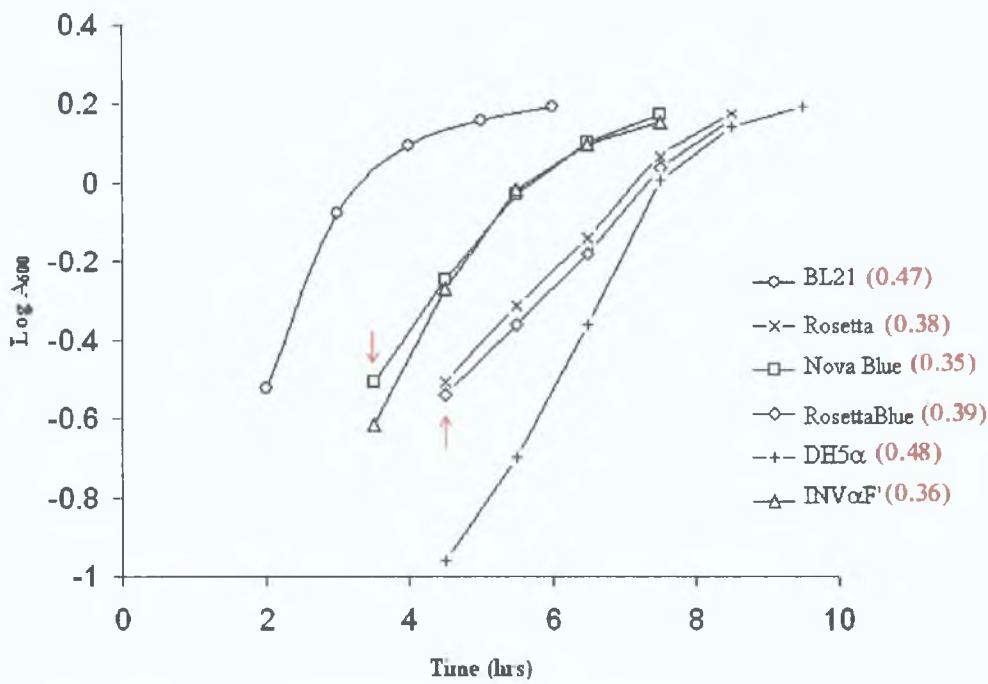


Figure 3.17 Growth curves for expression of *rHsaPAP1_{6H}* in *E. coli*
 Log absorbance readings at 600nm (Log *A*₆₀₀) taken during expression (Section 2.12) of *rHsaPAP1_{6H}* in *E. coli* DH5α and INVαF' (Figure 3.18); BL21 and Rosetta (Figure 3.19); Nova Blue and RosettaBlue (Figure 3.20). Induction of Nova Blue and RosettaBlue with 50 μM IPTG as indicated by arrows. Growth rates (increase of *A*₆₀₀ per hour) during exponential phase are given in red.

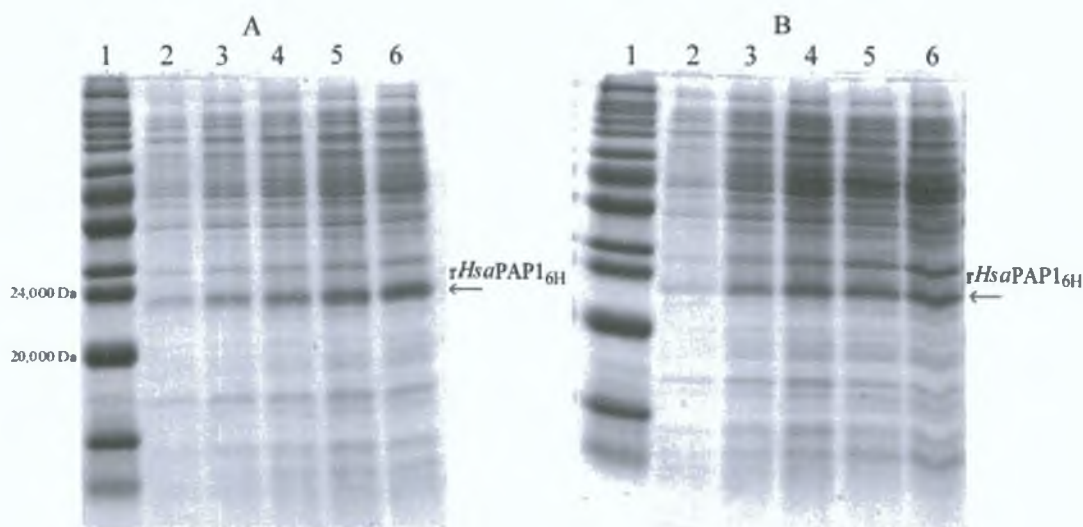


Figure 3.18 Expression of *rHsaPAP1_{6H}* in *E. coli* DH5α and INVαF'

Analysis of *rHsaPAP1_{6H}* expression in *E. coli* (Section 2.12) from pRV5 (Figure 3.8) by 15% SDS-PAGE (Section 2.16). *rHsaPAP1_{6H}* given as percentage of total protein, as determined by densitometry (Section 2.16.5). Growth curves shown in Figure 3.17. (A) DH5α: Lane 1, *M_r* (sizes given in Figure 2.6); Lane 2, 5.5 hrs (8%); Lane 3, 6.5 hrs (11%); Lane 4, 7.5 hrs (11.9%); Lane 5, 8.5 hrs (13.3%); Lane 6, 9.5 hrs (13.7%). (B) INVαF': Lane 1, *M_r*; Lane 2, 3.5 hrs (6%); Lane 3, 4.5 hrs (9.4%); Lane 4, 5.5 hrs (11.3%); Lane 5, 6.5 hrs (11.2%); Lane 6, 7.5 hrs (11.8%).

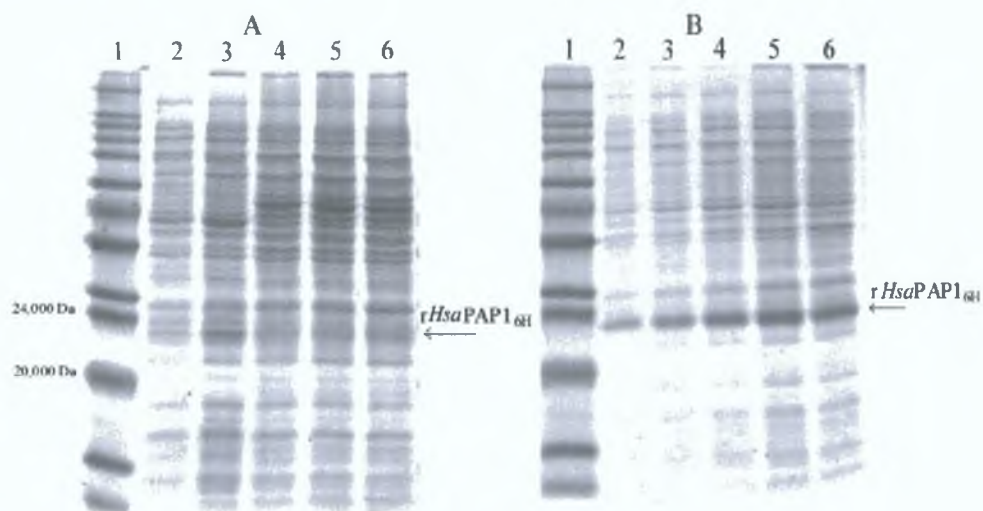


Figure 3.19 Expression of *rHsaPAP1_{6H}* in *E. coli* BL21 and Rosetta

Analysis of *rHsaPAP1_{6H}* expression in *E. coli* (Section 2.12) from pRV5 (Figure 3.8) by 15% SDS-PAGE (Section 2.16). *rHsaPAP1_{6H}* given as percentage of total protein, as determined by densitometry (Section 2.16.5). Growth curves shown in Figure 3.17. (A) BL21: Lane 1, *M_r* (sizes given in Figure 2.6); Lane 2, 2 hrs (4.6%); Lane 3, 3 hrs (6.8%); Lane 4, 4 hrs (3.8%); Lane 5, 5 hrs (5.4%); Lane 6, 6 hrs (5.5%). (B) Rosetta: Lane 1, *M_r*; Lane 2, 4.5 hrs (10.5%); Lane 3, 5.5 hrs (12.2%); Lane 4, 6.5 hrs (13%); Lane 5, 7.5 hrs (13.2%); Lane 6, 8.5 hrs (14.2%).

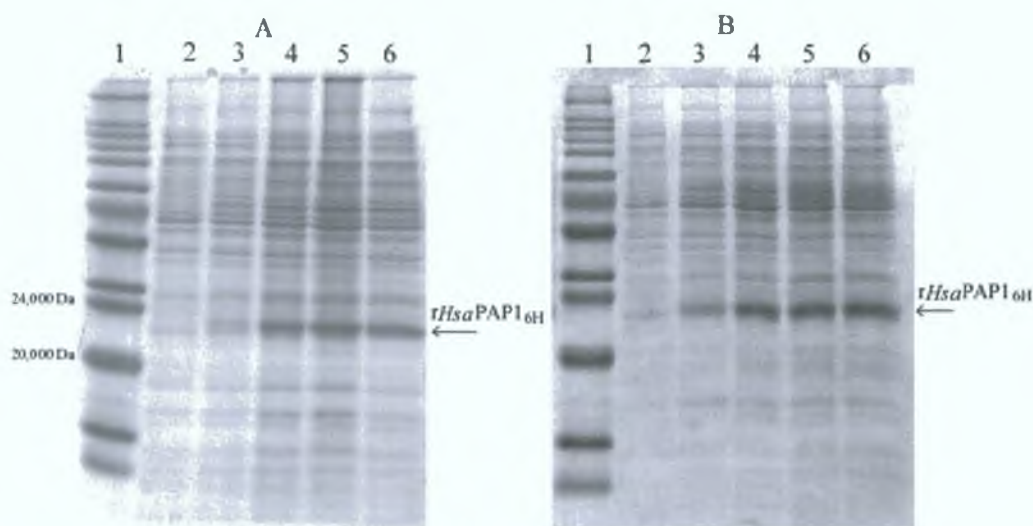


Figure 3.20 Expression of $rHsaPAP1_{6H}$ in *E. coli* Nova Blue and RosettaBlue

Analysis of $rHsaPAP1_{6H}$ expression in *E. coli* (Section 2.12) from pRV5 (Figure 3.8) by 15% SDS-PAGE (Section 2.16). $rHsaPAP1_{6H}$ given as percentage of total protein, as determined by densitometry (Section 2.16.5). Growth curves shown in Figure 3.17. (A) Nova Blue: Lane 1, M_r (sizes given in Figure 2.6); Lane 2, 3.5 hrs (0%) – point of induction with 50 mM IPTG; Lane 3, 4.5 hrs (3.7%); Lane 4, 6.5 hrs (10.2%); Lane 5, 7.5 hrs (10.2%); Lane 6, 8.5 hrs (11.4%). (B) RosettaBlue: Lane 1, M_r ; Lane 2, 4.5 hrs (0%) – point of induction with 50 mM IPTG; Lane 3, 5.5 hrs (10.6%); Lane 4, 6.5 hrs (13.5%); Lane 5, 7.5 hrs (13.5%); Lane 6, 8.5 hrs (13%).

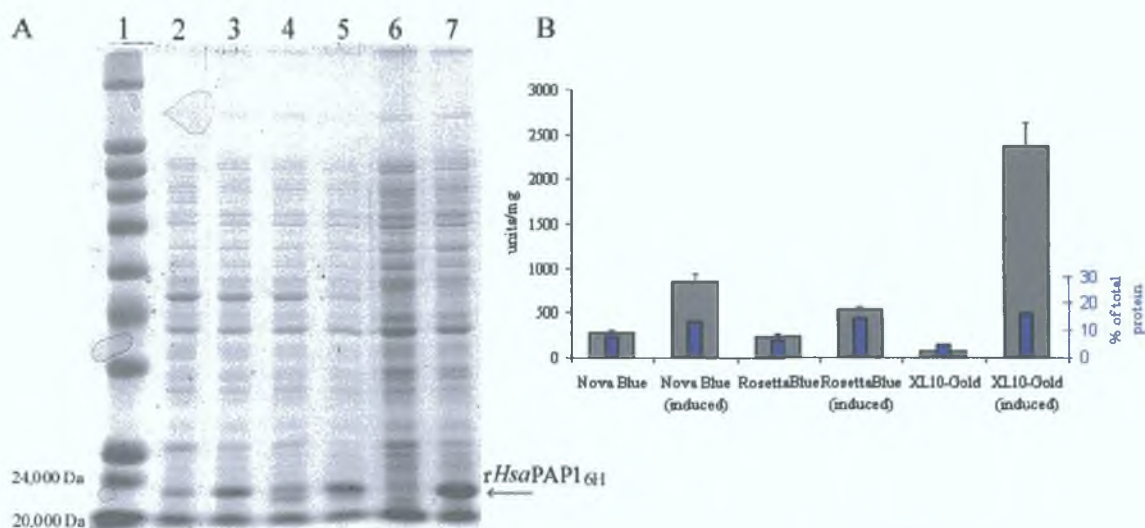


Figure 3.21 Expression control in *E. coli* Nova Blue, RosettaBlue and XL10-Gold

(A): Analysis of $rHsaPAP1_{6H}$ expression in *E. coli* (Section 2.12) from pRV5 (Figure 3.8) by 10% SDS-PAGE (Section 2.16). Samples taken 4 hrs after induction with 50 mM IPTG. Lane 1, M_r (sizes given in Figure 2.6); Lane 2, Nova Blue (non-induced); Lane 3, Nova Blue (induced); Lane 4, RosettaBlue (non-induced); Lane 5, RosettaBlue (induced); Lane 6, XL10-Gold (non-induced); Lane 7, XL10-Gold (induced). (B): $rHsaPAP1_{6H}$ activity (units per mg total protein) of each sample (Section 2.15.2). $rHsaPAP1_{6H}$ given as percentage of total protein, as determined by densitometry (Section 2.16.5).

3.6.1 Expression of recombinant human PAP1 in *E. coli* XL10-Gold

Of the various *E. coli* strains investigated (Section 3.6) for the expression of recombinant human PAP1 from pRV5 (Figure 3.8), XL10-Gold was considered the most efficient. The growth curve of an *E. coli* XL10-Gold culture, expressing (Section

2.12) *rHsaPAP1_{6H}* is shown in Figure 3.22. The culture was induced with 50 μ M IPTG when the A_{600} had reached 0.5. The specific *rHsaPAP1_{6H}* activity (units per mg of total protein) of hourly samples was calculated by the method outlined in Section 2.15.2, using protein concentrations determined by methods described in Section 2.14. Also shown in Figure 3.22 are growth curves of two *E. coli* XL10-Gold control cultures, one with pRV5 which was not induced with IPTG, the other with the negative control plasmid pPC223. Figure 3.23 shows the SDS-PAGE analysis (Section 2.16) of the *rHsaPAP1_{6H}* expression run. The prominent protein band, attributed to *rHsaPAP1_{6H}*, is estimated as percentage of total protein as determined by densitometry (Section 2.16.5). Figure 3.24 is an SDS-PAGE analysis comparing samples taken from an induced (50 μ M IPTG) and a non-induced cultures expressing *rHsaPAP1_{6H}* in *E. coli* XL10-Gold. To optimise the induction with IPTG, cultures of *E. coli* XL10-Gold containing the *rHsaPAP1_{6H}* expression construct pRV5, similar to those described above, were induced with various concentrations of IPTG (up to 1 mM), at A_{600} of 0.3. The SDS-Page analysis of these cultures is shown in Figure 3.25.

In order to investigate what consequences, if any, the time of induction has on expression, cultures of *E. coli* XL10-Gold containing the *rHsaPAP1_{6H}* expression construct pRV5 were induced (with 50 μ M IPTG) at various times from A_{600} of 0.3 to 0.5. The SDS-PAGE analysis of samples, taken 4 hrs after induction, is shown in Figure 3.26.

The length of sonication time used in the preparation of cleared lysate (Section 2.12.2) was optimised. Several samples from a culture of *E. coli* XL10-Gold expressing *rHsaPAP1_{6H}* from the construct pRV5 were subjected to increasing times of sonication from 15 to 180 sec. SDS-PAGE analysis of this trial is shown in Figure 3.27A. The *rHsaPAP1_{6H}* activities of the resulting cleared lysate samples (units per mg total protein) are shown in Figure 3.27B.

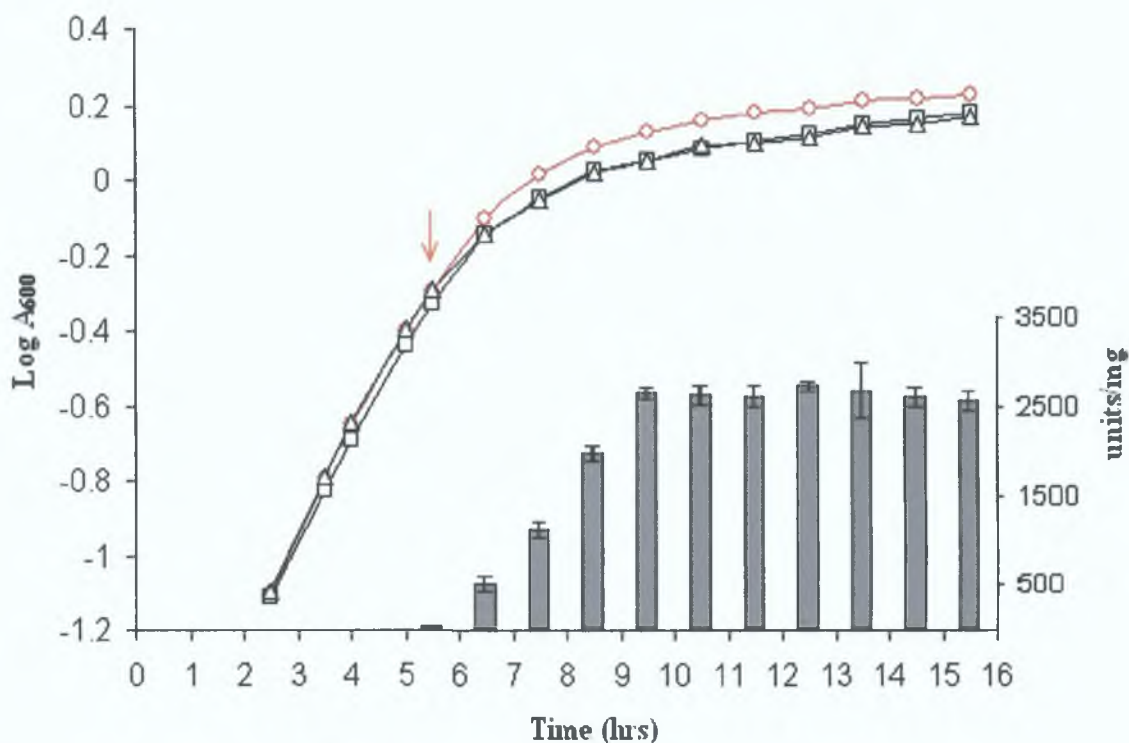


Figure 3.22 Growth curve for expression of *rHsaPAP1_{6H}* in *E. coli* XL10-Gold

Log absorbance readings at 600nm (Log A₆₀₀) taken during expression of *rHsaPAP1_{6H}* in *E. coli* XL10-Gold from pRV5 with (O) and without (O) induction using 50 μ M IPTG (indicated by red arrow). A culture with the vector pPC223 (Δ) served as a negative control for expression. Growth rates (increase in A₆₀₀ per hour) during exponential phase were of 0.27 for O, 0.19 for O and 0.22 for Δ . Specific activity of *rHsaPAP1_{6H}* (units per mg of total protein) shown by bar chart.

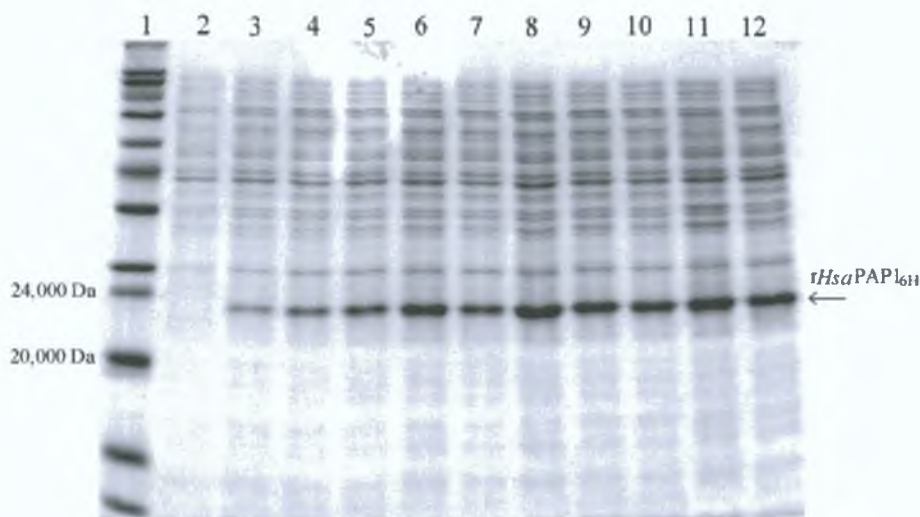


Figure 3.23 Expression of *rHsaPAP1_{6H}* in *E. coli* XL10-Gold

Analysis of *rHsaPAP1_{6H}* expression in *E. coli* XL10-Gold (Section 2.12) from pRV5 (Figure 3.8) by 15% SDS-PAGE (Section 2.16). *rHsaPAP1_{6H}* given as percentage of total protein, as determined by densitometry (Section 2.16.5). Growth curve shown in Figure 3.22. Lane 1, M_r (sizes given in Figure 2.6); Lane 2, 5.5 hrs (0%) – point of induction with 50 mM IPTG; Lane 3, 6.5 hrs (8.7%); Lane 4, 7.5 hrs (13.2%); Lane 5, 8.5 hrs (16.7%); Lane 6, 9.5 hrs (16.3%); Lane 7, 10.5 hrs (16.7%); Lane 8, 11.5 hrs (17.2%); Lane 9, 12.5 hrs (16.4%); Lane 10, 13.5 hrs (17.9%); Lane 11, 14.5 hrs (15.5%); Lane 12, 15.5 hrs (17.6%).

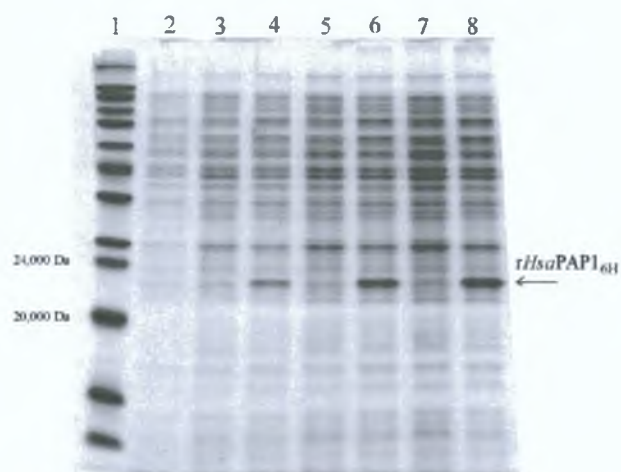


Figure 3.24 Expression control of *rHsaPAP1_{6H}* in *E. coli* XL10-Gold

Analysis of *rHsaPAP1_{6H}* expression in *E. coli* XL10-Gold (Section 2.12) from pRV5 (Figure 3.8) by 15% SDS-PAGE (Section 2.16). Induction with 50 μ M IPTG at A_{600} of 0.5. Lane 1, M_r (sizes given in Figure 2.6); Lane 2, point of induction; Lane 3, non-induced for 1 hr; Lane 4, induced for 1 hr; Lane 5, non-induced for 3 hrs; Lane 6, induced for 3 hrs; Lane 7, non-induced for 4 hrs; Lane 8, induced for 4 hrs.

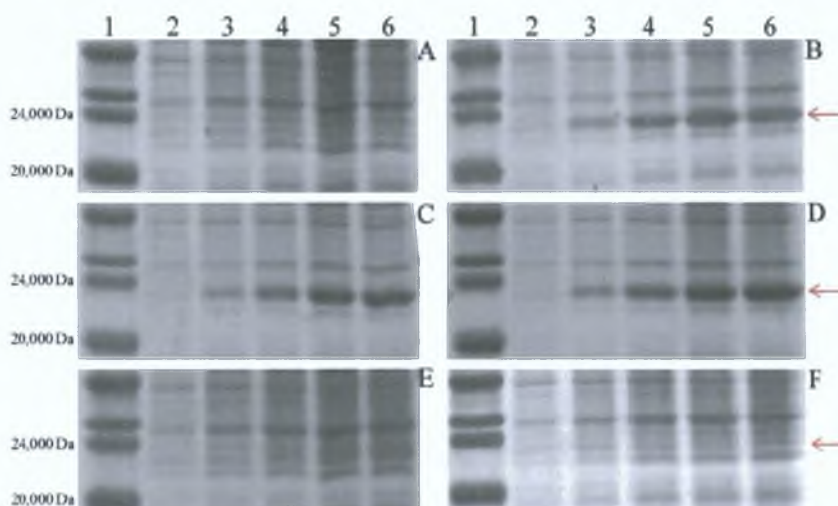


Figure 3.25 Effect of IPTG on *rHsaPAP1_{6H}* expression in *E. coli* XL10-Gold

Analysis of *rHsaPAP1_{6H}* expression in *E. coli* XL10-Gold (Section 2.12) from pRV5 (Figure 3.8) by 15% SDS-PAGE (Section 2.16). Induction at A_{600} of 0.3 with 0 μ M (A), 1 μ M (B), 5 μ M (C), 50 μ M (D), 500 μ M (E) and 1000 μ M (F) IPTG. Lane 1, M_r (sizes given in Figure 2.6); Lane 2, point of induction; Lane 3, 1 hr induction; Lane 4, 2 hrs induction; Lane 5, 3 hrs induction; Lane 6, 4 hrs induction. Location of *rHsaPAP1_{6H}* is indicated by red arrows.

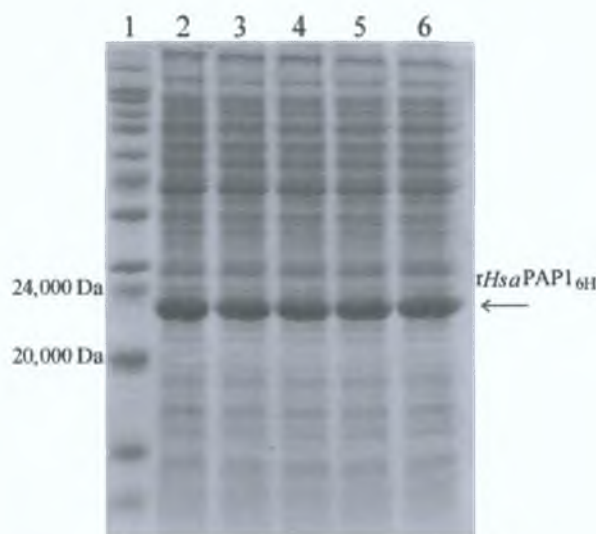


Figure 3.26 Effect of induction time on *rHsaPAP1_{6H}* expression

Analysis of *rHsaPAP1_{6H}* expression in *E. coli* XL10-Gold (Section 2.12) from pRV5 (Figure 3.8) by 15% SDS-PAGE (Section 2.16). Induction with 50 μ M IPTG at various A_{600} points, samples taken 4 hrs after induction. *rHsaPAP1_{6H}* given as percentage of total protein, as determined by densitometry (Section 2.16.5). Lane 1, M_r (sizes given in Figure 2.6); Lane 2, A_{600} of 0.3 (14.8%); Lane 3, A_{600} of 0.378 (15.6%); Lane 4, A_{600} of 0.409 (15.4%); Lane 5, A_{600} of 0.449 (16.7%); Lane 6, A_{600} of 0.505 (16.7%).

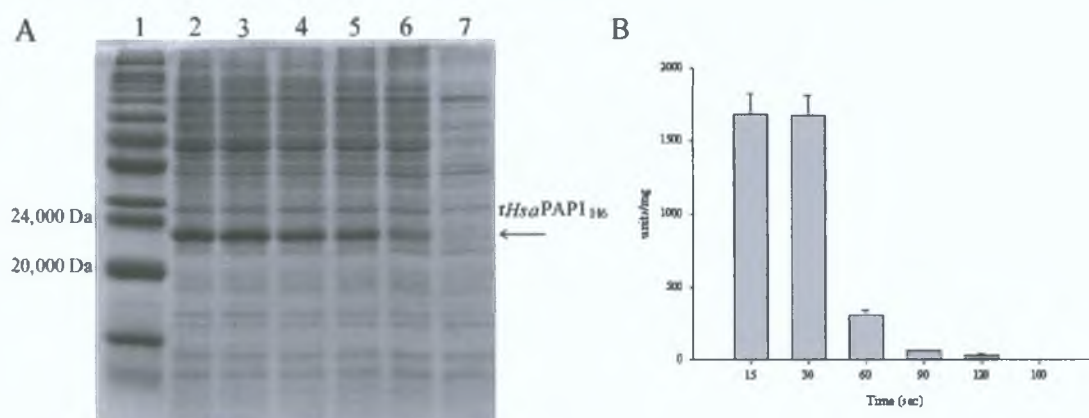


Figure 3.27 Effect of sonication time on lysis

Identical samples taken from *E. coli* XL10-Gold culture, expressing (Section 2.12) *rHsaPAP1_{6H}* from pRV5 (Figure 3.8), after 4 hrs induction with 50 μ M IPTG. Lysis carried out by various times of sonication. (A): Analysis of cleared lysate samples by 15% SDS-PAGE (Section 2.16). Lane 1, M_r (sizes given in Figure 2.6); Lane 2, 15 sec; Lane 3, 30 sec; Lane 4, 60 sec; Lane 5, 90 sec; Lane 6, 120 sec; Lane 7, 180 sec. (B): *rHsaPAP1_{6H}* activity (units per mg total protein) of each cleared lysate sample.

3.6.2 Addressing codon bias of *E. coli* by site-specific mutagenesis

As described in Section 3.6, the effect that codon bias may have on the expression of *rHsaPAP1_{6H}* was addressed using the pRARE plasmid system, supplementing limiting tRNAs to the *E. coli* phenotype. It was decided to complement this study with an alternative approach: using site-specific mutagenesis to substitute rare codons, in particular those at the 5' end of the gene, for those more commonly used in *E. coli* (Baneyx, 1999). It is known that the presence of rare codons near the 5' end of a

transcript can affect translational efficiency (Makrides, 1996; Hannig and Makrides, 1998) and it has been shown that 5' codon substitutions in a recombinant gene can improve heterologous protein expression (Nilsson and Mannervik, 2001).

At the 5' end of the sequence coding for *rHsaPAP1_{6H}* (Figure 3.16), Arg5[AGG] is the first codon considered rare in *E. coli*. It is followed by Lys6[AAG] which, while not considered a rare codon, is not the Lys codon optimally-used in *E. coli*. It was decided to create a mutant (Section 2.9.2) which has the wild-type Arg5 and Lys6 codons substituted with [CGT] and [AAA] respectively (mutant construct pRV5_R5^SK6^S, Table 2.3). This was achieved using primers PAPmR5K6-f & PAPmR5K6-r (Table 2.2). To confirm successful mutagenesis, prior to DNA sequencing, gradient PCR was used to differentiate between pRV5 and pRV5_R5^SK6^S (Figure 3.28), as described by Padmakumar and Varadarajan (2003). This screening method utilises the forward mutagenesis primer, together with a reverse primer specific to *amp*^R common to both plasmids. The differential hybridisation of the mutagenesis primer at higher temperatures provided an effective means of identifying the pRV5_R5^SK6^S plasmid. The DNA sequencing (Section 2.10) data for pRV5_R5^SK6^S is given in Appendix B.

Standard cultures of *E. coli* XL10-Gold, transformed with both pRV5 and pRV5_R5^SK6^S, were carried out as described in Section 2.12.1. SDS-PAGE (Section 2.16) analysis of these expression cultures is shown in Figure 3.29. Compared to the prominent *rHsaPAP1_{6H}* band visible for the pRV5 culture, the yield of *rHsaPAP1_{6H}* was greatly reduced for pRV5_R5^SK6^S.

Two further mutant derivatives of pRV5 were created, having only the Arg5 codon changed to each of the two optimally-used Arg codons in *E. coli*, namely: [CGC] (pRV5_R5^Sa) and [CGT] (pRV5_R5^Sb) (Table 2.3). The primer sets used were PAPmR5a-f/PAPmR5a-r and PAPmR5b-f/PAPmR5b-r respectively (Table 2.2). The DNA sequencing data for pRV5_R5^Sa and pRV5_R5^Sb is given in Appendix B. Expression cultures were analysed as before and the mutant constructs were found to yield less *rHsaPAP1_{6H}* product than obtained from pRV5 (Figure 3.30). These results were contrary to an anticipated increase (or even no change) in protein yield. To determine if this drop in protein yield was due to a decrease in stable mRNA, RT-PCR analysis (Sections 2.8.1 and 2.8.2) was performed on samples taken from each expression culture four hours after induction with 50 μ M IPTG. As can be seen in Figure 3.31, the level of stable *rHsaPAP1_{6H}* mRNA obtained from both pRV5_R5^Sa and

pRV5_R5^Sb was comparable to that from pRV5. Such analysis was also carried out on pRV5_R5^SK6^S, with similar results (data not shown). It may be that the reduction in rHsaPAP1_{6H} yield is due to adverse folding of the mRNA, leading to retarded translation.

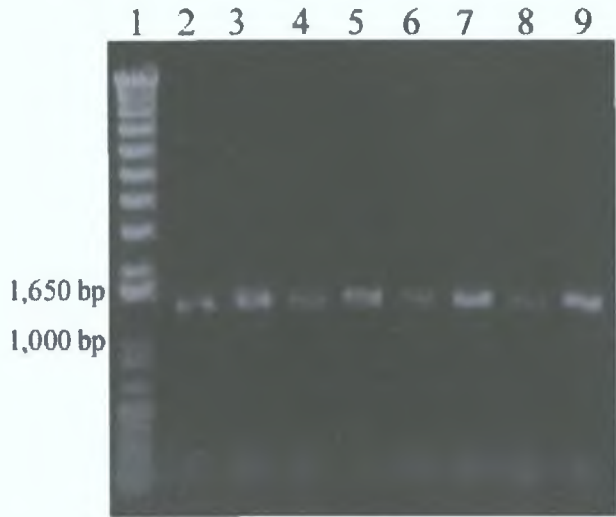


Figure 3.28 Confirmation of pRV5_R5^SK6^S by gradient PCR screening
 Differentiation between templates pRV5 and pRV5_R5^SK6^S using PCR (Section 2.8.2), analysed by 0.7% agarose gel electrophoresis (Section 2.5). Primers PAPmR5K6-f & pKKtestA (Table 2.2), giving expected product of 1458 bp, with 100 ng template, 15 cycles and increasing T_{ann}. Lane 1, DNA Ladder (sizes as in Figure 2.5); Lane 2, template 1 56.1°C; Lane 3, template 2 56.1°C; Lane 4, template 1 68.8°C; Lane 5, template 2 68.8°C; Lane 6, template 1 69.9°C; Lane 7, template 2 69.9°C; Lane 8, template 1 70.3°C; Lane 9, template 2 70.3°C.

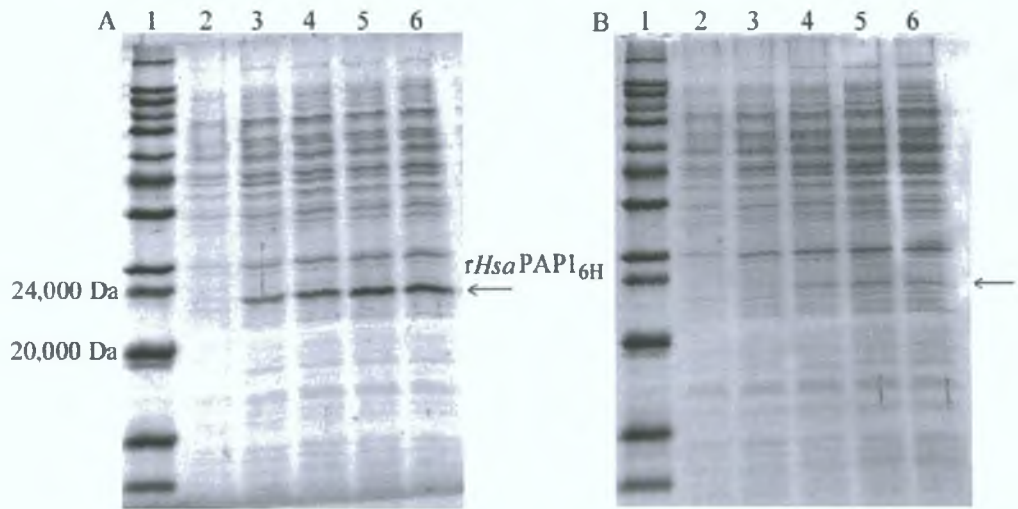


Figure 3.29 Effect of Arg5/Lys6 codon substitution on rHsaPAP1_{6H} expression
 Analysis of rHsaPAP1_{6H} (indicated by arrow) expression in *E. coli* (Section 2.12) from pRV5 (A, Figure 3.8) and pRV5_R5^SK6^S (B, Table 2.3) by 15% SDS-PAGE (Section 2.16). Lane 1, M_r (sizes given in Figure 2.6); Lane 2, time of induction with 50 µM IPTG; Lane 3, 1 hr; Lane 4, 2 hrs; Lane 5, 3 hrs; Lane 6, 4 hrs.

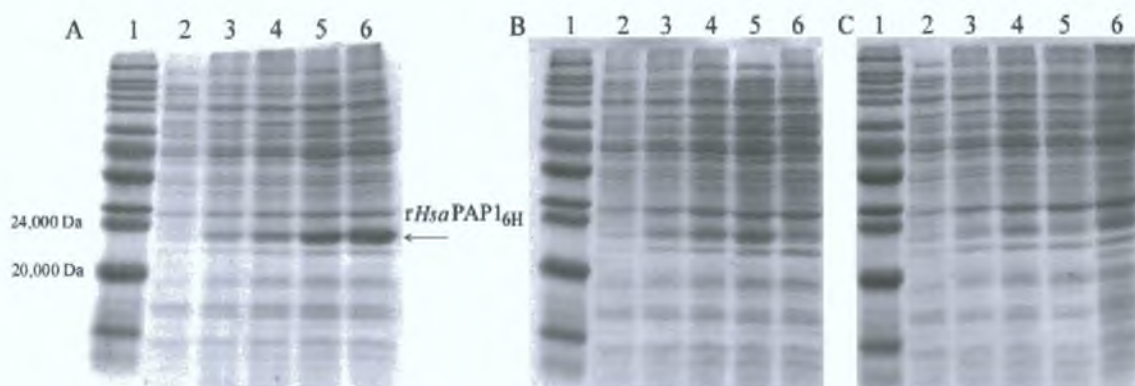


Figure 3.30 Effect of Arg5 codon substitution on *rHsaPAP1_{6H}* expression

Analysis of *rHsaPAP1_{6H}* expression in *E. coli* (Section 2.12) from pRV5 (A, Figure 3.8), pRV5_R5^Sa (B, Table 2.3) and pRV5_R5^Sb (C, Table 2.3) by 15% SDS-PAGE (Section 2.16). Lane 1, *M_r* (sizes given in Figure 2.6); Lane 2, time of induction with 50 μ M IPTG; Lane 3, 1 hr; Lane 4, 2 hrs; Lane 5, 3 hrs; Lane 6, 4 hrs.

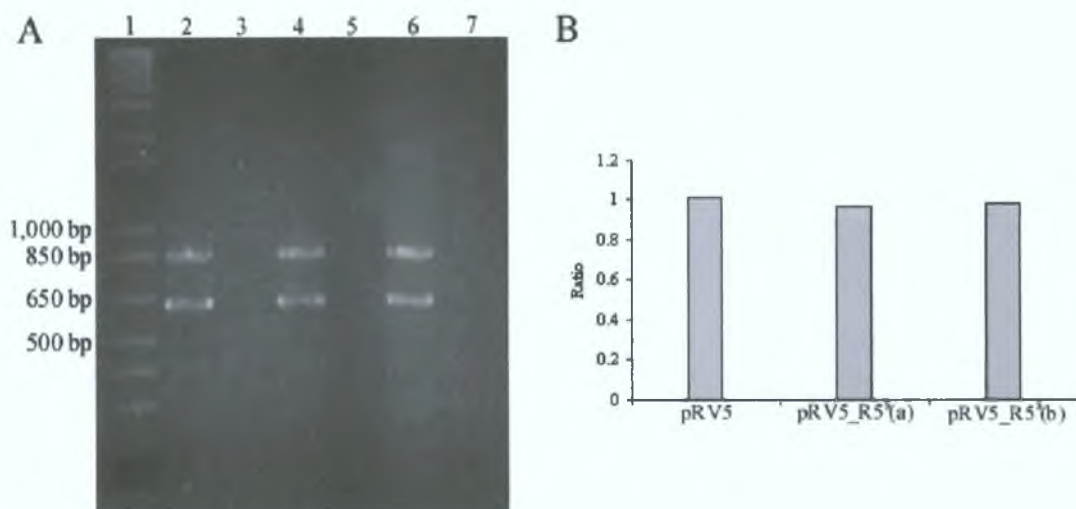


Figure 3.31 Level of transcription from pRV5, pRV5_R5^Sa and pRV5_R5^Sb

(A): RT-PCR analysis (Sections 2.8.1 to 2.8.2) of RNA isolated (Section 2.4.4) from *E. coli* XL10-Gold cultures harbouring pRV5, pRV5_R5^Sa and pRV5_R5^Sb (Table 2.3). *rHsa-pap1* specific primers PAPHsM & PAPHsN (627 bp product) and *amp^R* specific primers Amp-for & Amp-rev (856 bp product) were used (Table 2.2). Negative controls carried out by omitting RTase. Lane 1, DNA Ladder (sizes as in Figure 2.5); Lane 2, pRV5; Lane 3, pRV5 -ve control; Lane 4, pRV5_R5^Sa; Lane 5, pRV5_R5^Sa -ve control; Lane 6, pRV5_R5^Sb; Lane 7, pRV5_R5^Sb -ve control. Ratios of *rHsa-pap1* to *amp^R* product, determined by densitometry (Section 2.16.5), are shown in bar chart (B).

3.7 Purification of recombinant human PAP1

Optimisation of the expression system for recombinant human PAP1 in *E. coli* is described in Section 3.6. Since the addition of the His residues to the C-terminal end of the enzyme did not appear to alter its catalytic properties, it was decided to scale-up the production in order to carry out a detailed biochemical and kinetic study (Chapter 4) of the recombinant forms of the enzyme, both with (*rHsaPAP1_{6H}*) and without

(*rHsaPAP1*) the addition of the His-tag. *rHsaPAP1*_{6H} could be purified by single-stage immobilised metal affinity chromatography (IMAC).

IMAC was traditionally used with the chelating ligand iminodiacetic acid, charged with metal ions such as Ni^{2+} , Zn^{2+} or Cu^{2+} (Porath *et al.*, 1975). This ligand, however, only has three metal chelating sites, leading to weak binding and frequent leaching of the ions, resulting in impurity and metal-ion contamination. Nitrilotriacetic acid (NTA) is a tetradentate chelator, allowing it to occupy four of the six ligand binding sites in the coordination sphere of Ni^{2+} , therefore binding the metal ion far more stably, retaining it even under stringent wash conditions. In the Ni-NTA resin, two ligand binding sites of each Ni^{2+} are free to interact with the six consecutive imidazole groups of a 6xHis tag. Ni-NTA was used in this work to purify *rHsaPAP1*_{6H} as reported in the following Section 3.7.1.

3.7.1 IMAC purification of recombinant human PAP1

The standard procedure to purify *rHsaPAP1*_{6H} from cleared lysate, obtained from *E. coli* cultures expressing *rHsaPAP1*_{6H} from the construct pRV5 (Section 3.6), using Ni-NTA is described in Section 2.13.1. Both washing and elution was carried out using free imidazole. In order to determine the elution profile within the standard buffer of 50 mM potassium phosphate at pH 8.0, a series of small-scale elutions were set up. Following the mixing of 200 μl cleared lysate with 50 μl of Ni-NTA, the resin was collected by centrifugation and then mixed with 200 μl buffer containing increasing concentrations of imidazole (up to 250 mM). The *rHsaPAP1*_{6H} activity within each elution was determined as described in Section 2.15.2. The resulting imidazole elution profile is shown in Figure 3.32.

Figure 3.33A shows the SDS-PAGE analysis (Section 2.16) of a routine IMAC purification of *rHsaPAP1*_{6H}, using two 20 mM imidazole washes and two 200 mM imidazole elutions. By densitometry (Section 2.16.5) it was determined that *rHsaPAP1*_{6H} was more than 90% pure in both elutions. The protein concentration and specific activity values for the initial cleared lysate and both elutions are given in Figure 3.33B. Zymogram analysis (Section 2.17) was carried out on purified *rHsaPAP1*_{6H}, shown in Figure 3.34.

To determine suitable storage conditions for purified *rHsaPAP1*_{6H}, aliquots of purified enzyme (250 $\mu\text{g/ml}$) were stored at both 4°C and -20°C. Additional samples were stored

at -20°C with 10% and 40% glycerol. Samples were assayed for enzyme activity (Section 2.15.2) over the course of ten days. The samples stored at -20°C without glycerol suffered extreme precipitation. The accumulated storage profile data is presented in Figure 3.35. A similar sample of purified $rHsaPAP1_{6H}$, which was stored for two months at -20°C with 40% glycerol, retained almost two thirds of its original activity (data not shown).

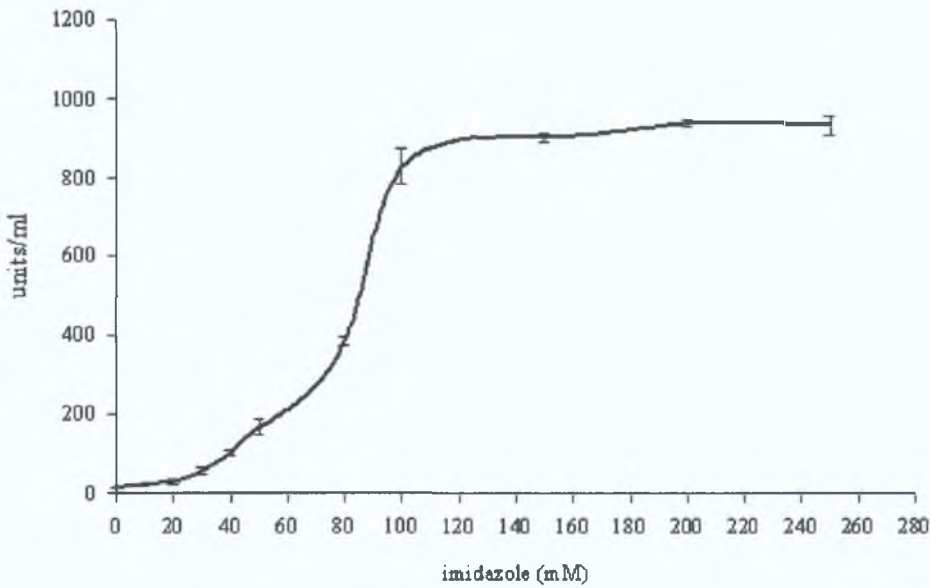


Figure 3.32 Imidazole elution profile

$rHsaPAP1_{6H}$ activity of samples eluted from Ni-NTA resin by 50 mM potassium phosphate, pH 8.0 containing increasing concentrations of imidazole.

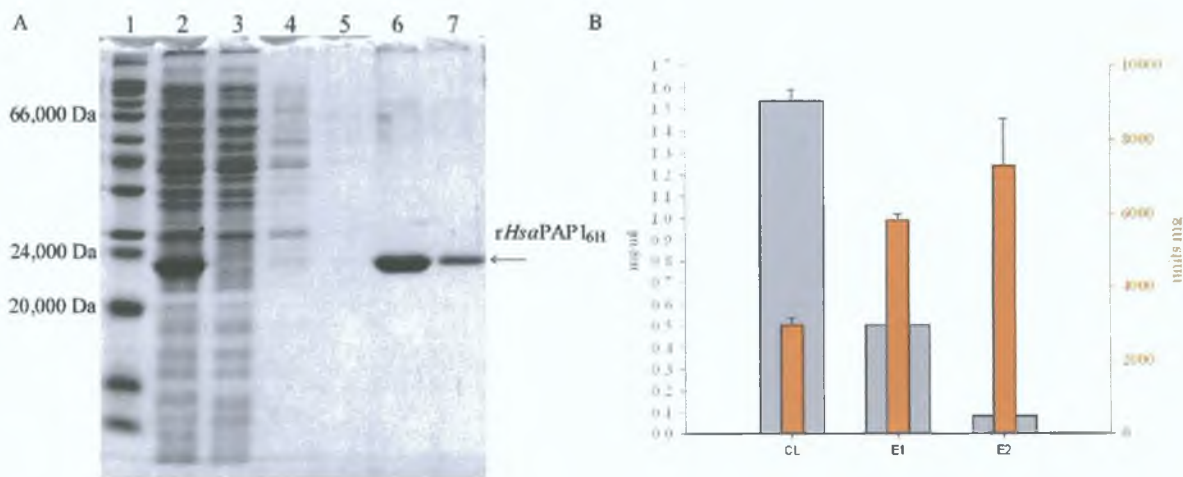


Figure 3.33 Purification of $rHsaPAP1_{6H}$

(A): Analysis by 15% SDS-PAGE (Section 2.16), of $rHsaPAP1_{6H}$ purification using IMAC (Section 2.13.1). Lane 1, Mr (sizes given in Figure 2.6); Lane 2, cleared lysate; Lane 3, flow-through; Lane 4, first wash; Lane 5, second wash; Lane 6, first elution; Lane 7, second elution. (B): Protein concentration (mg/ml) and specific activity (units/mg) of cleared lysate (CL), first elution (E1) and second elution (E2).

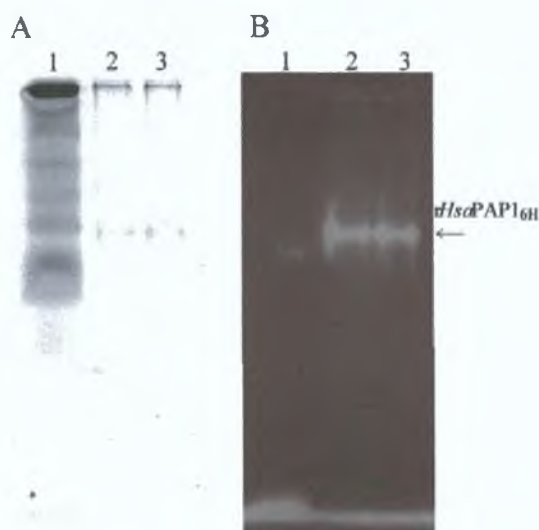


Figure 3.34 Zymogram analysis of purified *rHsaPAP1_{6H}*

Native PAGE (A) and subsequent zymogram analysis (B, Section 2.17) of purified *rHsaPAP1_{6H}*. Lane 1, M_r (Figure 2.6); Lanes 2 & 3, purified *rHsaPAP1_{6H}* (20 μ l of 0.195 mg/ml).

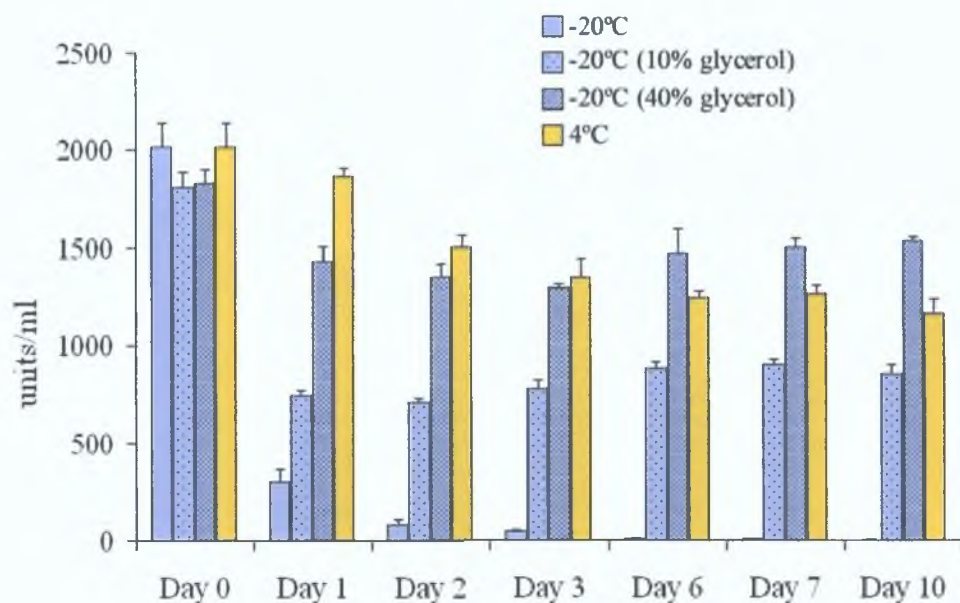


Figure 3.35 Storage profile of *rHsaPAP1_{6H}*

Profile of *rHsaPAP1_{6H}* activity (units/ml) in samples stored for up to 10 days at both 4°C and -20°C. Samples stored at -20°C with and without 10% and 40% glycerol.

3.8 Discussion

The human PAP1 gene was successfully cloned from human mRNA (plasmid pRV1, Section 3.2). Although many mammalian proteins cannot be expressed in an active form in *E. coli*, our hope was that, since human PAP1 shared many biochemical characteristics and sequence homology with its prokaryotic counterparts (Chapter 1), *E. coli* could be used successfully to express active recombinant human PAP1. As outlined in Section 3.5, *E. coli* could indeed be used as host, expressing catalytically active

recombinant human PAP1. Significant quantities of enzyme were produced as shown in Figure 3.13 and using zymography (Figure 3.14) PAP1 activity was correlated to this protein band.

To achieve this expression in *E. coli*, the recombinant human PAP1 gene (*rHsa-pap1*) was cloned within a suitable expression vector. The *rHsa-pap1* sequence was sub-cloned into the pQE-60 vector to yield the construct pRV8 (Section 3.4.1), and into the pPC225 vector to yield the constructs pRV5 (Section 3.4.2) and pRV5_Δ6H (Section 3.4.3). pQE-60 features the phage T5 transcriptional promoter and a *lac* operator sequence while pPC225 features *Ptac*, a hybrid of the *trp* and *lac* promoters. Both are strong promoter systems that can be repressed by the LacI protein.

Both of these expression vectors provide ribosomal binding sites (RBS). As a result of the cloning strategy used to create pRV5 (Figure 3.8), there are two RBS downstream of *Ptac* in this construct (Figure 3.10). One of the two RBS is separated from the *rHsa-pap1* start codon by 9 nucleotides, while the other is about 20 nucleotides further upstream. It has been reported that the optimal distance between a RBS and the start codon is seven plus or minus two nucleotides (Kozak, 1983). At least one of these RBS is functionally relevant in pRV5.

Using *E. coli* XL10-Gold as expression host, the pRV5 and pRV8 expression constructs produced similar quantities of rHsaPAP1_{6H} (Figure 3.13B). pRV5 was chosen for optimisation of expression conditions and to establish an efficient host-vector combination, to determine if some genotypes are more suitable for expression of this particular recombinant protein. Several *E. coli* strains were transformed with pRV5 to express rHsaPAP1_{6H} (Section 3.6).

In addition to investigating the merit of different *E. coli* host strains (Table 2.1), the effect codon bias may have on recombinant human PAP1 expression was considered. As described in Section 3.6, *E. coli* Rosetta and RosettaBlue (derivatives of BL21 and NovaBlue, respectively) carry the pRARE plasmid (Table 2.3, Figure 2.4). The pRARE system represents a possible strategy to overcome *E. coli* codon bias, by raising the concentration of the tRNAs that complement rare codons (Novy *et al.*, 2001). The data presented in Figure 3.19 would suggest that the pRARE strategy indeed improves rHsaPAP1_{6H} expression. A distinctly greater yield of rHsaPAP1_{6H} was obtained in *E. coli* Rosetta than in BL21. However, this effect is not so evident from the data presented in Figure 3.20, where rHsaPAP1_{6H} yield is only slightly more efficient in *E. coli* RosettaBlue than Nova Blue. By comparison, expression in *E. coli* DH5α and INVαF'

(Figure 3.18), not harbouring the pRARE plasmid, was relatively good. Furthermore, the data presented in Figure 3.21 shows an even greater expression level was achieved by *E. coli* XL10-Gold, a strain also lacking any additional tRNA genes. It should be noted that the transcriptional repression in XL10-Gold was more effective than in Nova Blue or RosettaBlue.

Although the supplementation of limiting tRNAs can improve heterologous recombinant protein expression to some extent, it is only one aspect of choosing a suitable expression host. It would appear that the gain achieved by using a system such as pRARE can be made somewhat obsolete by using a more efficient *E. coli* expression strain. It was decided to employ the XL10-Gold strain for all subsequent rHsaPAP1_{6H} expression. The production of functionally active enzyme was very satisfactory and justified the choice of using *E. coli* as expression host for this eukaryotic enzyme.

An alternative strategy to overcome the codon bias of *E. coli* would be to use site-specific mutagenesis to replace rare codons with *E. coli* optimally-used counterparts (Baneyx, 1999). This approach was undertaken, as is reported in Section 3.6.2. However, this resulted in a decrease in protein yield, rather than an anticipated increase. Replacing of the first rare codon Arg5[AGG] with [CGT], together with changing Lys6[AAA] to the more optimally-used codon [AAA], resulted in reduced yield of rHsaPAP1_{6H} protein, as seen in Figure 3.29. Focusing only on the Arg5 codon (Figure 3.30), lower levels of rHsaPAP1_{6H} were obtained in the case of both optimal *E. coli* codons [CGT] and [CGC]. RT-PCR analysis (Figure 3.31) showed that the level of transcription obtained for pRV5_R5^Sa and pRV5_R5^Sb, were comparable to wild type (pRV5), indicating that the reduced yield in protein for these two mutants is not a result of reduced message. One possibility is that making the above nucleotide substitutions has resulted in a change in mRNA structure. Such an alteration in the mRNA secondary structure, although not obviously affecting RNA stability, may be affecting the efficiency of translation. A software package which calculates possible secondary structures of single stranded nucleic acids (Mfold, Section 2.11) was used to predict how the mRNA coding for rHsaPAP1_{6H} might fold (Figure 3.36). For each mRNA sequence, up to 35 possible structures were calculated. Figure 3.36 shows those structures which had the lowest calculated free energy (ΔG), therefore being the most likely structures. These three structures varied from one another only in the region of the translational start. It is possible that ribosome binding may be affected by this structural variation, causing retarded translation.

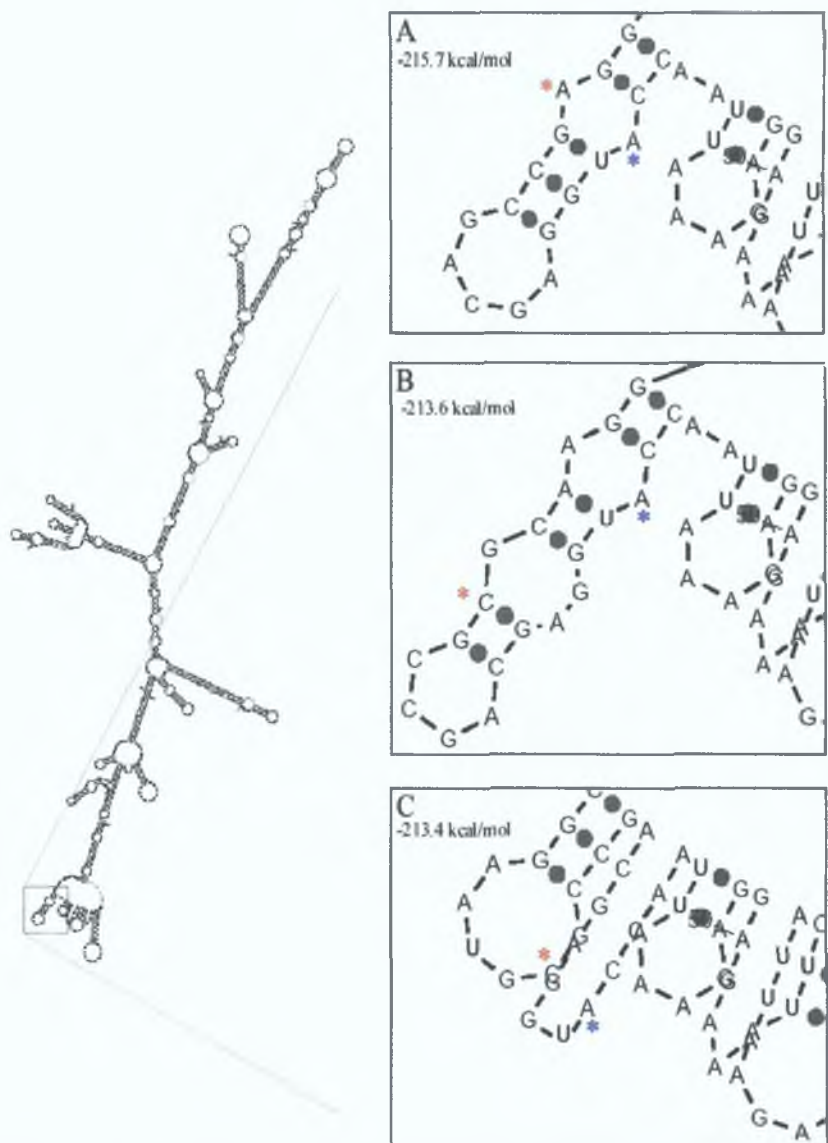


Figure 3.36 Potential structure of mRNA from pRV5, pRV5_R5^Sa and pRV5_R5^Sb
Possible secondary structures of mRNA, predicted using Mfold (Section 2.11), coding for *rHsaPAP1*_{6H}, transcribed from pRV5 (Figure 3.10). Substitution of Arg5 codon on pRV5_R5^Sa and pRV5_R5^Sb (Table 2.3) causes change in structure at translation start site (insets). (A): pRV5 (Arg5 AGG). (B): pRV5_R5^Sa (Arg5 AGG to CGC). (C): pRV5_R5^Sb (Arg5 AGG to CGU). Calculated free energy (ΔG) is given for each structure. Translation start indicated by *. Arg5 codon indicated by *. The predicted structures contained a small number of mismatched U-C and U-G base pairs. The validity of such pairing has been supported by a study of water-mediated U-C pairings in RNA structures (Brandl *et al.*, 1998).

While the pRV5_Δ6H construct was created for the expression of native recombinant human enzyme (*rHsaPAP1*), the pRV5 and pRV8 constructs were designed to express a His-tagged enzyme (*rHsaPAP1*_{6H}), having a six histidine peptide (6xHis) fused to the C-terminal of the protein to aid in purification. A short Gly-Ser codon linker sequence (Figure 3.10) exists between *rHsa-pap1* and 6xHis, a result of the *Bam*H1/*Bgl*II fusion (Figure 3.5). Since there was a possibility that such a peptide fusion would affect

enzyme structure and/or activity (Ledent *et al.*, 1997), the C-terminal location was chosen, it being the terminal of human PAP1 which exhibits the least homology with its prokaryotic counterparts (Figure 5.6). The hope was that this end of the protein is less likely to involve critical regions of the enzyme. As this C-terminal tag did not interfere with *rHsaPAP1_{6H}* activity (Section 3.5), one-step purification by metal affinity chromatography was possible, as reported in Section 3.7. The application of a *Strep*-tag for affinity purification (Skerra and Schmidt, 1999), as an alternative to the His-tag, was considered but not carried out in this work.

The analysis of *rHsaPAP1_{6H}* expression in *E. coli* XL10-Gold, shown in Figures 3.22 and 3.23, shows clearly the production of recombinant enzyme. Upon induction with 50 μ M IPTG, the level of *rHsaPAP1_{6H}* rises rapidly until it constitutes 15-18% of total protein. The increase continues for 4 hrs at which time it levels off, approximately matching the progression into stationary phase for the culture. Interestingly, and somewhat unexpected in the context of expressing recombinant proteins, the presence of the human PAP1 gene results in a higher cell density (Figure 3.22). The effectiveness of transcriptional repression shown in Figure 3.21 can again be seen in Figure 3.24. This repression is attributed to the *lacI^q* allele, part of the XL10-Gold genotype, resulting in high levels of LacI repressor protein (Section 3.6). A study to determine the optimal concentration of IPTG used for induction (Figure 3.25) shows that this is a very important parameter to consider. Very low levels of recombinant protein were obtained when using 0.5 or 1 mM IPTG. 50 μ M IPTG was deemed the most suitable concentration. Similar expression was obtained when inducing at various times during early exponential phase (A_{600} of 0.3 to 0.5) as shown in Figure 3.26. The standard induction parameters used in this work were 50 μ M IPTG at A_{600} of 0.3.

The effect of sonication was dramatically demonstrated by the analysis shown in Figure 3.27. Sonication of longer than 30 sec resulted in a sharp reduction of *rHsaPAP1_{6H}* activity. This reduction in activity correlates with a reduction in the amount of *rHsaPAP1_{6H}* present. It is assumed that prolonged sonication denatures the protein sufficiently to be susceptible to rapid proteolysis. The standard sonication time used in this work was 30 sec.

The imidazole elution profile (Figure 3.32) was used to establish suitable concentrations of imidazole for the washing and elution steps during IMAC purification of *rHsaPAP1_{6H}* (Section 2.13.1). Since it was desirable to maximise the yield of pure

enzyme, a concentration of 20 mM imidazole was used for washing. 200 mM imidazole was used for elution, ensuring greatest recovery from the Ni-NTA resin.

SDS-PAGE analysis of the standard IMAC purification procedure employed in this work (Figure 3.33A) shows that exceptionally pure *rHsaPAP1_{6H}* was obtained. The enzyme had in excess of 90% purity, demonstrating the benefit of this one-stage method of protein purification. The purified protein was confirmed as *rHsaPAP1_{6H}* by zymography (Figure 3.34) and western blotting (Figure 5.19). Even with added wash steps and/or increased imidazole concentrations, very small amounts of contaminating proteins of about 45,000 and 66,000 Da were routinely observed (see also Figure 5.15). The increase in specific activity correlating to the purification can be seen in Figure 3.33B, *rHsaPAP1_{6H}* having a specific activity of around 7000 units/mg. By the standard expression (Section 2.12.1) and purification (Section 2.13.1) procedures, *rHsaPAP1_{6H}* concentrations of around 500 µg/ml were routinely achieved in the first 5 ml elution. Therefore, by scaling the expression culture up to 1 L, yields of up to 25 mg pure enzyme could be achieved.

Suitable storage conditions for purified *rHsaPAP1_{6H}* were investigated as shown in Figure 3.35. Storage at -20°C without glycerol resulted in almost complete loss of activity, correlating to the substantial precipitation observed. Inclusion of glycerol at this temperature protected the protein from precipitating. When using 10% glycerol the activity initially dropped to half its original value but did not decrease further during the ten-day trial. Using 40% glycerol proved even more preservative, conserving almost 80% of activity. Even after two months of storage, two thirds of activity remained. In contrast, storage (without glycerol) at 4°C resulted in a less immediate but more constant drop in activity, falling below the value retained by -20°C/40% glycerol storage, by the end of the ten-day study. When samples of pure enzyme were to be used for various procedures within a few days, storage at 4°C was usually satisfactory. Some stock sample was habitually stored at -20°C with 40% glycerol.

4.0 Characteristics of recombinant human PAP1

The successful cloning and expression of catalytically active, recombinant human PAP1 (*rHsaPAP1*) and the purification of the His-tagged enzyme (*rHsaPAP1*_{6H}), is reported in Chapter 3. Subsequently, a range of studies to determine the biochemical (Section 4.1) and kinetic (Section 4.2) characteristics of *rHsaPAP1*_{6H} were carried out. A preliminary study was carried out to establish conditions under which *rHsaPAP1*_{6H} would form protein crystals (Section 4.3).

4.1 Biochemical properties of recombinant human PAP1

4.1.1 Relative molecular mass

As outlined in Section 2.19.1, 50 µl of purified *rHsaPAP1*_{6H}, at a concentration of 250 µg/ml, was analysed by size exclusion chromatography. Figure 4.1 shows the protein concentration profile corresponding to the elution of *rHsaPAP1*_{6H} and the protein standards BSA, Carbonic Anhydrase and Lysozyme. Also shown is the profile of PAP1 activity corresponding to the elution of *rHsaPAP1*_{6H}. The inset in Figure 4.1 shows a plot of elution volume over void volume (V_e/V_o) versus the log of molecular weight. From this plot the relative molecular weight of *rHsaPAP1*_{6H} was determined as 24.5 kDa.

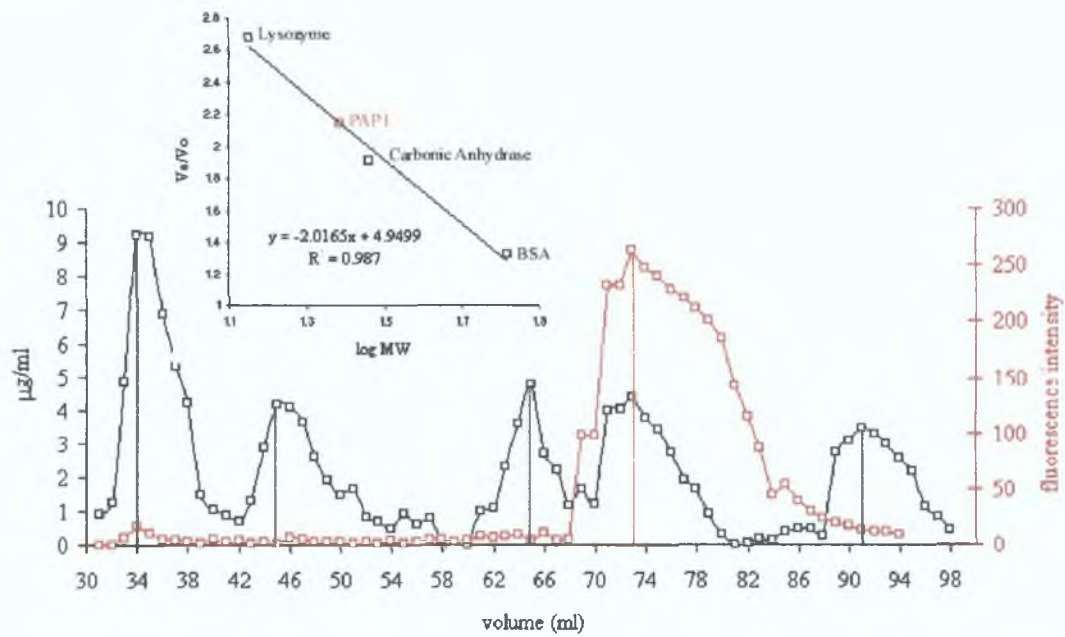


Figure 4.1 Determination of *rHsaPAP1*_{6H} native size

Size exclusion chromatography (Section 2.19.1) of *rHsaPAP1*_{6H}. Void volume of 34 ml was ascertained using blue dextran. Elution of *rHsaPAP1*_{6H} at 73 ml was determined using PAP1 activity assay (Section 2.15.3) and is represented by fluorescent intensity, shown in red. Elution volumes of molecular weight standards BSA (45 ml), Carbonic Anhydrase (65 ml) and Lysozyme (91 ml), were determined by protein assay (Section 2.14.3) and used to construct a plot of V_e/V_o versus log MW (inset). This plot was used to establish the relative molecular weight of *rHsaPAP1*_{6H} to be 24.5 kDa.

4.1.2 Isoelectric point

The isoelectric points of *rHsaPAP1* and *rHsaPAP1_{6H}* were predicted as 5.54 and 5.97 respectively, using online tools (Section 2.11). To determine the isoelectric point of *rHsaPAP1_{6H}* experimentally, the isoelectric focusing (IEF) procedure described in Section 2.19.6 was carried out on extensively dialysed, purified *rHsaPAP1_{6H}*. Figure 4.2A shows the IEF gel stained by the coomassie blue method (Section 2.16.4). By this method of staining an extremely faint band was detectable corresponding to the 6.9 pI marker. The gel was additionally treated by the silver staining method (Section 2.16.4), resulting in Figure 4.2B. By now the faint band, attributed to *rHsaPAP1_{6H}* had been significantly over-stained. As an estimation, the pI of *rHsaPAP1_{6H}* was taken to be 6.9.

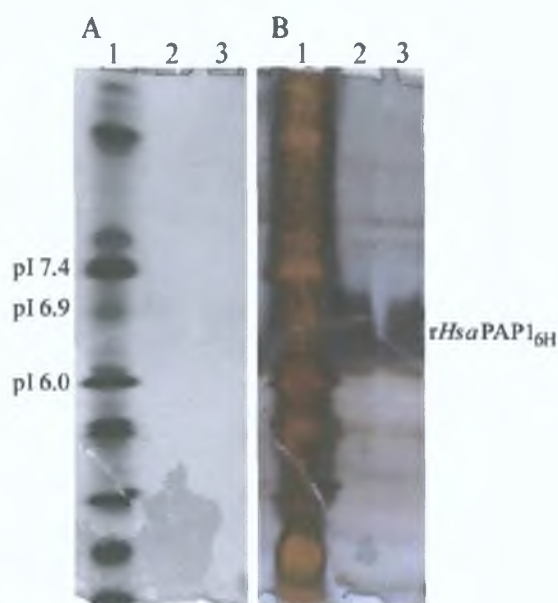


Figure 4.2 Iso-electric point of *rHsaPAP1_{6H}*

Isoelectric focusing (IEF) of purified *rHsaPAP1_{6H}* as described in Section 2.19.6. Gel visualised after both coomassie (A) and silver staining (B) as described in Section 2.16.4. Lane 1, IEF markers (sizes as in Figure 2.8); Lane 2, 5 µl purified *rHsaPAP1_{6H}* (150 µg/ml); Lane 3, 2.5 µl purified *rHsaPAP1_{6H}* (150 µg/ml).

4.1.3 Influence of pH

The influence of pH on the activity of *rHsaPAP1* and *rHsaPAP1_{6H}* was investigated as outlined in Section 2.19.2. Figure 4.3 shows the pH profile for purified *rHsaPAP1_{6H}* which is compared to that of both *rHsaPAP1* and *rHsaPAP1_{6H}* in cleared lysate. The three profiles are consistent across the range except at pH 8.5, where purified *rHsaPAP1_{6H}* shows optimum activity.

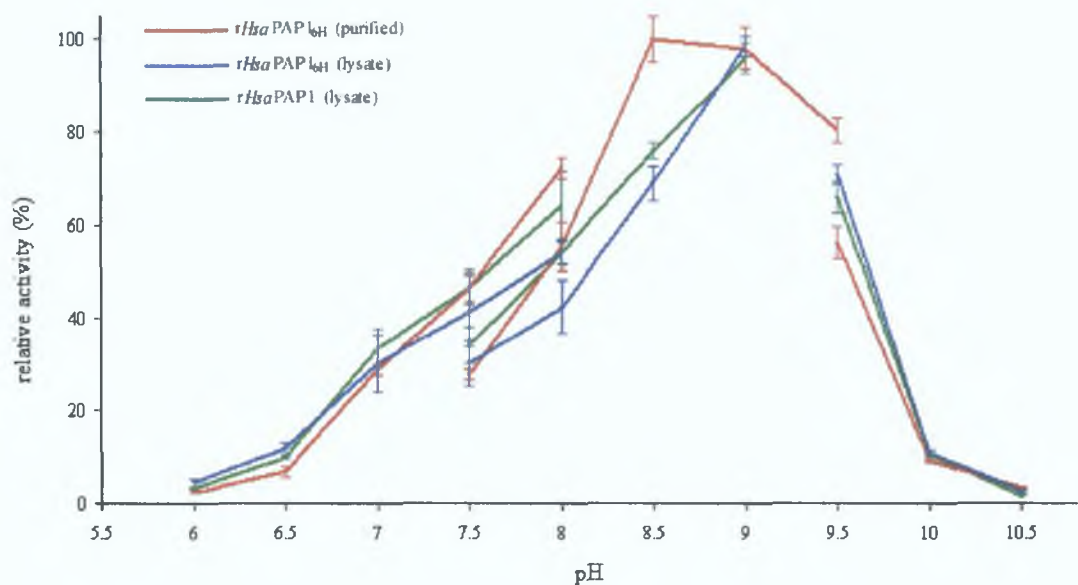


Figure 4.3 Influence of pH on *rHsaPAP1_{6H}* and *rHsaPAP1_{6H}* activity
 Relative enzyme activity over pH range 6.0 to 10.5. Different buffers used for specific ranges as described in Section 2.19.2. Both *rHsaPAP1* and *rHsaPAP1_{6H}* in cleared lysate (Section 3.5) are compared to purified *rHsaPAP1_{6H}* (Section 3.7.1).

4.1.4 Influence of temperature

The influence of temperature on the activity of *rHsaPAP1_{6H}* was investigated. As outlined in Section 2.15.2, the standard activity assay was carried out at a range of temperatures. The resulting 15 min temperature profile is shown in Figure 4.4. Maximal activity was observed at 50°C.

The temperature stability of *rHsaPAP1_{6H}* is illustrated in Figure 4.5, showing the effect that pre-incubation of *rHsaPAP1_{6H}* has on its activity, which was subsequently determined using the standard assay (Section 2.15.2) at 37°C.

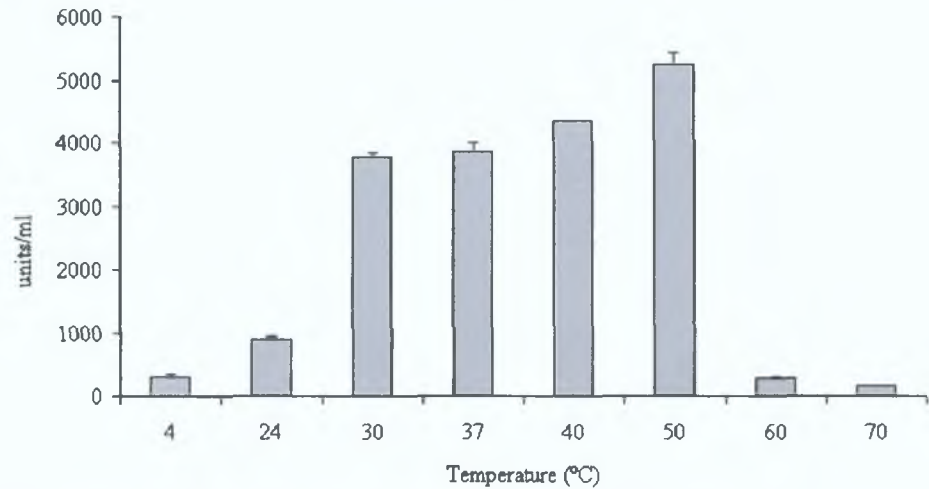


Figure 4.4 Influence of temperature on *rHsaPAP1_{6H}* activity
 The effect of temperature on the activity of *rHsaPAP1_{6H}* activity was determined by carrying out the standard assay (Section 2.15.2) at 4 to 70°C for 15 min.

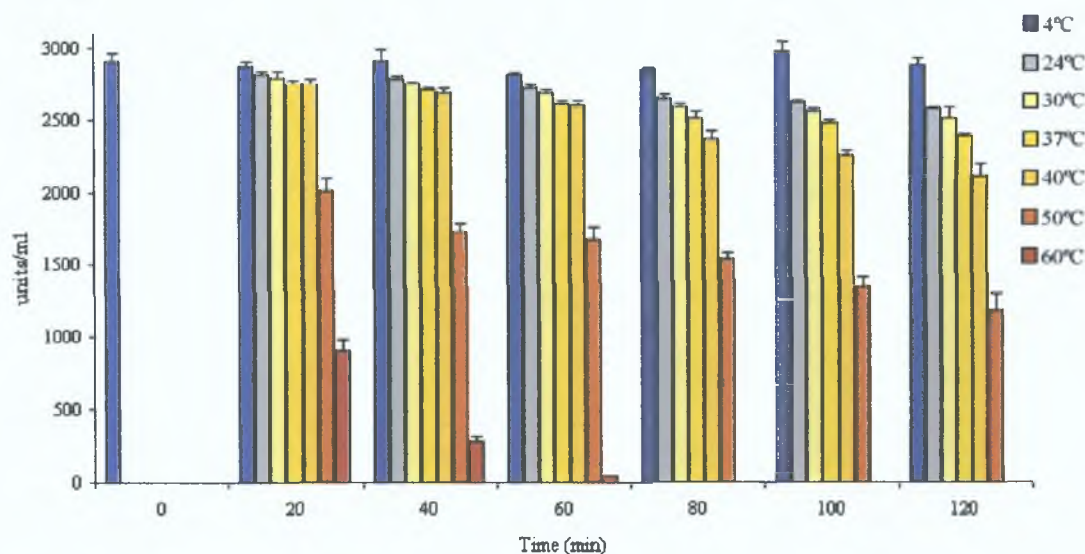


Figure 4.5 Temperature stability of *rHsaPAP1_{6H}*

The effect of pre-incubating *rHsaPAP1_{6H}* at a range of temperatures (4 to 60°C). Activity determined by standard assay at 37°C (Section 2.15.2).

4.1.5 Influence of DTT and EDTA

DTT and EDTA are essential components of the standard PAP1 activity assay (Section 2.15.2). In order to optimise the assay concentrations, their effect on *rHsaPAP1_{6H}* activity was determined. Figures 4.6 and 4.7 show the influence of DTT and EDTA, respectively, on *rHsaPAP1_{6H}* activity. The optimal DTT concentration for activity was found to be 10 mM, while EDTA had only a marginal positive effect on activity as the concentration was increased to 20mM.

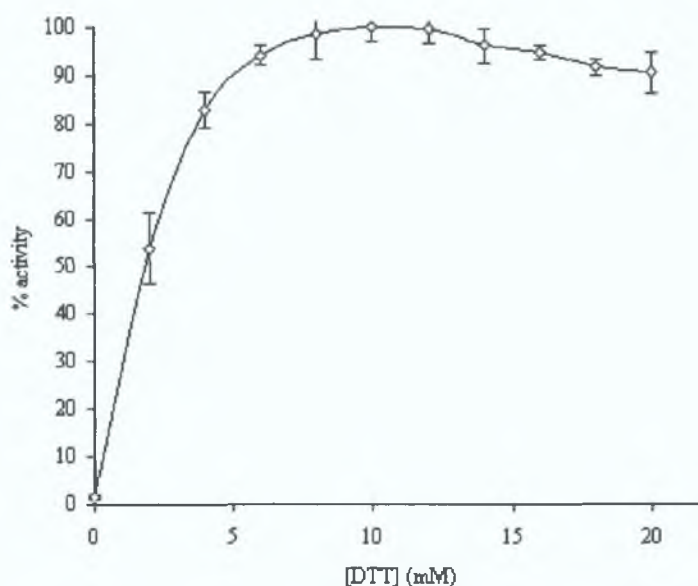


Figure 4.6 Influence of DTT on *rHsaPAP1_{6H}* activity

The effect of DTT concentration on the activity of *rHsaPAP1_{6H}* activity was determined by carrying out the standard assay (Section 2.15.2) using a range of DTT concentrations up to 20 mM.

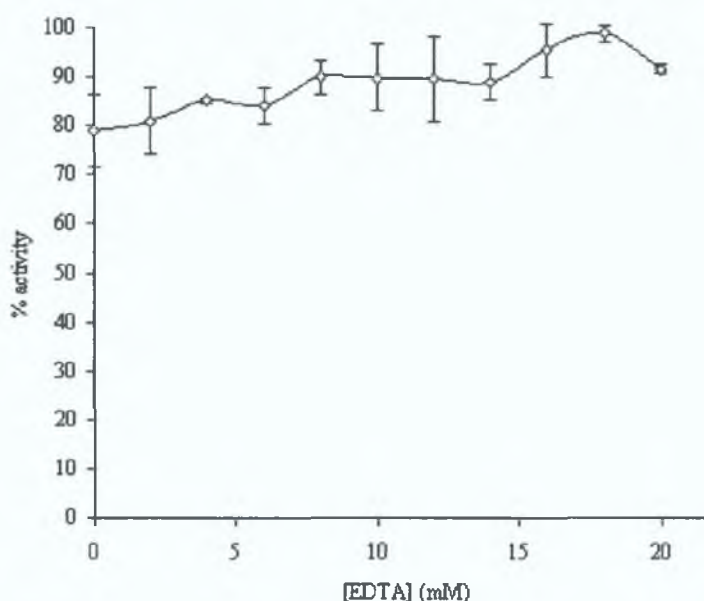


Figure 4.7 Influence of EDTA on *rHsaPAP1_{6H}* activity

The effect of EDTA concentration on the activity of *rHsaPAP1_{6H}* activity was determined by carrying out the standard assay (Section 2.15.2) using a range of EDTA concentrations up to 20 mM.

4.1.6 Inhibition

A selection of compounds were investigated for their possible inhibitory effect on the pGlu-AMC degrading activity of *rHsaPAP1_{6H}* (Section 2.19.5).

The thiol-blocking compound iodoacetate caused complete inhibition of activity at a concentration of 100 μ M (data not shown), consistent with the literature as described in Section 1.3.7.

The inhibition caused by free *L*-pGlu and the substrate analogue 2-pyrrolidone (Section 1.3.8) is shown in Figures 4.8 and 4.9. IC_{50} values of 2.75 and 1.0 mM, respectively, were determined. IC_{50} being the inhibitor concentration resulting in a 50% loss in enzymatic activity.

The broad spectrum serine protease inhibitor diisopropyl fluorophosphate (DFP) had no discernable inhibitory effect on activity (Figure 4.10). Phenylmethylsulphonyl fluoride (PMSF), which inhibits serine and some cysteine proteases, had an IC_{50} of 4.1 mM (Figure 4.11).

The compound 1,10-phenanthroline was inhibitory with an IC_{50} of 2.6 mM (Figure 4.12) while 4,7-phenanthroline was found to exert no inhibitory effect (Figure 4.13).

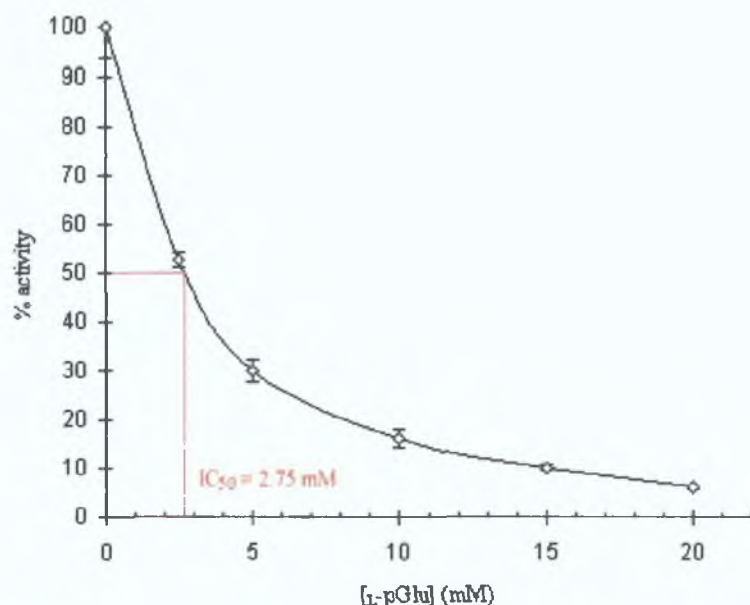


Figure 4.8 Effect of L-pGlu on rHsaPAP1_{6H} activity

The effect of L-pGlu on the activity of rHsaPAP1_{6H} was determined by the procedure outlined in Section 2.19.5, using L-pGlu at concentration range up to 20 mM.

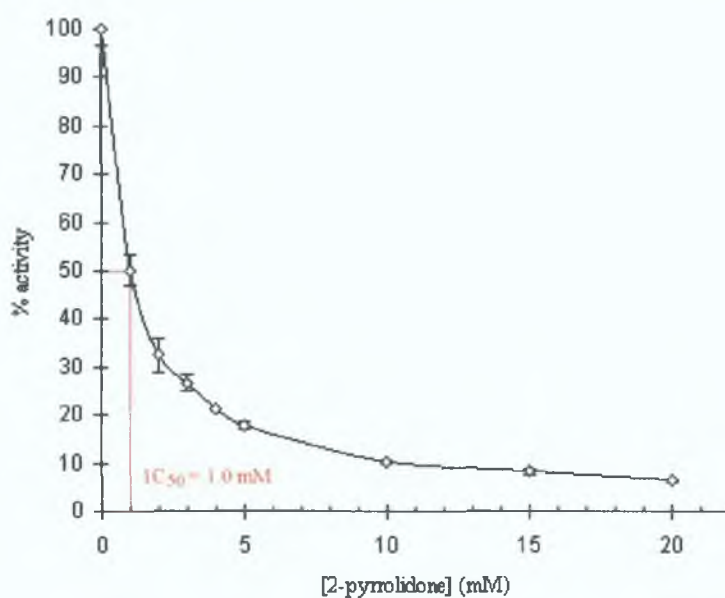


Figure 4.9 Effect of 2-pyrrolidone on rHsaPAP1_{6H} activity

The effect of 2-pyrrolidone on the activity of rHsaPAP1_{6H} was determined by the procedure outlined in Section 2.19.5, using 2-pyrrolidone at concentration range up to 20 mM.

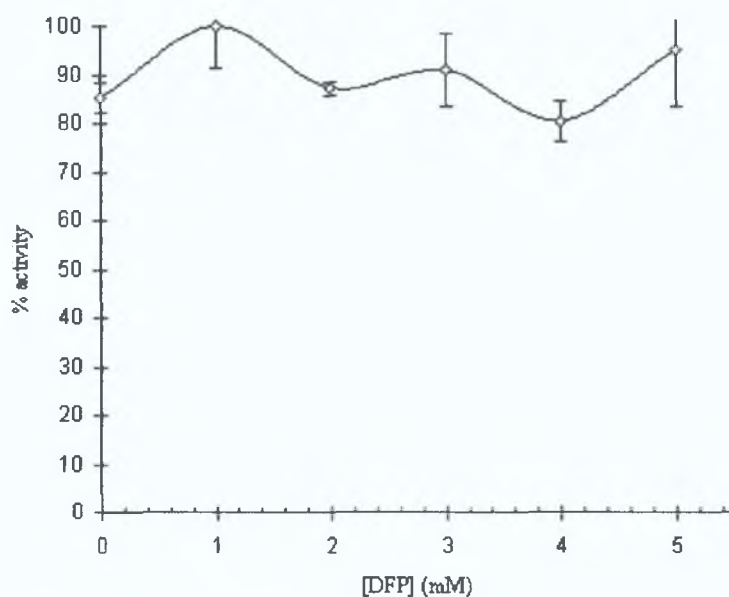


Figure 4.10 Effect of DFP on *rHsaPAP1_{6H}* activity

The effect of DFP on the activity of *rHsaPAP1_{6H}* was determined by the procedure outlined in Section 2.19.5, using DFP at concentration range up to 5 mM.

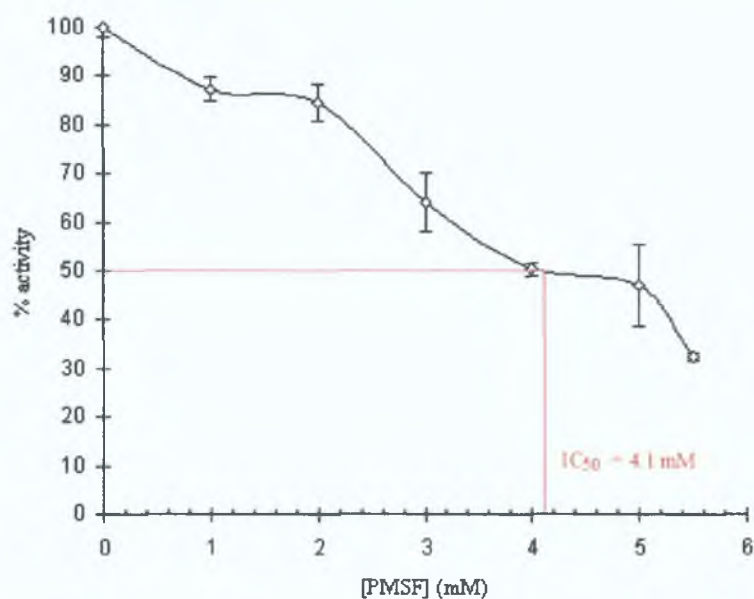


Figure 4.11 Effect of PMSF on *rHsaPAP1_{6H}* activity

The effect of PMSF on the activity of *rHsaPAP1_{6H}* was determined by the procedure outlined in Section 2.19.5, using PMSF at concentration range up to 5.5 mM.

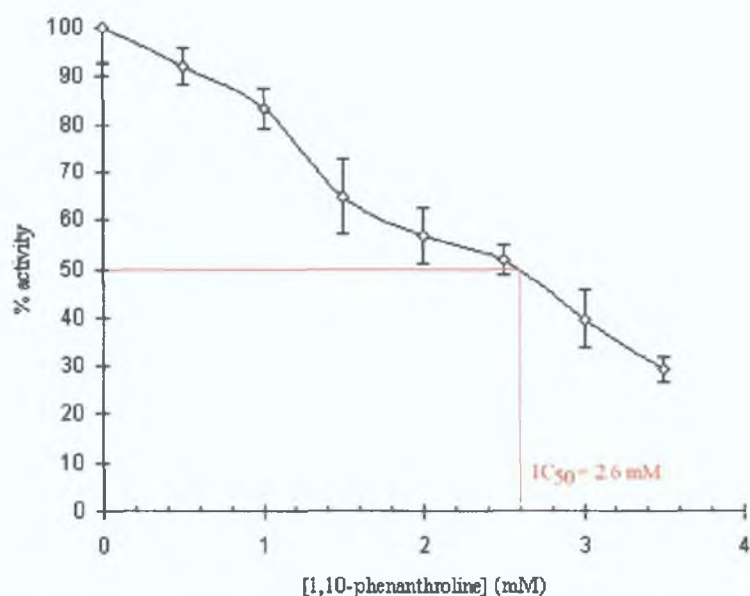


Figure 4.12 Effect of 1,10-phenanthroline on *rHsaPAP1_{6H}* activity

The effect of 1,10-phenanthroline on the activity of *rHsaPAP1_{6H}* was determined by the procedure outlined in Section 2.19.5, using 1,10-phenanthroline at concentration range up to 3.5 mM.

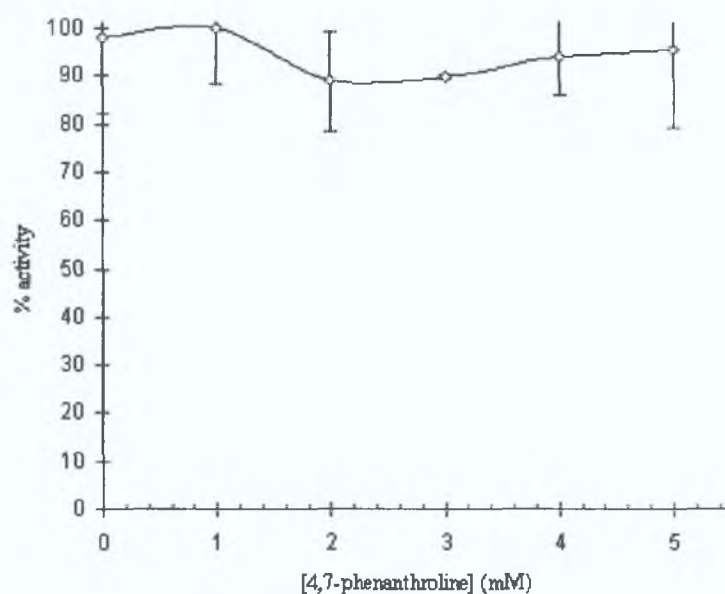


Figure 4.13 Effect of 4,7-phenanthroline on *rHsaPAP1_{6H}* activity

The effect of 4,7-phenanthroline on the activity of *rHsaPAP1_{6H}* was determined by the procedure outlined in Section 2.19.5, using 4,7-phenanthroline at concentration range up to 5 mM.

4.2 Kinetic parameters of recombinant human PAP1

Purified *rHsaPAP1_{6H}* needed to be diluted to a concentration of 1-5 $\mu\text{g/ml}$ to ensure the fluorescent intensity values fell within the measurable range of the Perkin-Elmer LS-50 fluorescence spectrophotometer used in this work (Section 2.15.2).

The procedure to assay PAP1 activity at various pGlu-AMC concentrations, as outlined in Section 2.19.3, was carried out with purified *rHsaPAP1_{6H}* (Section 3.7.1), as well as

with *rHsaPAP1* and *rHsaPAP1*_{6H} in cleared lysate (Section 3.5). The resulting Michaelis-Menten curve for purified *rHsaPAP1*_{6H} is shown in Figure 4.14 and the curves for *rHsaPAP1*_{6H} and *rHsaPAP1* in cleared lysate are shown in Figure 4.15.

The data sets from Figures 4.14 and 4.15 were applied to the three kinetic models described in Appendix C, namely: Lineweaver-Burk (Figures 4.16 and 4.17), Eadie-Hofstee (Figures 4.18 and 4.19) and Hanes-Woolf (Figures 4.20 and 4.21). From these plots the maximal enzyme velocity (V_{max}) and the Michaelis constant (K_m) were determined, which are given in Table 4.1. Also given in Table 4.1 is the turnover number (k_{cat}) for purified *rHsaPAP1*_{6H}, which was determined as described in Appendix C.

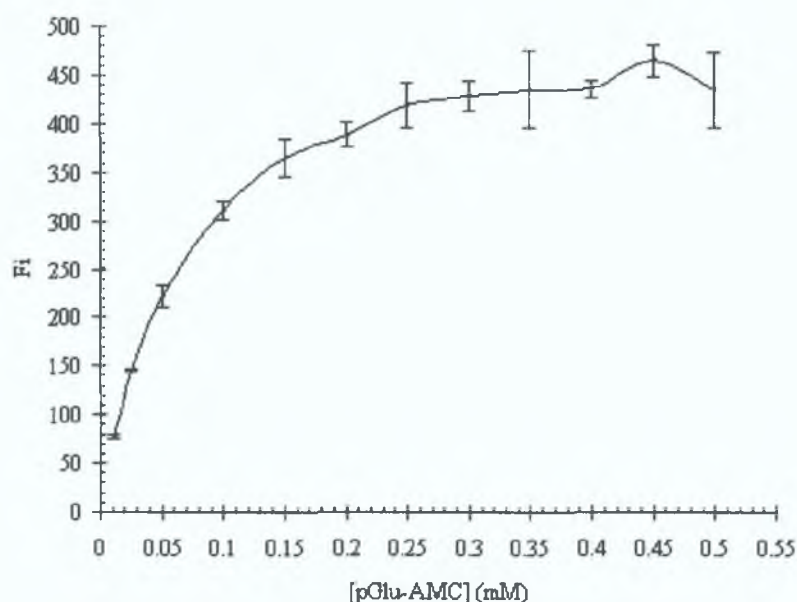


Figure 4.14 Michaelis-Menten curve for purified *rHsaPAP1*_{6H}

Plot of substrate concentration ([pGlu-AMC]) versus reaction rate, represented by fluorescence intensity (Fi), for purified *rHsaPAP1*_{6H} diluted to 2.5 $\mu\text{g/ml}$. Data obtained by the procedure described in Section 2.19.3.

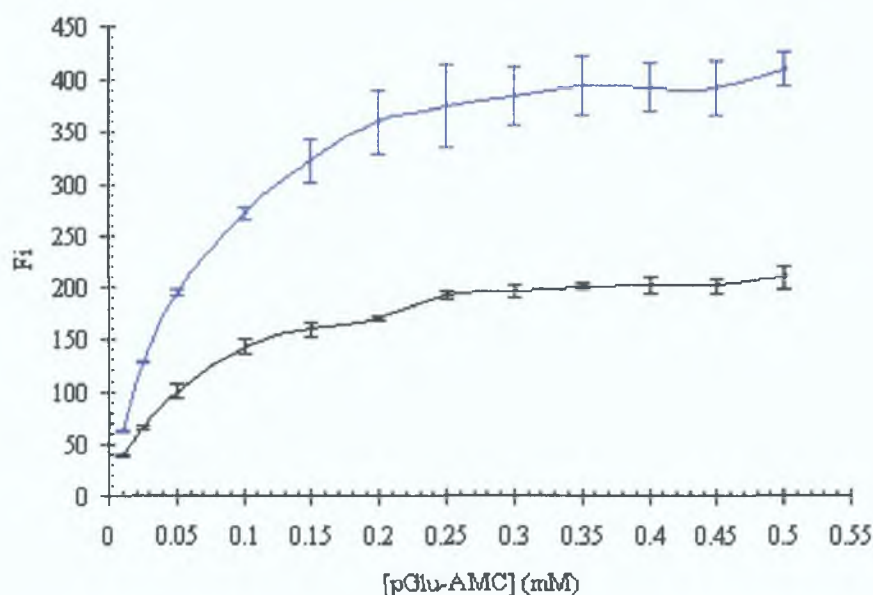


Figure 4.15 Michaelis-Menten curve for *rHsaPAP1_{6H}* and *rHsaPAP1* in lysate
 Plot of substrate concentration ([pGlu-AMC]) versus reaction rate, represented by fluorescence intensity (Fi), for both *rHsaPAP1* (blue) and *rHsaPAP1_{6H}* (black) in lysate. Data obtained by the procedure described in Section 2.19.3.

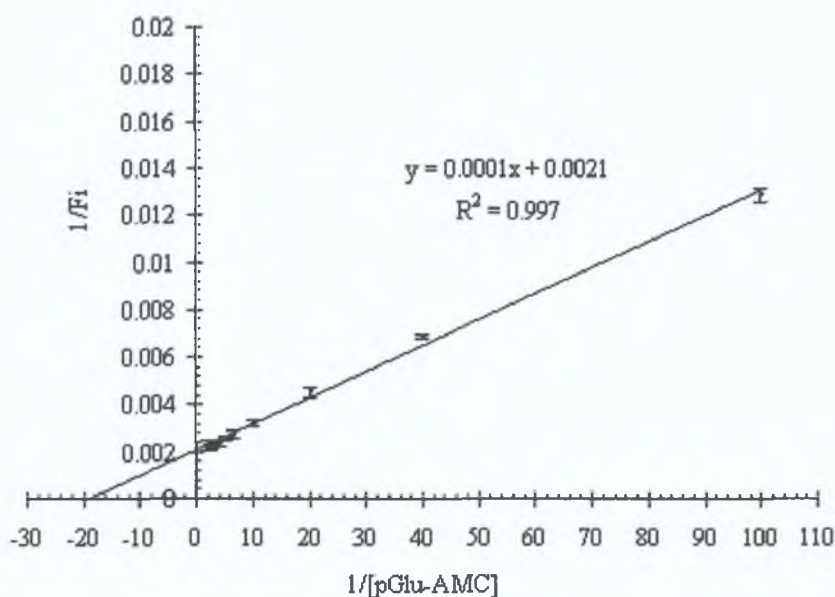


Figure 4.16 Lineweaver-Burk plot for purified *rHsaPAP1_{6H}*
 Data from Figure 4.14 fitted to the Lineweaver-Burk model for the determination of kinetic parameters, as outlined in Appendix C. Equation and regression are shown. K_m and V_{max} values given in Table 4.1.

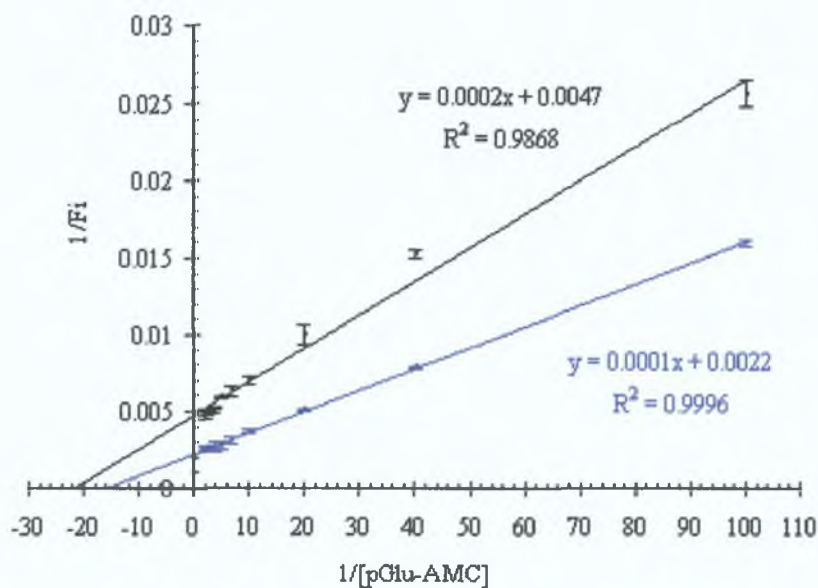


Figure 4.17 Lineweaver-Burk plot for *rHsaPAP1_{6H}* and *rHsaPAP1* in lysate
 Data from Figure 4.15 (same colour association) fitted to the Lineweaver-Burk model for the determination of kinetic parameters, as outlined in Appendix C. Equation and regression are shown. K_m and V_{max} values given in Table 4.1.

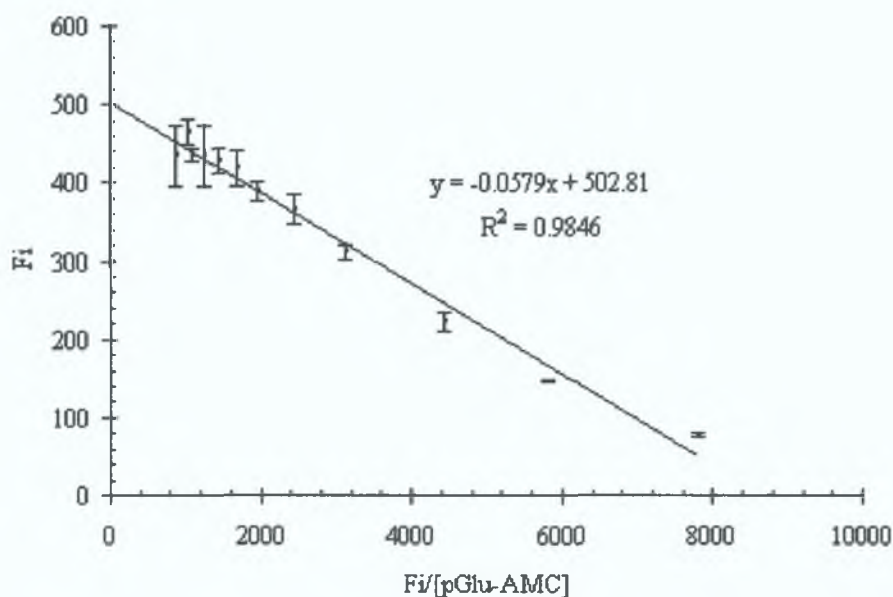


Figure 4.18 Eadie-Hofstee plot for purified *rHsaPAP1_{6H}*
 Data from Figure 4.14 fitted to the Eadie-Hofstee model for the determination of kinetic parameters, as outlined in Appendix C. Equation and regression are shown. K_m and V_{max} values given in Table 4.1.

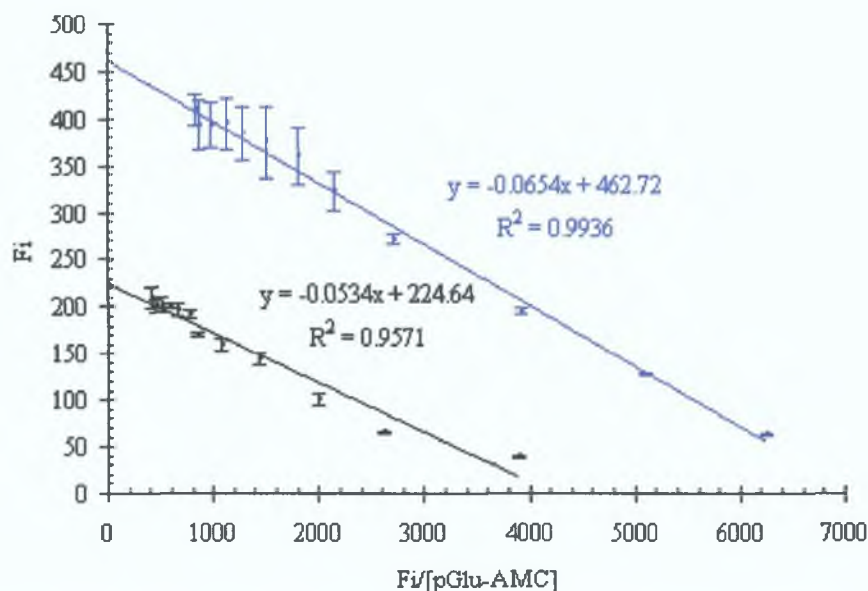


Figure 4.19 Eadie-Hofstee plot for *rHsaPAP1_{6H}* and *rHsaPAP1* in lysate
 Data from Figure 4.15 (same colour association) fitted to the Eadie-Hofstee model for the determination of kinetic parameters, as outlined in Appendix C. Equation and regression are shown. K_m and V_{max} values given in Table 4.1.

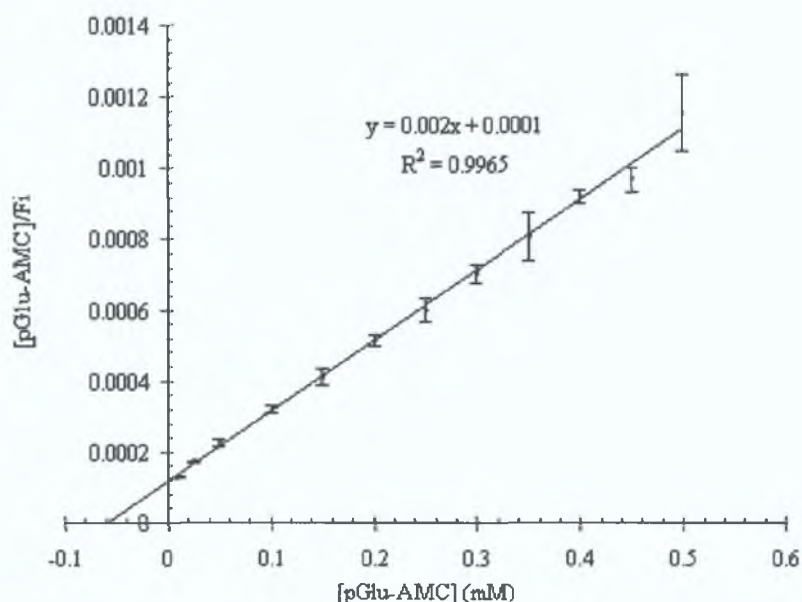


Figure 4.20 Hanes-Woolf plot for purified *rHsaPAP1_{6H}*
 Data from Figure 4.14 fitted to the Hanes-Woolf model for the determination of kinetic parameters, as outlined in Appendix C. Equation and regression are shown. K_m and V_{max} values given in Table 4.1.

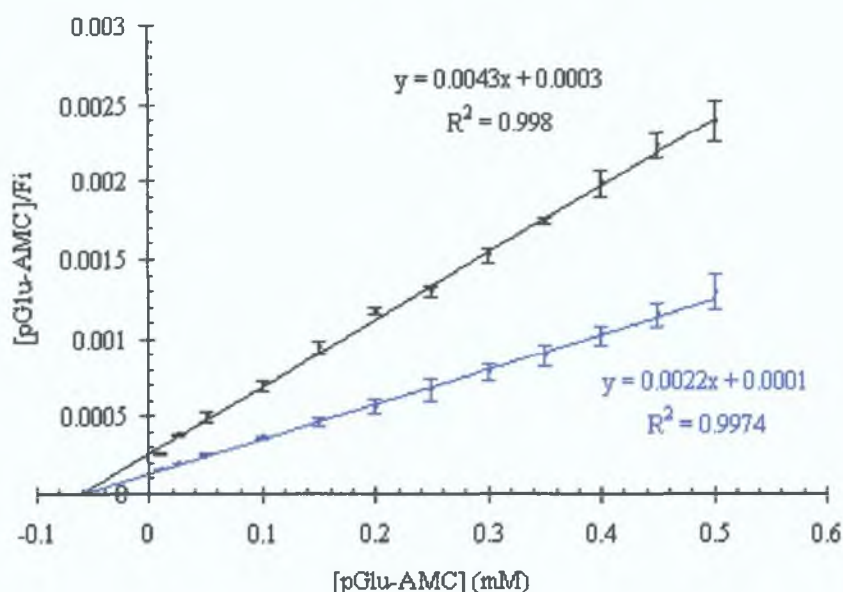


Figure 4.21 Hanes-Woolf plot for *rHsaPAP1*_{6H} and *rHsaPAP1* in lysate

Data from Figure 4.15 (same colour association) fitted to the Hanes-Woolf model for the determination of kinetic parameters, as outlined in Appendix C. Equation and regression are shown. K_m and V_{max} values given in Table 4.1.

Table 4.1 Kinetic parameters for *rHsaPAP1*_{6H} and *rHsaPAP1*

Sample	Lineweaver-Burk	Eadie-Hofstee	Hanes-Woolf	Average
K_m (μM)				
<i>rHsaPAP1</i> _{6H} (purified)	47.6	57.9	50	51.8
<i>rHsaPAP1</i> _{6H} (lysate)	42.6	53.4	69.8	55.3
<i>rHsaPAP1</i> (lysate)	45.5	65.4	45.5	52.1
V_{max} (units/ml)				
<i>rHsaPAP1</i> _{6H} (purified)	15.1	15.9	15.8	15.6
<i>rHsaPAP1</i> _{6H} (lysate)	6.7	7.1	7.4	7.1
<i>rHsaPAP1</i> (lysate)	14.4	14.6	14.4	14.5
k_{cat} (s^{-1})				
<i>rHsaPAP1</i> _{6H} (purified)	2.43	2.56	2.54	2.51

[K_m and V_{max} values obtained from Figures 4.16 to 4.21. k_{cat} calculated as outlined in Appendix C].

4.2.1 Competitive inhibition of recombinant human PAP1 by pGlu-peptides

The competitive influence of selected peptide substrates pGlu-Ala, pGlu-Val, pGlu-His-Gly and pGlu-His-Pro-NH₂ (TRH) on the pGlu-AMC degrading activity of *rHsaPAP1*_{6H} was investigated as described in Section 2.19.4. The resulting data sets were applied to the aforementioned kinetic models: Lineweaver-Burk (Figure 4.22), Eadie-Hofstee (Figure 4.23) and Hanes-Woolf (Figure 4.24). From these plots the apparent K_m (K_m^{app}) was determined in the presence of each peptide. These K_m^{app} values were used to calculate the dissociation constants (K_i) as outlined in Appendix C, which are given in Table 4.2. For example the average K_m^{app} for TRH, from the three kinetic models (Figures 4.22 to 4.24), was 370.1 μM . Using the formula described in Appendix C the K_i was calculated to be 40.7 μM .

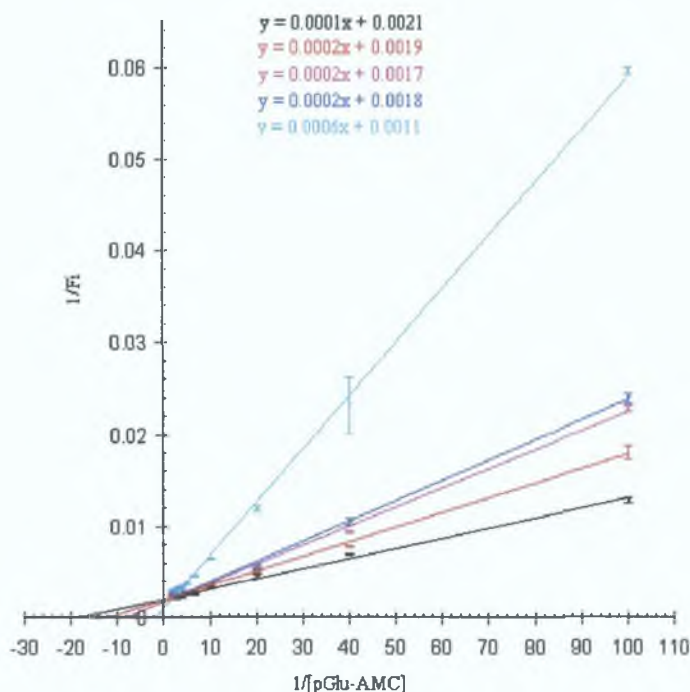


Figure 4.22 K_i determination for selected peptides using Lineweaver-Burk plot
 Data from peptide competition assays (Section 2.19.4) fitted to Lineweaver-Burk model, as outlined in Appendix C, for the determination of apparent K_m (K_m^{app}). Peptides used: pGlu-His-Pro-NH₂ (TRH), pGlu-Ala, pGlu-His-Gly and pGlu-Val. Non-inhibited data plotted in black. Equations are shown. K_i values given in Table 4.2.

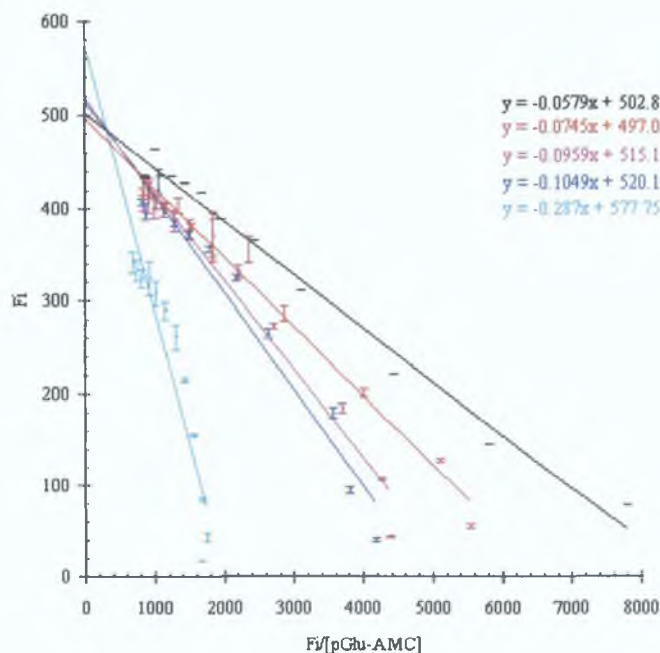


Figure 4.23 K_i determination for selected peptides using Eadie-Hofstee plot
 Data from peptide competition assays (Section 2.19.4) fitted to Eadie-Hofstee model, as outlined in Appendix C, for the determination of apparent K_m (K_m^{app}). Peptides used: pGlu-His-Pro-NH₂ (TRH), pGlu-Ala, pGlu-His-Gly and pGlu-Val. Non-inhibited data plotted in black. Equations are shown. K_i values given in Table 4.2.

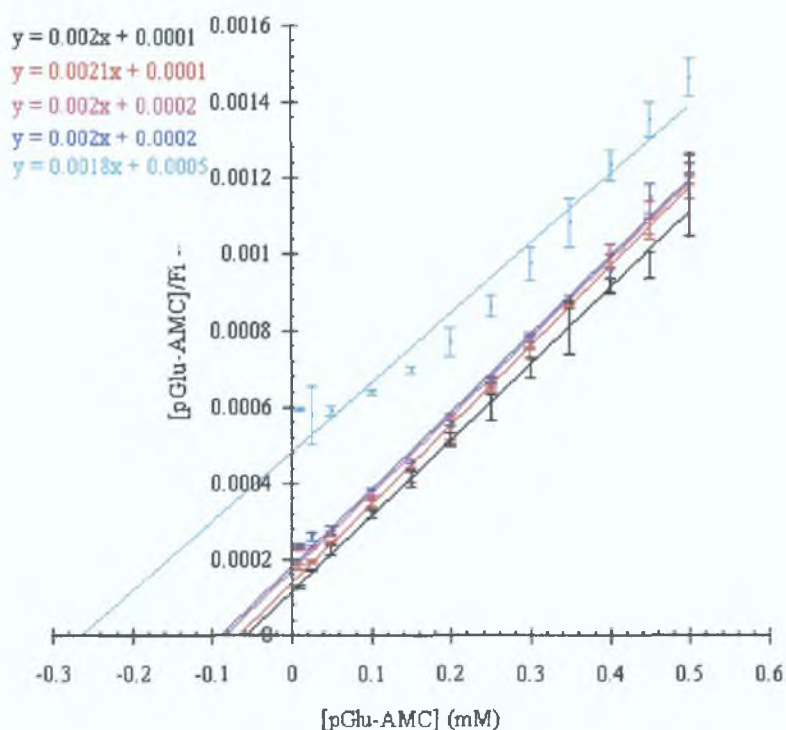


Figure 4.24 K_i determination for selected peptides using Hanes-Woolf plot

Data from peptide competition assays (Section 2.19.4) fitted to Hanes-Woolf model, as outlined in Appendix C, for the determination of apparent K_m (K_m^{app}). Peptides used: pGlu-His-Pro-NH₂ (TRH), pGlu-Ala, pGlu-His-Gly and pGlu-Val. Non-inhibited data plotted in black. Equations are shown. K_i values given in Table 4.2.

Table 4.2 K_i values for selected peptides

Peptide	K_m^{app} (μ M)			Average	K_i (μ M)
	Lineweaver-Burk	Eadie-Hofstee	Hanes-Woolf		
None	47.6	57.9	50	51.8	-
pGlu-His-Pro (TRH)	545.5	287	277.8	370.1	40.7
pGlu-Ala	111.1	104.9	100	105.3	242.1
pGlu-His-Gly	117.7	95.9	100	104.5	245.7
pGlu-Val	105.3	74.5	47.6	76.1	532.9

[K_m^{app} values obtained from Figures 4.22 to 4.24. K_i calculated as outlined in Appendix C].

4.3 Determining crystallisation conditions for recombinant human PAP1

A preliminary study was carried out to determine conditions under which rHsaPAP1_{6H} could form protein crystals. The methodology of sparse matrix screening (Jancarik and Kim, 1991; Cudney *et al.*, 1994) was employed as outlined in Section 2.19.7. Purified rHsaPAP1_{6H} (Section 3.7.1), which has been extensively dialysed, was used at various concentrations from 1.5 up to 12 mg/ml. Crystals were routinely observed after 48 hrs, in Screen 1 reagents 1, 24 and 45 (Table 2.7). Figure 4.25 shows typical examples of these crystals. Reagent 1 crystals were usually of a multiple *slate* morphology, while Reagent 24 crystals were also *slates* but seemed to be composed of several fused layers. Reagent 45 crystals had a *star* morphology, with multiple shards joining centrally.

In a collaboration, these crystals were initially analysed, using X-ray diffraction, by Dr. John Gallagher (DCU, Ireland) during his time at the Department of Structural Biology and Biochemistry, Howell Laboratories, Hospital for Sick Children, Toronto, Canada. Crystals were manipulated into capillary tubes and placed in an X-ray beam, resulting in a diffraction pattern. Unfortunately, preliminary reports show that these diffraction patterns are typical of mixed salt and protein crystals. Despite the extensive dialysis carried out on the *rHsaPAP1*_{6H} samples, conditions of even lower salt content would have to be achieved. Screening for further crystallisation conditions is ongoing.

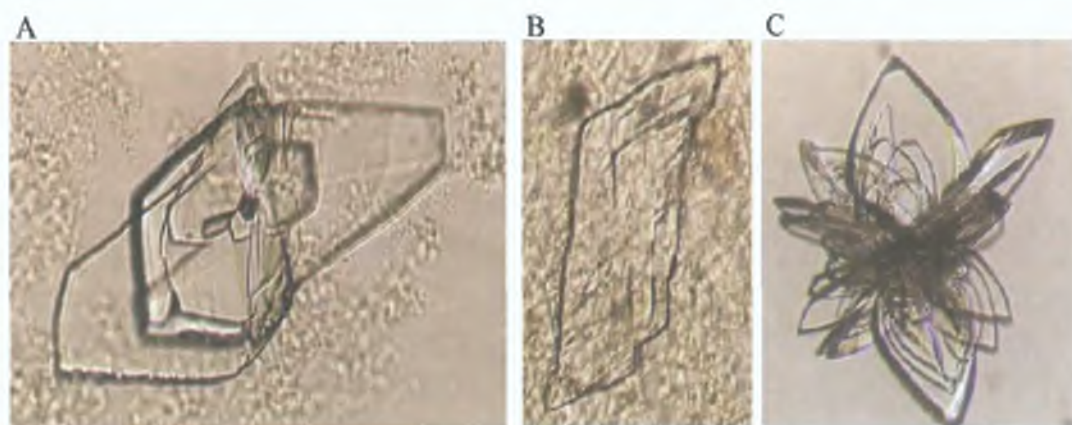


Figure 4.25 Crystals observed while attempting crystallisation of *rHsaPAP1*_{6H}

Typical examples of crystals observed during the study to determine crystallisation conditions for *rHsaPAP1*_{6H} (Section 2.19.7). Crystals were observed within 48 hrs in Screen 1, reagent 1 (A), reagent 24 (B) and reagent 45 (C) for *rHsaPAP1*_{6H} concentrations ranging from 1.5 to 12 mg/ml. Images viewed at 10x magnification.

4.4 Discussion

The catalytically active *rHsaPAP1*_{6H} was shown to have a relative molecular mass of 24.5 kDa (Section 4.1.1). This value correlates very well with the molecular weight of 24,105, deduced from the amino acid sequence for *rHsaPAP1*_{6H} as mentioned in Section 3.5. An independent study (Dando *et al.*, 2003), published shortly after this experiment was completed, reported that the non His-tagged, catalytically active, recombinant human PAP1 has a relative molecular mass of 24 kDa. The size difference attributed to the His-tag is merely ~1 kDa as shown in Section 3.5 (Figure 3.13A). This combined data concludes that recombinant human PAP1 is active as a monomeric enzyme in both the native and His-tagged form. This is in agreement with the findings (Table 1.6) that eukaryotic PAP1 is generally monomeric.

The isoelectric point (pI), the pH at which the protein has an equal number of positive and negative charges, was predicted as 5.97 for *rHsaPAP1*_{6H} (5.54 for *rHsaPAP1*) using

online tools (Section 2.11). As shown in Section 4.1.2 (Figure 4.2) the experimentally determined isoelectric point for *rHsaPAP1_{6H}* lies between 6.0 and 6.9. The pI for PAP1 enzymes has previously been reported within the range of 4.9 to 5.5 (Table 1.6).

The pH profiles in Figure 4.3 show that both *rHsaPAP1* and *rHsaPAP1_{6H}* have a pH optimum between 8.0 and 9.5. The samples in cleared lysate correlate very closely, while the purified *rHsaPAP1_{6H}* shows a particular peak of activity at pH 8.5. Throughout this work, enzymatic analysis was carried out at pH 8.0, being a compromise between the physiological pH 7.4 and the optimum determined in the above study. pH optima for various eukaryotic and prokaryotic PAP1s have been published (Table 1.6), falling in the range of 6.0 to 9.5.

The activity of *rHsaPAP1_{6H}* was shown to increase with temperature, having an optimum at 50°C (Figure 4.4), for a 15 min assay. Above 50°C however, activity was dramatically reduced. A similar 50°C optimum has been reported for non His-tagged, recombinant human PAP1 (Dando *et al.*, 2003). Throughout this work, activity assays were carried out at the physiological temperature of 37°C. Even though higher activity was achieved at 50°C, this corresponds to a short 15 min assay and is not necessarily reflective of the true physiological optimum. This was supported by the thermal stability of *rHsaPAP1_{6H}*, which was shown to be robust up to 40°C (Figure 4.5) for two hours. Above this temperature the activity rapidly declined. At 60°C the activity had reduced to about 30% after 20 min and was almost completely lost after 60 min.

rHsaPAP1_{6H} was shown have an absolute requirement for a thiol reducing agent (Figure 4.6). This cysteine protease characteristic of PAP1 has been widely reported (Section 1.3.7). The optimum concentration of DTT, used throughout this work, was 10 mM. The thiol-dependent nature of *rHsaPAP1_{6H}* was further demonstrated by the complete inhibition of activity by 100 µM of the thiol-blocking agent iodoacetate (Section 4.1.6). In contrast, the serine inhibitor DFP had no effect on activity (Figure 4.10). Another serine inhibitor PMSF, known to inhibit some cysteine proteases, exhibited inhibition with an IC₅₀ of 4.1 mM (Figure 4.11).

The chelating agent EDTA, although not a requirement for PAP1 activity (Figure 4.7), was routinely included in the standard PAP1 activity assay (Section 2.15.2), as a protective agent for DTT, according to the modified method described by Browne and O’Cuinn (1983).

The inhibitory effect by L-pGlu ($IC_{50} = 2.75$ mM, Figure 4.8) and substrate analogue 2-pyrrolidone ($IC_{50} = 1.0$ mM Figure 4.9) on the pGlu-AMC degrading activity of rHsaPAP1_{6H} is due to the competitive binding into the catalytic site, as described in Section 1.3.8.

The chelating agent 1,10-phenanthroline caused inhibition with an IC_{50} of 2.6 mM (Figure 4.12) while the structurally similar, yet non-chelating, 4,7-phenanthroline did not affect activity (Figure 4.13). It is not known why 1,10-phenanthroline caused inhibition. These results would suggest that inhibition is caused by its chelating property. This is unlikely however, given the lack of inhibition by the chelator EDTA (Figure 4.7). Rather, inhibition may be due to some structural property which 1,10-phenanthroline does not share with 4,7-phenanthroline.

The Michaelis-Menten curves (Figures 4.14 and 4.15) show that up to about 200 μ M pGlu-AMC, the PAP1 substrate concentration is limiting to the enzymatic reaction. For this reason, the standard activity assay used in this work (Section 2.15.2) used 250 μ M pGlu-AMC, ensuring a substrate excess.

The kinetic data in Table 4.1 shows that relatively little variance was observed between the three kinetic models plotted in Figures 4.16 to 4.21. The average K_m determined for purified rHsaPAP1_{6H} was 51.8 μ M (Table 4.1). The average K_m determined for rHsaPAP1_{6H} and rHsaPAP1 in cleared lysate were 55.3 and 52.1 μ M respectively. The presence of the His-tag clearly did not affect the K_m of recombinant human PAP1. A K_m of 50 μ M for the same substrate was independently reported for non His-tagged, recombinant human PAP1 (Dando *et al.*, 2003). Similar values for K_m were determined for various eukaryotic PAP1 enzymes, as listed in Table 1.6.

As can be seen in Table 4.1, a higher V_{max} was observed for rHsaPAP1 than rHsaPAP1_{6H} in cleared lysate. This was due to different concentrations of enzyme in the samples. While the K_m is independent of enzyme concentration, V_{max} increases the more enzyme is present. Since the concentration of purified rHsaPAP1_{6H} was known, the turnover number (k_{cat}), the number of substrate molecules converted into product by one enzyme molecule per second, could be calculated as 2.51 s⁻¹.

The K_i values (Table 4.2) exhibited by rHsaPAP1_{6H} for the selected peptide substrates, pGlu-Ala, pGlu-Val, pGlu-His-Gly and pGlu-His-Pro-NH₂ (TRH), give an indication of their ability to act as competitive inhibitors. With the exception of TRH, the values are in agreement with the reported preference PAP1 has for the residue directly adjacent to

pGlu (Section 1.3.8), namely: Ala>His>>Val (K_i : 242.1<245.7<<532.9 μM). The K_i for TRH (pGlu-His-Pro-NH₂) was 40.7 μM , indicating rHsaPAP1_{6H} has a relatively higher specificity for this than the other pGlu peptides. This increased specificity is probably not simply due to TRH being a tripeptide (considering the K_i for pGlu-His-Gly), but rather the identity of the third residue: proline.

The crystallisation work carried out in this thesis was unable to provide protein crystals which would be suitable for X-ray diffraction analysis. However, this study is currently ongoing in Dr. Brendan O'Connor's Neuroscience laboratory, DCU, Ireland. It is hoped that human PAP1 protein crystals of the required quality will soon be available. X-ray diffraction is a critical step towards elucidating the three-dimensional structure of a protein. Once the structure of human PAP1 has been determined, it will be of great benefit to further work, as discussed in Section 5.3.1.

5.0 Structural analysis of recombinant human PAP1

5.1 Modelling the structure of human PAP1

To date, there has been no publication of the three-dimensional (3D) structure of human or other eukaryotic PAP1 enzymes. However, the 3D structures of four prokaryotic PAP1 enzymes have recently been solved using X-ray crystallography. These PAP1 structures, namely for *B. amyloliquefaciens* (*Bam*PAP1), *T. litoralis* (*Tli*PAP1), *P. furiosus* (*Pfu*PAP1) and *P. horikoshii* (*Pho*PAP1) are described in detail in Section 1.3.10. *Hsa*PAP1 has considerable homology to prokaryotic PAP1 enzymes, exhibiting greatest homology with *Bam*PAP1, as shown in Figure 1.6. Based on this homology a 3D structure model of *Hsa*PAP1 was created, using the prokaryotic structures as templates. The amino acid sequence for human PAP1 (*Hsa*PAP1, Figure 3.10) was submitted to the SWISS-MODEL web server (Section 2.11), which provides an automated protein homology-modelling service (Guex and Peitsch, 1997; Schwede *et al.*, 2003). In short, SWISS-MODEL performs a search of the Protein Data Bank (PDB, Berman *et al.*, 2000) for primary sequence homology to the target (*Hsa*PAP1). The PDB archive contains all known 3D protein crystal structures. If sufficient homology is found with one or more primary sequences in the archive, the 3D structure of the target is predicted by using the structures corresponding to the homologous sequences as a template. The resulting 3D model is returned by e-mail as a structural coordinate file. In the case of *Hsa*PAP1, SWISS-MODEL automatically selected the PDB entries for *Bam*PAP1, *Tli*PAP1 and *Pfu*PAP1 as template structures.

A ribbon diagram of the *Hsa*PAP1 3D structure model is shown in Figure 5.1, in comparison to the three template structures. The α -helices and β -strands in the model have been indicated by colouration. The structural pattern of *Hsa*PAP1 closely resembles the three prokaryotic structures. The structure is a single α/β globular domain, having a central β -sheet surrounded by α -helices. The location of these structural features on the primary sequence are indicated on an alignment of *Hsa*PAP1 with *Bam*PAP1, *Tli*PAP1 and *Pfu*PAP1 (Figure 5.2). The same alignment, showing primary sequence homology, is given in Figure 5.3. Homology domains indicated on alignments of prokaryotic and eukaryotic PAP1 sequences in Section 1.3.6 (Figures 1.4 to 1.6), have again been indicated in Figures 5.2 and 5.3.

SWISS-MODEL excluded three N-terminal and twenty-four C-terminal residues from the *Hsa*PAP1 model. These residues are indicated in Figures 5.2 and 5.3. The C-terminal region is where the least homology is observed (Figure 5.3). In the prokaryotic enzymes the C-terminal region contributes greatly toward interactions between

monomeric subunits resulting in a tetrameric arrangement, as reported in Section 1.3.10.1. Experimental evidence so far to date (Section 1.3.9), including data presented in this thesis (Section 4.1.1), supports a monomeric structure for active eukaryotic PAP1 enzymes. This would be in keeping with the observed divergence of sequence homology between prokaryotic and eukaryotic PAP1 enzymes at the C-terminal.

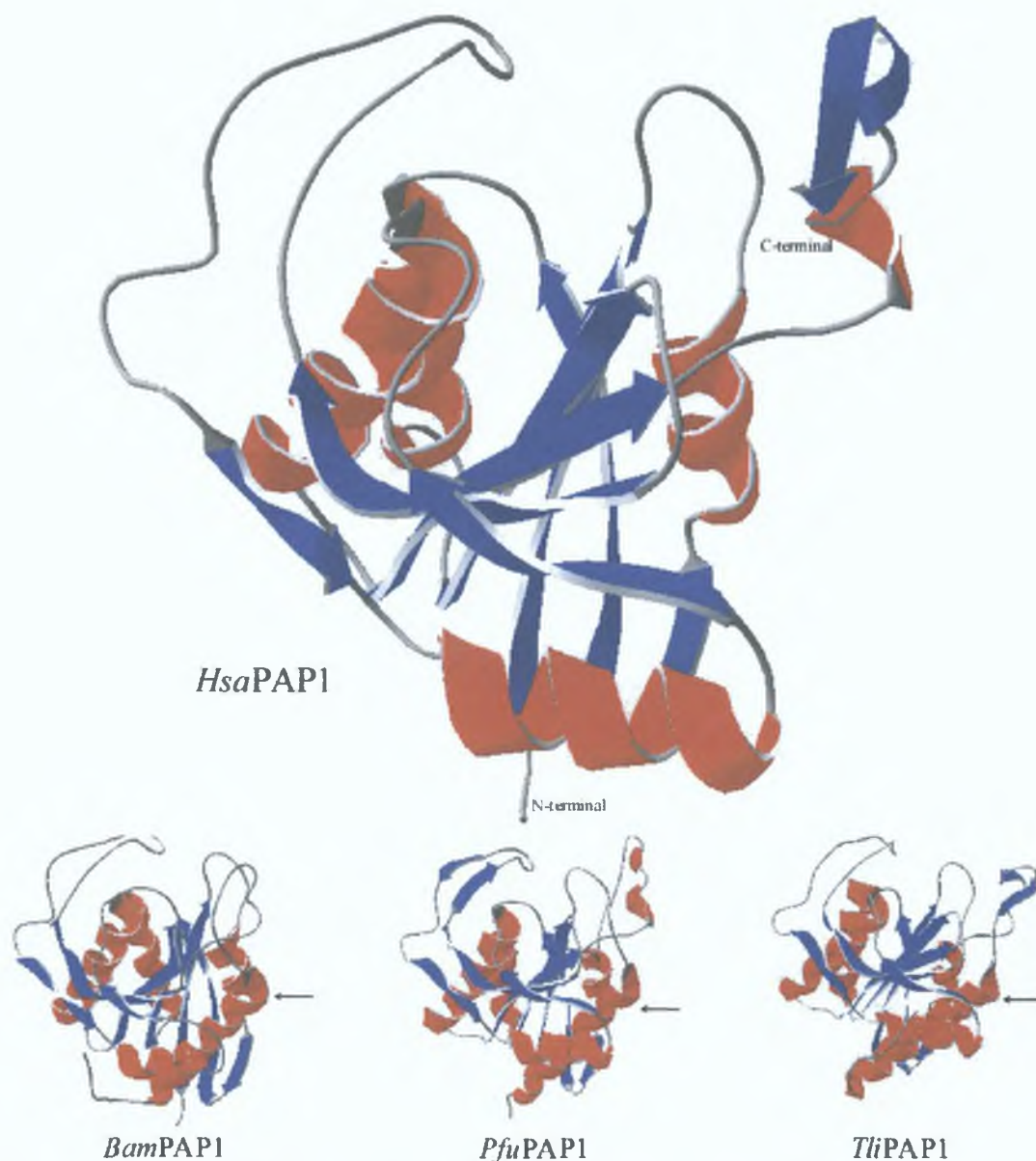


Figure 5.1 *HsaPAP1* 3D model compared with structural templates

Ribbon diagram of the 3D *HsaPAP1* model, which was predicted using the SWISS-MODEL server (Section 2.11), taking the structures of *BamPAP1*, *PfuPAP1* and *TliPAP1* (Section 1.3.10) as templates. α -helices are coloured red and β -strands are coloured blue. 3 N-terminal and 24 C-terminal *HsaPAP1* residues were excluded from the model. The residues of *BamPAP1*, *PfuPAP1* and *TliPAP1* correlating to the excluded C-terminal residues form α -helices, indicated by arrows. Generated using DeepView (Section 2.11).



Figure 5.2 Primary sequence alignment showing structural features

Alignment of *HsaPAP1* with *BamPAP1*, *TliPAP1* and *PfuPAP1*, identical to the homology alignment in Figure 5.3. Residues forming α -helices and β -strands (as shown in Figure 5.1) are coloured red and blue respectively. Residues which are not included in the *HsaPAP1* model are marked by black line. Homology domains reported in Figures 1.4 to 1.6 are indicated in red. For amino acid information see Appendix D. Created using MultAlin with Blossum62-12-2 parameters and edited using GenDoc (Section 2.11).

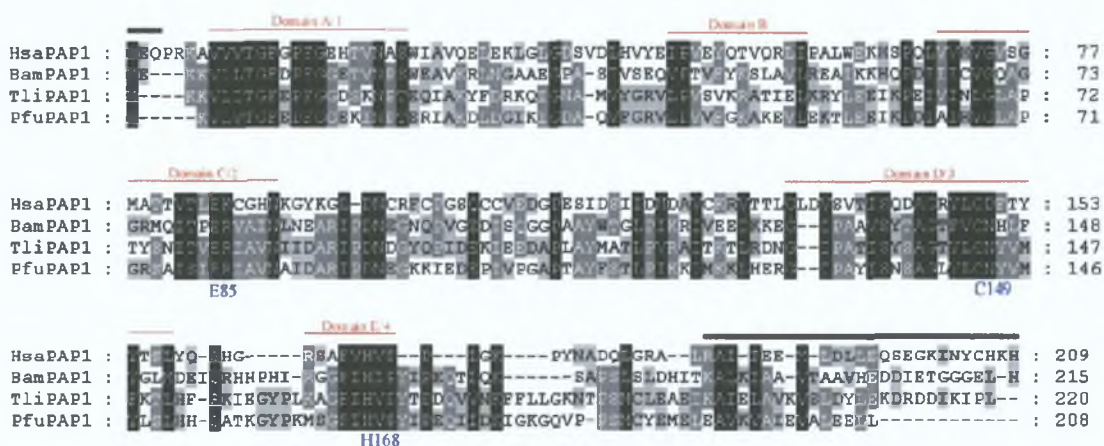


Figure 5.3 Primary sequence homology alignment

Homology alignment of *HsaPAP1* with *BamPAP1*, *TliPAP1* and *PfuPAP1*. Sequence homology is represented by grey scale shading, with black being the highest homology. Residues which are not included in the *HsaPAP1* model are marked by black line. Homology domains reported in Figures 1.4 to 1.6 are indicated in red. Catalytic triad residues Glu85, Cys149 and His168 are marked in blue (numbering applies to *HsaPAP1*). For amino acid information see Appendix D. Created using MultAlin with Blossum62-12-2 parameters and edited using GenDoc (Section 2.11).

5.1.1 Validation of the human PAP1 structure model

As described in Section 1.3.7 the catalytic triad of prokaryotic PAP1 enzymes was shown to consist of Glu-Cys-His. By primary sequence alignment (Figure 5.3) these residues correspond to Glu85, Cys149 and His168 in *HsaPAP1*. Strikingly, in the *HsaPAP1* model of Figure 5.1, the orientation of these residues relative to each other is very similar to that observed for *BamPAP1*, *TliPAP1* and *PfuPAP1* (see Figure 5.4).

If these residues constitute the catalytic site of *HsaPAP1*, one would expect the enzyme activity to be affected dramatically by introducing miss-sense mutations into the codons for these amino acids. Using a site-specific mutagenesis approach Cys149 and His168 were substituted with tryptophan and aspartic acid respectively (Figure 5.5). This was carried out as described in Section 2.9.2, using primers listed in Table 2.2. Codon changes to the recombinant human PAP1 DNA sequence (*rHsa-pap1*) were made on the *rHsaPAP1*_{6H} expression plasmid pRV5 (Figures 3.8 and 3.10), as it had been demonstrated that the His-tag does not affect the kinetic parameters of the enzyme (Table 4.1). The choice of these amino acid replacements was based on radically altering the side-chains. The resulting mutant constructs are listed in Table 5.1. Both mutations caused a complete loss of activity, supporting the catalytic significance of these residues and also lending support for the *HsaPAP1* 3D structure (Figure 5.1) obtained via SWISS-MODEL.

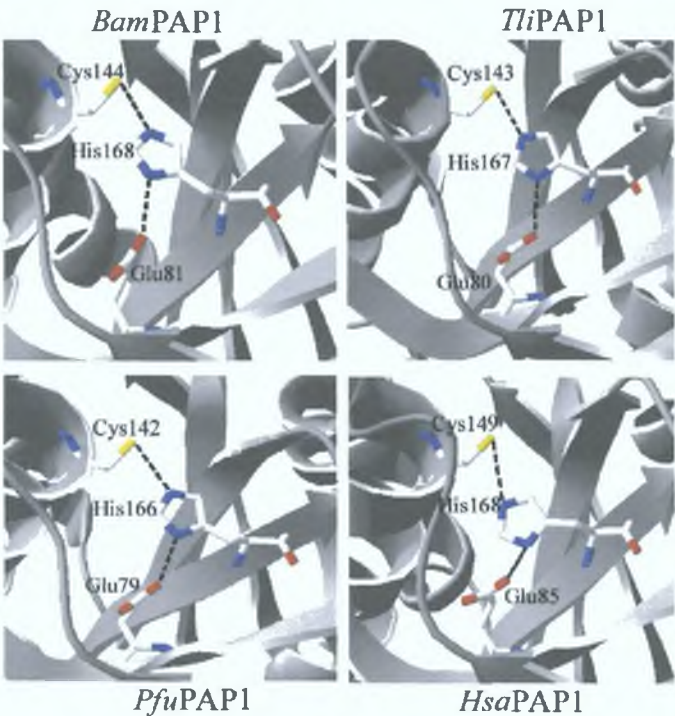


Figure 5.4 Comparison of PAP1 active sites

The side-chain orientation of the catalytic residues Glu, Cys and His for *BamPAP1*, *TliPAP1*, *PfuPAP1* and the *HsaPAP1* model. Hydrogen bonds are indicated by broken lines. Colouration of residues by CPK colour scheme (C = white, O = red, N = blue and S = yellow). For amino acid information see Appendix D. Generated using DeepView (Section 2.11).

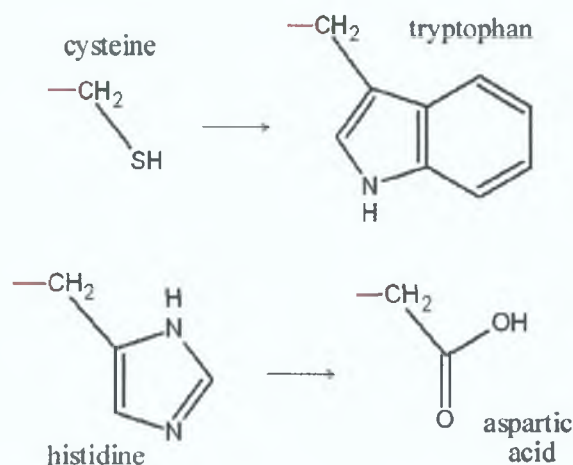


Figure 5.5 Substitution of *Hsa*PAP1 residues Cys149 and His168

Amino acid side-chains of the active site residues Cys149 and His168, which were substituted with tryptophan and aspartic acid respectively. For amino acid information see Appendix D. Illustrated using ChemSketch (Section 2.11).

5.2 Functional analysis of recombinant human PAP1

Continuing to use the 3D structure model of *Hsa*PAP1 produced by SWISS-MODEL (Section 5.1), further residues of the enzyme (Section 5.2.1 to 5.2.3) were targeted for functional analysis using site-specific mutagenesis. Figure 5.6 shows the location of these residues on a primary sequence alignment and Figure 5.7 shows how these residues are positioned in the *Hsa*PAP1 model. As with the catalytic residue mutants (Section 5.1.1) the codon changes were introduced on the expression plasmid pRV5 (Figures 3.8 and 3.10) by the method described in Section 2.9.2, using primers listed in Table 2.2. Table 5.1 gives a list of all mutant pRV5 constructs as well as names of the corresponding mutant r*Hsa*PAP1_{6H} enzymes.

Where possible, silent mutations were co-introduced to create a diagnostic restriction site. The silent mutations introduced in each construct are listed in Table 2.3 and the diagnostic restriction sites are indicated in Table 2.2. These restriction sites allowed for verification of mutant pRV5 constructs prior to DNA sequencing. For example, with the creation of pRV5_I120mG (Table 5.1), having the Ile120 codon changed to a Gly codon (Section 5.2.2), a *Pvu*I site was introduced (see Figure 5.8B). The pRV5 construct already has one *Pvu*I site within *amp*^R. A *Pvu*I restriction of pRV5_I120mG gave a 1237 bp band, whereas pRV5 was linearised (see Figure 5.8A). Obtaining a band of the expected size confirmed successful mutagenesis.

However, this approach may need to be reassessed in light of some of the data presented in Section 3.6.2, where silent codon mutations were found to result in a significant drop in recombinant protein production.

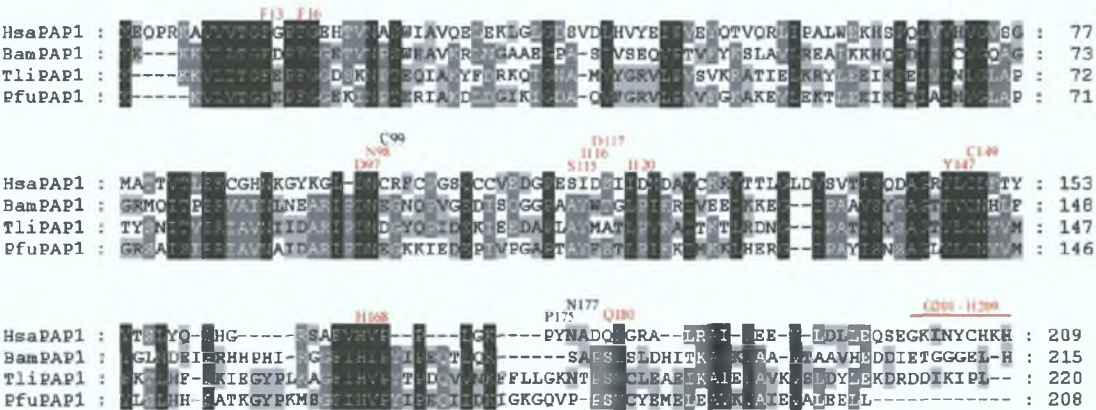


Figure 5.6 Primary sequence alignment showing residues targeted for mutagenesis
Homology alignment of *HsaPAP1* with *BamPAP1*, *TliPAP1* and *PfuPAP1*. Sequence homology is represented by grey scale shading, with black being the highest homology. Residues which were targeted for site-specific mutagenesis are marked in red. Residues affected by unintentional mutagenesis are marked in black. For amino acid information see Appendix D. Created using MultAlin with Blosum62-12-2 parameters and edited using GenDoc (Section 2.11).

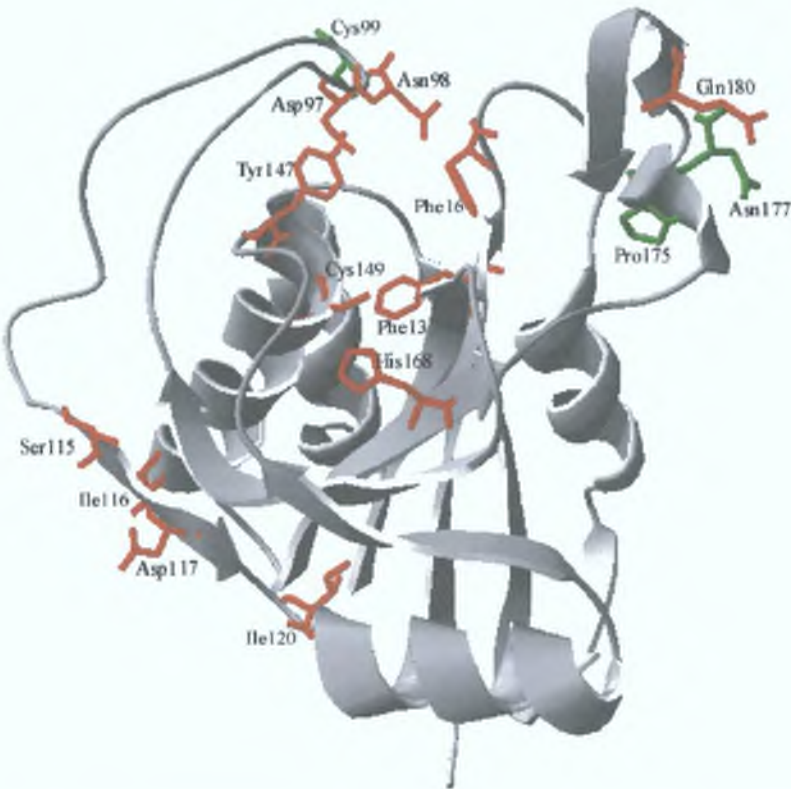


Figure 5.7 *HsaPAP1* 3D model showing residues targeted for mutagenesis
Ribbon diagram of the 3D *HsaPAP1* model. Residues targeted for site-specific mutagenesis are visualised in red. Residues affected by unintentional mutagenesis are marked in green. For amino acid information see Appendix D. Generated using DeepView (Section 2.11).

Table 5.1 Mutant derivatives of rHsaPAP1_{6H}

Mutation	Plasmid	Enzyme
Wild Type	pRV5	rHsaPAP1 _{6H}
C149 to W	pRV5_C149mW	rHsaPAP1_C149mW _{6H}
H168 to D	pRV5_H168mD	rHsaPAP1_H168mD _{6H}
F13 to W	pRV5_F13mW	rHsaPAP1_F13mW _{6H}
F13 to Y	pRV5_F13mY	rHsaPAP1_F13mY _{6H}
F13 to Y, P175 to S	pRV5_F13mY.P175mS	rHsaPAP1_F13mY.P175mS _{6H}
F13 to L	pRV5_F13mL	rHsaPAP1_F13mL _{6H}
F13 to L, F16 to L	pRV5_F13mL.F16mL	rHsaPAP1_F13mL.F16mL _{6H}
F16 to L	pRV5_F16mL	rHsaPAP1_F16mL _{6H}
F16 to Y	pRV5_F16mY	rHsaPAP1_F16mY _{6H}
Y147 to F	pRV5_Y147mF	rHsaPAP1_Y147mF _{6H}
D97 to E	pRV5_D97mE	rHsaPAP1_D97mE _{6H}
D97 to E, N98 & C99 to M	pRV5_D97mE.N98C99mM	rHsaPAP1_D97mE.N98C99mM _{6H}
N98 to Q	pRV5_N98mQ	rHsaPAP1_N98mQ _{6H}
N98 to V	pRV5_N98mV	rHsaPAP1_N98mV _{6H}
S115 to A, I116 to Y, D117 to F	pRV5_SID115mAYF	rHsaPAP1_SID115mAYF _{6H}
GGG added after D117	pRV5_D117+GGG	rHsaPAP1_D117+GGG _{6H}
I120 to G	pRV5_I120mG	rHsaPAP1_I120mG _{6H}
Q180 to E	pRV5_Q180mE	rHsaPAP1_Q180mE _{6H}
Q180 to E, N177 to Y	pRV5_Q180mE.N177mY	rHsaPAP1_Q180mE.N177mY _{6H}
C-terminal deletion: G201 to H209	pRV5_Δ3'	rHsaPAP1_ΔC _{6H}
	(pRV5_Δ3'Δ6H - without 6xHis)	(rHsaPAP1_ΔC)

[On three occasions, unintended mutations (indicated in red) were obtained alongside the planned substitutions. For amino acid information see Appendix D. The DNA sequencing data for all mutant constructs is given in Appendix B.]

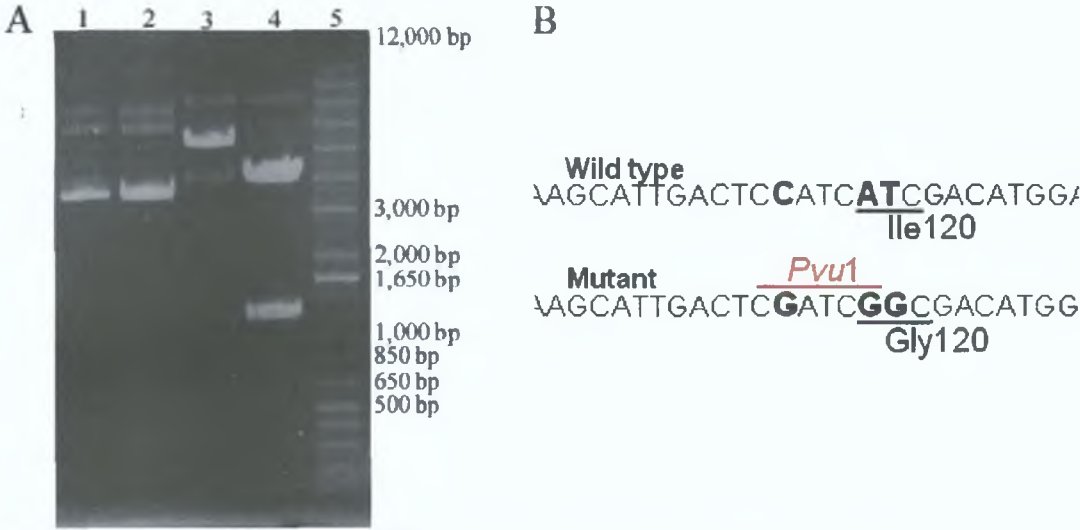


Figure 5.8 Verification of mutation using diagnostic restriction analysis

(A): Restriction digests of pRV5 and pRV5_I120mG (Table 2.3) using *Pvu*I, analysed on 0.7% agarose gel (Section 2.5). pRV5 has single *Pvu*I site, within amp^R (Figure 3.8). A second *Pvu*I site was introduced into pRV5_I120mG by mutation (B). A band of 1237 bp is expected from a *Pvu*I restriction of pRV5_I120mG. Lane 1, pRV5 (uncut); Lane 2, pRV5_I120mG (uncut); Lane 3, pRV5 (*Pvu*I); Lane 4, pRV5_I120mG (*Pvu*I); Lane 5, DNA Ladder (sizes as in Figure 2.5).

5.2.1 Substrate binding pocket residues of human PAP1

A common feature of peptidases is a hydrophobic 'binding pocket' that typically surrounds the catalytic residues. This type of binding pocket is known to help orientate the substrate for hydrolysis. The *Hsa*PAP1 residues Phe13, Phe16 and Tyr147 (Figure 5.7) correspond to the *Bam*PAP1 residues Phe10, Phe13 and Phe142 respectively (Figure 5.6). A recent mutational analysis study has highlighted the essential role these residues have within the substrate binding pocket of *Bam*PAP1 (Section 1.3.11). In the present study Phe13 and Phe16 of *Hsa*PAP1 were changed to Leu (both separately and simultaneously), thereby keeping the hydrophobic property but replacing the large aromatic structure with a smaller side-chain. Both residues were also changed to Tyr, conserving the aromatic structure but introducing a polar property. More dramatically, Phe13 was changed to Trp, having an even larger aromatic side-chain. Tyr147 (*Hsa*PAP1) is commonly substituted with Phe in prokaryotic PAP1 enzymes. Therefore Tyr147 was changed to introduce the triple Phe arrangement to *Hsa*PAP1. The side-chain structure changes of these substitutions are shown in Figure 5.9.

As shown in Figure 5.7, Asp97 and Asn98 are located at the outer edge of the binding pocket. While Asp97 is highly conserved (Figures 1.4 to 1.6 and 5.4), Asn98 is substituted with various residues in some of the lower eukaryotes (Figure 1.5). The charged Asp97 was conservatively changed to Glu. The polar Asn98 was conservatively changed to Gln and more dramatically to the hydrophobic Val. The side-chain structure changes of these substitutions are shown in Figure 5.10. Analysis of these mutants is presented in Section 5.2.4.

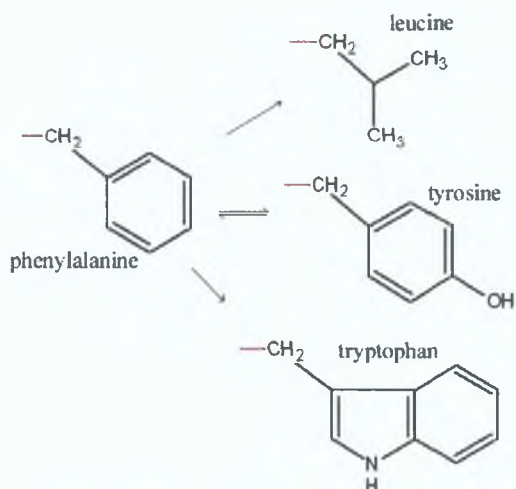


Figure 5.9 Substitution of *Hsa*PAP1 residues Phe13, Phe16 and Tyr147

Amino acid side-chains of the substrate binding pocket residues Phe13, Phe16 and Tyr147, which were substituted variably with leucine, tyrosine and tryptophan. For amino acid information see Appendix D. Illustrated using ChemSketch (Section 2.11).

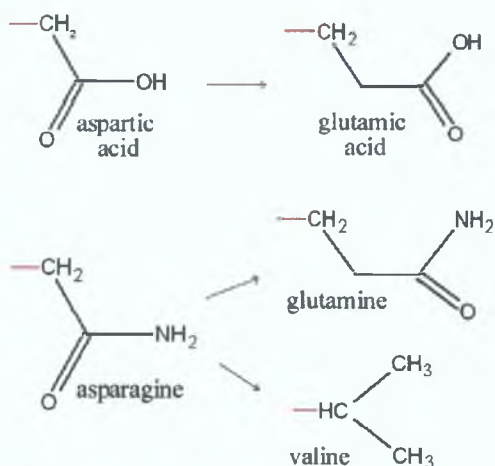


Figure 5.10 Substitution of *Hsa*PAP1 residues Asp97 and Asn98

Amino acid side-chains of the substrate binding pocket residues Asp97 and Asn98, which were substituted variably with glutamic acid, and glutamine or valine respectively. For amino acid information see Appendix D. Illustrated using ChemSketch (Section 2.11).

5.2.2 Extended loop residues of human PAP1

On an extended outer loop of *Hsa*PAP1 (Figure 5.7) the motif Ser115-Ile116-Asp117 was changed to Ala-Tyr-Phe, the latter three amino acids being commonly found in prokaryotic PAP1 enzymes (Figure 1.4). In addition to this mutation, an insertion of three Gly residues was introduced at Asp117 to further address the structural significance of the loop. The position occupied by Ile120 in *Hsa*PAP1 is conserved as hydrophobic in most PAP1 enzymes (Figures 1.4 and 1.5). This Ile120 residue in *Hsa*PAP1 is positioned at the junction between the extended loop and an α -helix. To

address its importance at the N-terminal of an α -helix (Figure 5.5), it was changed to Gly. Analysis of these mutants is presented in Section 5.2.4.

5.2.3 C-terminal residues of human PAP1

The striking lack of homology between the human PAP1 sequence, at the C-terminal end, when compared with the prokaryotic sequences (Figures 1.4 to 1.6 and 5.4), is particularly interesting in the proposed structure for *Hsa*PAP1, in which the terminal 24 residues have been excluded from the model. In order to test the role, if any, in the catalytic activity and/or stability of the enzyme, of these terminal residues, a number of additional mutations were made. These included a deletion mutation (spanning the terminal nine residues) and a substitution involving residue Gln180. The strategy used to create the C-terminal deletion is outlined in Figure 5.11. Analysis of these mutants is presented in Section 5.2.4.

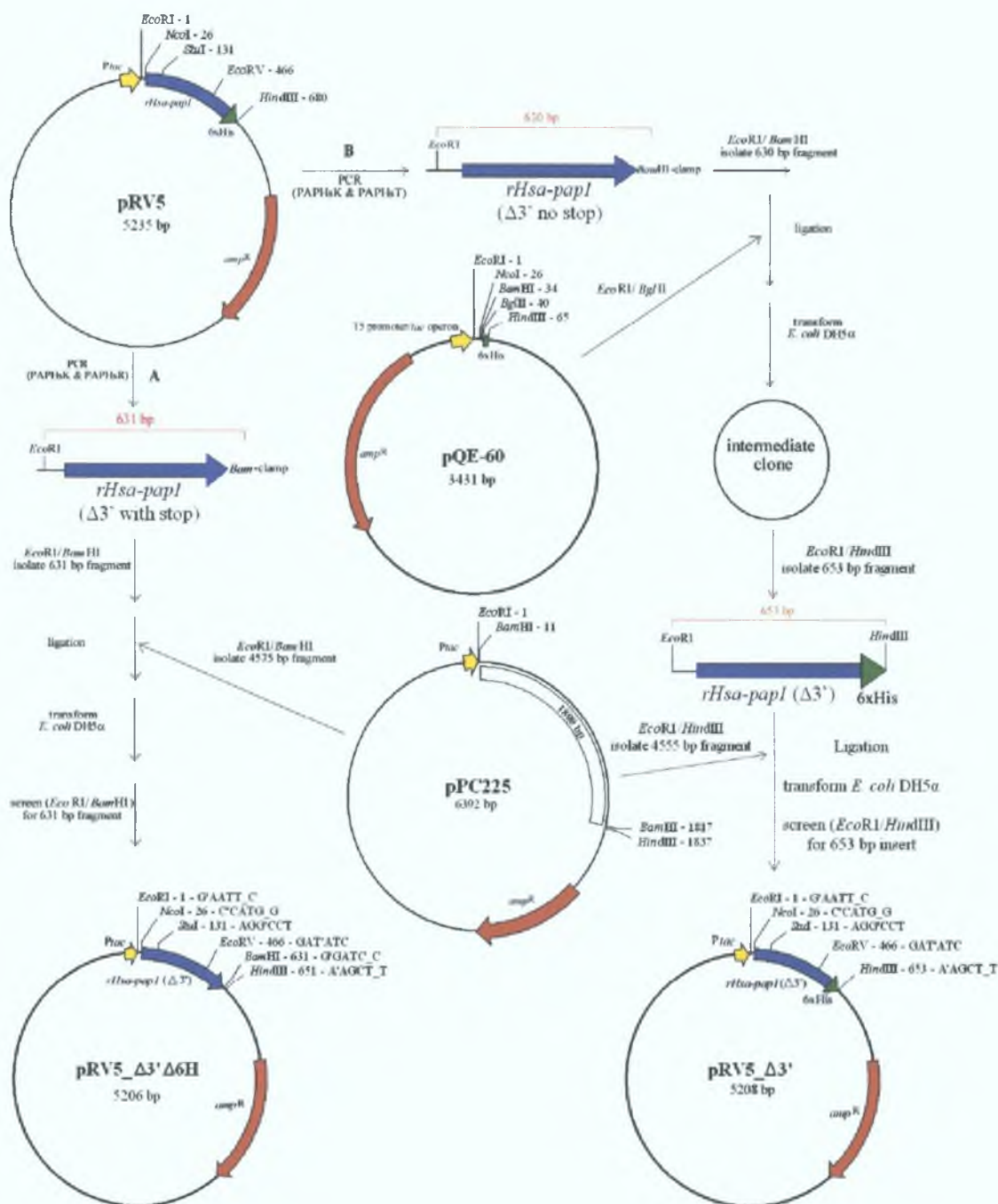


Figure 5.11 Cloning strategies for 3' deletion constructs

(A): PCR (Section 2.8.2) on pRV5 using primers PAPHsK and PAPHsR (Table 2.2) generates product with *EcoRI* site upstream from the *rHsa-pap1* having nine 3' codons deleted and the 6xHis fusion replaced by stop codon, followed by *BamHI* tag with clamp. Fragment is restricted (*EcoRI/BamHI*) and ligated into pPC225 (opened with *EcoRI/BamHI*). pRV5_Δ3'Δ6H has *rHsa-pap1* (3' deletion and without 6xHis) under control of *Ptac* promoter. (B): PCR on pRV5 using primers PAPHsK and PAPHsT (Table 2.2) generates product with *EcoRI* site upstream from the *rHsa-pap1* having 3' deletion and 6xHis fusion replaced by *BamHI* site with clamp. Fragment is restricted (*EcoRI/BamHI*) and ligated into pQE-60 (opened with *EcoRI/BglII*). Restriction (*EcoRI/HindIII*) on intermediate clone results in fragment having *rHsa-pap1* with 3' deletion, fused to 6xHis, which is ligated into pPC225 (opened with *EcoRI/HindIII*). pRV5_Δ3' has *rHsa-pap1* (with 3' deletion and 6xHis) under control of *Ptac* promoter. Restriction sites are indicated with positions (and sequences for final constructs). Illustrated using pDRAW32 (Section 2.11).

5.2.4 Analysis of recombinant human PAP1 mutants

The mutant plasmid derivatives of pRV5 (Table 5.1) were transformed into *E. coli* XL10-Gold as described in Section 2.6.3. Standard expression cultures were prepared as outlined in Section 2.12. Cleared lysate samples were assayed for PAP1 activity (Section 2.15.2). As reported in Section 5.1.1, no activity was detected for the putative catalytic triad mutants *rHsaPAP1_C149mW_{6H}* and *rHsaPAP1_H168mD_{6H}*. Nor was there any activity for *rHsaPAP1_F13mW_{6H}*, *rHsaPAP1_F13mL.F16mL_{6H}*, *rHsaPAP1_I120mG_{6H}* or *rHsaPAP1_D117+GGG_{6H}*.

All mutant proteins were found to be expressed at high levels comparable to *rHsaPAP1_{6H}*, except *rHsaPAP1_D117+GGG_{6H}*. No significant protein band of the expected molecular weight of 24,277 (deduced from amino acid sequence) could be seen for the latter. Figure 5.12 shows SDS-PAGE analysis (Section 2.16) of some cleared lysate samples from expression cultures, including from an *rHsaPAP1_D117+GGG_{6H}* culture. RT-PCR analysis (Sections 2.8.1 and 2.8.2) was performed on *rHsaPAP1_D117+GGG_{6H}* expression culture samples and a few control samples, similar to the approach in Section 3.6.2. As can be seen in Figure 5.13, the amount of mRNA being produced for *rHsaPAP1_D117+GGG_{6H}* was comparable to that of *rHsaPAP1_{6H}*. Thus, the most likely explanation for the lack of appearance of a protein band for *rHsaPAP1_D117+GGG_{6H}* is either due to adverse folding of the mRNA, leading to inefficient translation of protein, or, unstable conformation of protein structure which is particularly susceptible to degradation.

SDS-PAGE analysis (Figure 5.14) shows clearly the expected size difference between *rHsaPAP1_{6H}* and the mutant *rHsaPAP1_ΔC*, having nine C-terminal residues (and the His-tag) removed.

All mutants showing activity were purified according to the standard IMAC procedure outlined in Section 2.13.1 (Figure 5.15), except for *rHsaPAP1_ΔC*, which is lacking the His-tag necessary for IMAC purification. Purified, active *rHsaPAP1_{6H}* mutants were investigated for their kinetic parameters by the same procedures as used in Section 4.2. A summary of these parameters is given in Table 5.2, while graphical comparisons are given in Figures 5.16 to 5.18. Western blot analysis (Section 2.18) of *rHsaPAP1_{6H}* and selected mutants, in cleared lysate, is shown in Figure 5.19. Figure 5.20 shows Zymogram analysis (Section 2.17) of *rHsaPAP1_{6H}* and selected mutants in cleared lysate.

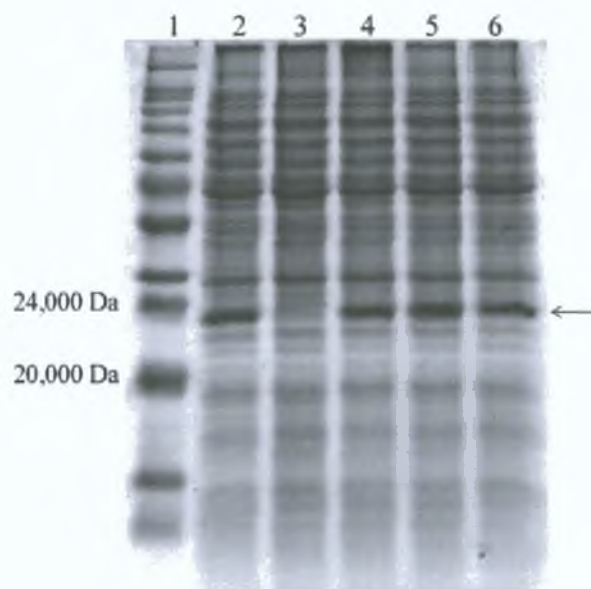


Figure 5.12 Expression of *rHsaPAP1*_{6H} mutants

Analysis of *rHsaPAP1*_{6H} mutant expression in *E. coli* XL10-Gold (Section 2.12) by 15% SDS-PAGE (Section 2.16). Samples taken 4 hrs after induction with 50 μ M IPTG. Lane 1, M_r (sizes given in Figure 2.6); Lane 2, *rHsaPAP1*_SID115mAYF_{6H}; Lane 3, *rHsaPAP1*_D117+GGG_{6H}; Lane 4, *rHsaPAP1*_I120mG_{6H}; Lane 5, *rHsaPAP1*_Q180mE; Lane 6, *rHsaPAP1*_Q180mE.N177mY_{6H}.

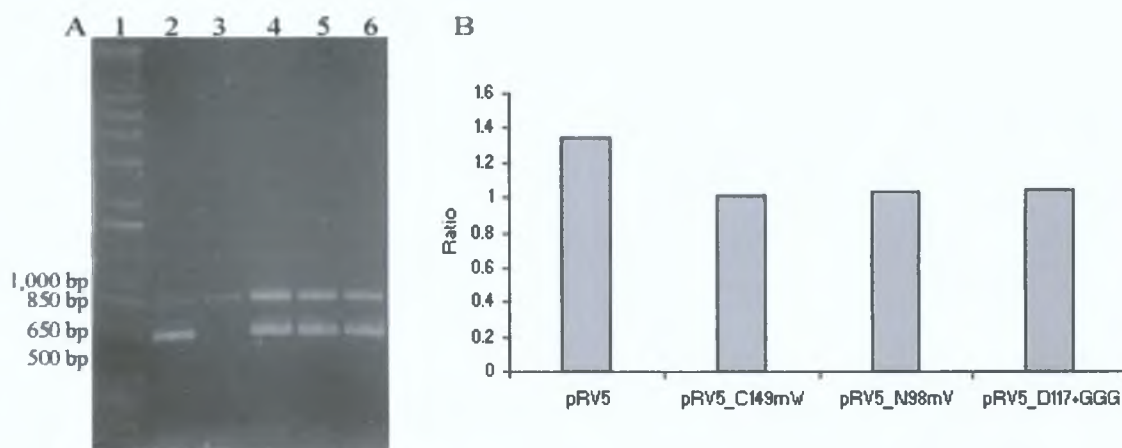


Figure 5.13 RT-PCR analysis of pRV5 mutants

(A): RT-PCR analysis (Sections 2.8.1 to 2.8.2) of RNA isolated (Section 2.4.4) from *E. coli* XL10-Gold cultures harbouring pRV5 and some mutant derivatives (Table 5.1). *rHsa-pap1* specific primers PAPHsM & PAPHsN (627 bp product) and *amp*^R specific primers Amp-for & Amp-rev (856 bp product) were used (Table 2.2). Negative controls carried out by omitting RTase (data not shown). Lane 1, DNA Ladder (sizes as in Figure 2.5); Lane 2, pRV5; Lane 3, pPC223 -ve control; Lane 4, pRV5_C149mW; Lane 5, pRV5_N98mV; Lane 6, pRV5_pRV5_D117+GGG. Ratios of *rHsa-pap1* to *amp*^R product, determined by densitometry (Section 2.16.5), are shown in bar chart (B).

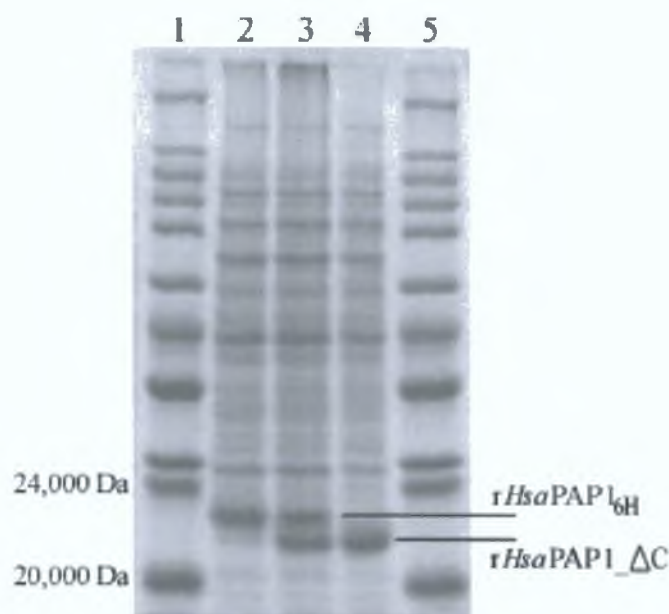


Figure 5.14 Comparison of *rHsaPAP1*_{6H} with *rHsaPAP1*_{ΔC}

Analysis of *rHsaPAP1*_{6H} compared to *rHsaPAP1*_{ΔC} in cleared lysate, by 15% SDS-PAGE (Section 2.16). Lane 1, *M_r* (sizes given in Figure 2.6); Lane 2, *rHsaPAP1*_{6H}; Lane 3, *rHsaPAP1*_{6H} & *rHsaPAP1*_{ΔC}; Lane 4, *rHsaPAP1*_{ΔC}; Lane 5, *M_r*.

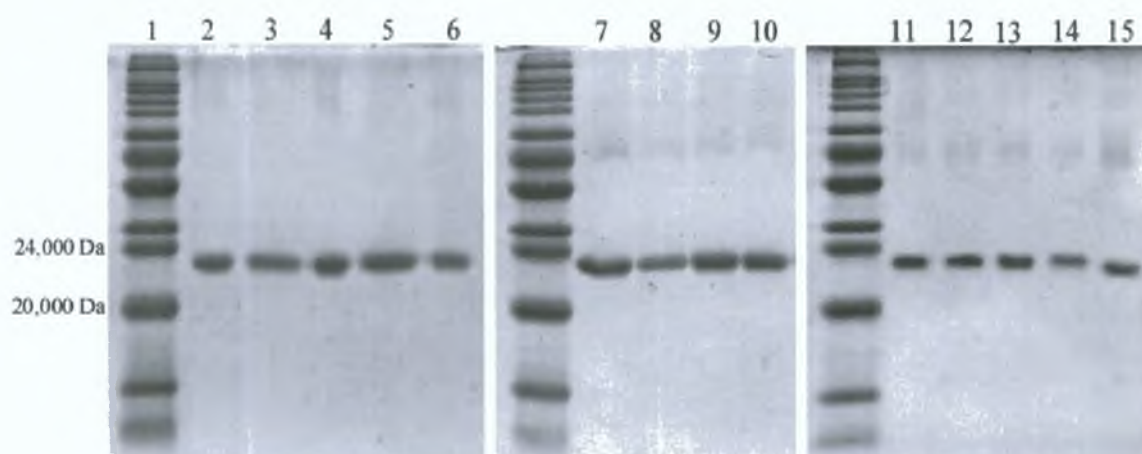


Figure 5.15 Purification of *rHsaPAP1*_{6H} mutants

Mutant derivatives of *rHsaPAP1*_{6H} (Table 5.1) purified by IMAC (Section 2.13.1) and analysed by 15% SDS-PAGE (Section 2.16). Lane 1, *M_r* (sizes given in Figure 2.6); Lane 2, *rHsaPAP1*_{6H}; Lane 3, *rHsaPAP1*_{D97mE_{6H}}; Lane 4, *rHsaPAP1*_{D97mE.N98C99mM_{6H}}; Lane 5, *rHsaPAP1*_{N98mQ_{6H}}; Lane 6, *rHsaPAP1*_{N98mV_{6H}}; Lane 7, *rHsaPAP1*_{F13mY_{6H}}; Lane 8, *rHsaPAP1*_{F16mY_{6H}}; Lane 9, *rHsaPAP1*_{F16mL_{6H}}; Lane 10, *rHsaPAP1*_{Y147mF_{6H}}; Lane 11, *rHsaPAP1*_{F13mL_{6H}}; Lane 11, *rHsaPAP1*_{F13mL_{6H}}; Lane 13, *rHsaPAP1*_{Q180mE_{6H}}; Lane 14, *rHsaPAP1*_{Q180mE.N177mY_{6H}}; Lane 15, *rHsaPAP1*_{ΔC}.

Table 5.2 Kinetic parameters of *rHsaPAP1*_{6H} mutants

Enzyme	Specific Activity (units/mg)	K _m (μM)	k _{cat} (s ⁻¹)
<i>rHsaPAP1</i> _{6H}	9334.3	51.8	3.75
<i>rHsaPAP1</i> _C149mW _{6H}	-	-	-
<i>rHsaPAP1</i> _H168mD _{6H}	-	-	-
<i>rHsaPAP1</i> _F13mW _{6H}	-	-	-
<i>rHsaPAP1</i> _F13mY _{6H}	4032.0	96.1	1.62
<i>rHsaPAP1</i> _F13mY.P175mS _{6H}	503.4	93.6	0.20
<i>rHsaPAP1</i> _F13mL _{6H}	472.6	78.3	0.19
<i>rHsaPAP1</i> _F13mL.F16mL _{6H}	-	-	-
<i>rHsaPAP1</i> _F16mL _{6H}	1213.7	161.0	0.49
<i>rHsaPAP1</i> _F16mY _{6H}	8405.4	24.9	3.38
<i>rHsaPAP1</i> _Y147mF _{6H}	9823.3	41.6	3.94
<i>rHsaPAP1</i> _D97mE _{6H}	902.3	29.9	0.36
<i>rHsaPAP1</i> _D97mE.N98C99mM _{6H}	814.2	48.2	0.33
<i>rHsaPAP1</i> _N98mQ _{6H}	5048.9	44.4	2.03
<i>rHsaPAP1</i> _N98mV _{6H}	10237.2	58.6	4.11
<i>rHsaPAP1</i> _SID115mAYF _{6H}	5742.3	27.7	2.31
<i>rHsaPAP1</i> _D117+GGG _{6H}	-	-	-
<i>rHsaPAP1</i> _I120mG _{6H}	-	-	-
<i>rHsaPAP1</i> _Q180mE _{6H}	5516.6	39.7	2.22
<i>rHsaPAP1</i> _Q180mE.N177mY _{6H}	5583.0	57.1	2.25
<i>rHsaPAP1</i> _ΔC _{6H}	12301.7	56.1	4.72

[K_m values are averages of those obtained from the three kinetic models described in Appendix C. k_{cat} values calculated as outlined in Appendix C.]

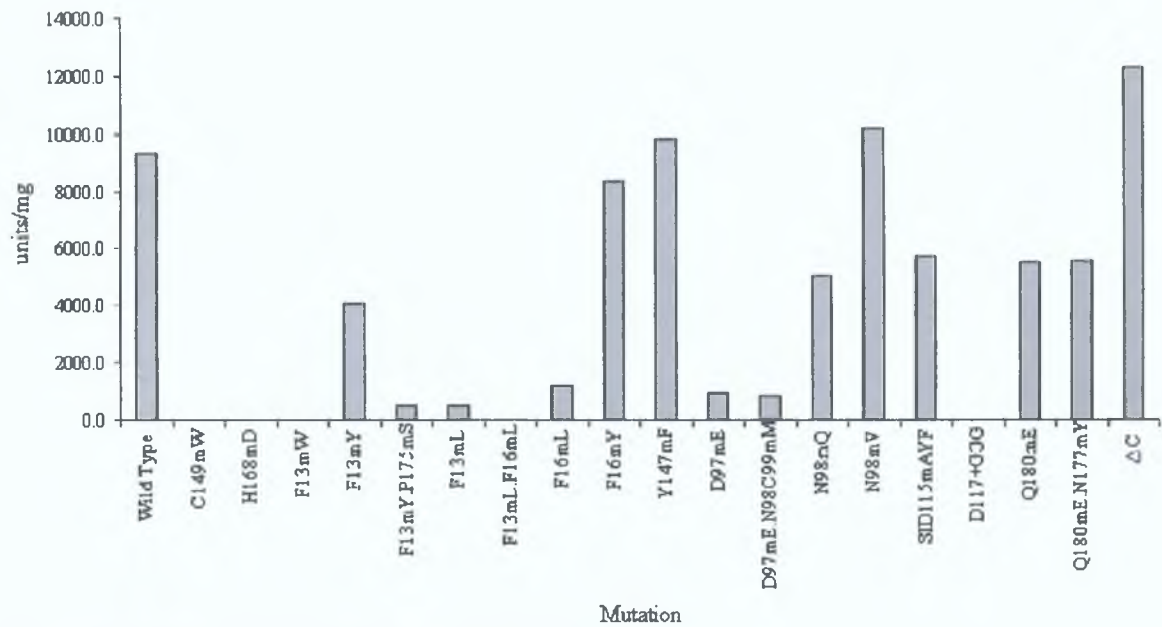


Figure 5.16 Effect of mutations on specific activity

Values for specific activity (units/mg) taken from Table 5.2

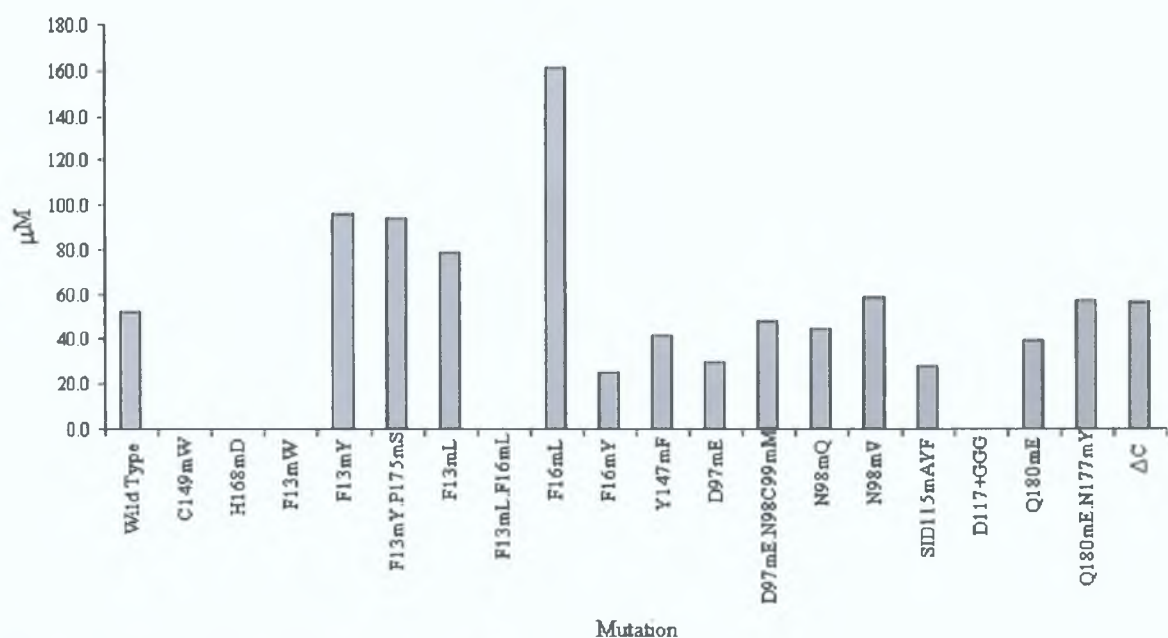


Figure 5.17 Effect of mutations on K_m
 Values for K_m (μM) taken from Table 5.2

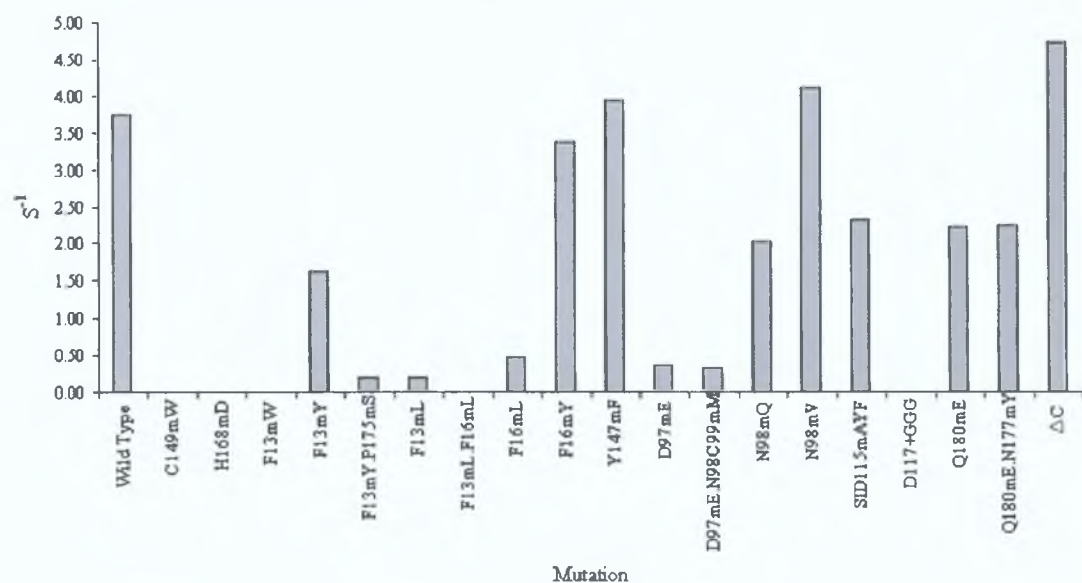


Figure 5.18 Effect of mutations on k_{cat}
 Values for k_{cat} (s^{-1}) taken from Table 5.2

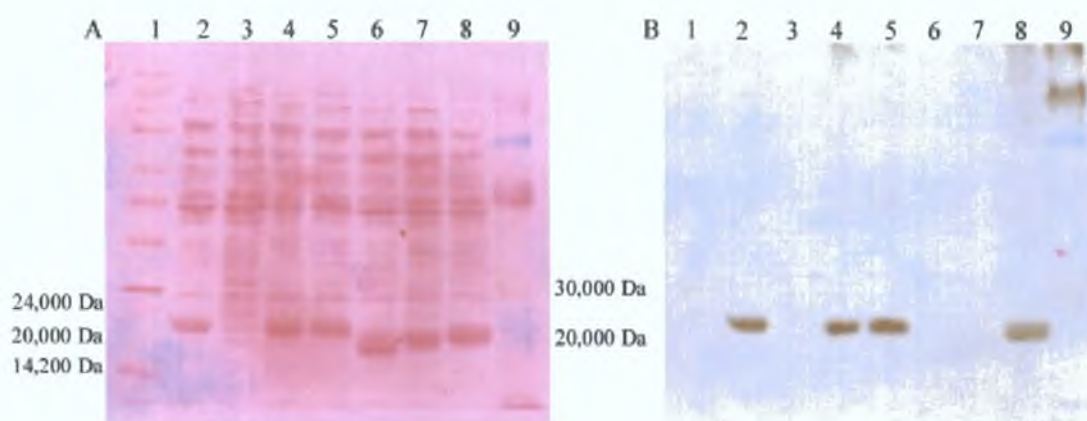


Figure 5.19 Western blot analysis of *rHsaPAP1* mutants
 Western blot (Section 2.18) analysis, using anti-His antibodies, of cleared lysate sample from *E. coli* XL10-Gold cultures expressing (Section 2.12) *rHsaPAP1* mutants (Table 5.1). (A): Ponceau S transfer. (B): Western blot. Lane 1, M_r (sizes given in Figure 2.6); Lane 2, *rHsaPAP1*_{6H}; Lane 3, pPC223 (–ve control); Lane 4, *rHsaPAP1*_C149mW_{6H}; Lane 5, *rHsaPAP1*_N98mV; Lane 6, *rHsaPAP1*_ΔC; Lane 7, *rHsaPAP1*; Lane 8, *rHsaPAP1*_ΔC_{6H}; Lane 9, CM_r (sizes given in Figure 2.7).

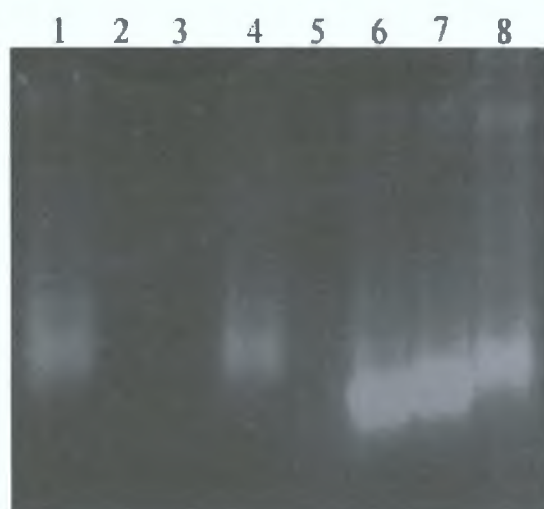


Figure 5.20 Zymogram analysis of *rHsaPAP1* mutants
 Zymogram analysis (Section 2.17) of cleared lysate samples from *E. coli* XL10-Gold cultures expressing (Section 2.12) *rHsaPAP1* mutants (Table 5.1). Lane 1, *rHsaPAP1*_{6H}; Lane 2, pPC223 (–ve control); Lane 3, *rHsaPAP1*_C149mW_{6H}; Lane 4, *rHsaPAP1*_N98mV; Lane 5, *rHsaPAP1*_F13mW_{6H}; Lane 6, *rHsaPAP1*_ΔC; Lane 7, *rHsaPAP1*; Lane 8, *rHsaPAP1*_ΔC_{6H}. The fluorescence bands correspond to the prominent protein bands visible after coomassie staining (data not shown).

5.3 Discussion

To date, no protein crystal structure has been published for human or any other eukaryotic PAP1. Although some preliminary work was done, as part of these studies on *HsaPAP1*, in an attempt to crystallise the protein (Section 4.3) to a quality fit for X-ray diffraction, this work is still ongoing. However, the availability of four prokaryotic PAP1 crystal structures (Section 1.3.10), together with the considerable homology observed between *HsaPAP1* and various prokaryotic PAP1 sequences, e.g. *BamPAP1*

(Figures 1.6 and 5.3), gave the impetus to attempt 3D homology modelling using the SWISS-MODEL server (Section 5.1).

SWISS-MODEL successfully predicted a 3D model for *Hsa*PAP1 by taking three of the existing prokaryotic PAP1 structures as templates (Figure 5.1). The *Hsa*PAP1 model exhibits a very similar folding pattern to the template structures, with α -helices and β -sheets in comparable locations along the primary sequence (Figure 5.2). As shown in Figure 5.4, the orientation of the catalytic triad (Glu-Cys-His, Section 1.3.7) of the prokaryotic structures is identical to the corresponding, conserved residues (Glu85, Cys149 and His168) in the *Hsa*PAP1 3D model. Mutation of the residues Cys149 and His168 individually (Section 5.1.1) in *Hsa*PAP1_{6H} resulted in complete loss of catalytic activity, strongly implicating them in the catalytic triad of *Hsa*PAP1 and supporting the use of the *Hsa*PAP1 3D model (Figure 5.1) as a reference structure to the mutational analysis in this work.

Figure 5.14 demonstrates the size difference between *rHsa*PAP1_{6H} and *rHsa*PAP1 Δ C, the mutant having nine C-terminal residues deleted and lacking a His-tag. The lack of a His-tag in *rHsa*PAP1 Δ C, as well as *rHsa*PAP1 (Figure 3.13), was confirmed by Western blot analysis (Figure 5.19) using anti-His antibodies. A considerable (24 residue) C-terminal domain of *Hsa*PAP1 was not incorporated into the 3D model. The most likely reason this domain was excluded is probably due to the observed reduced homology, between *Hsa*PAP1 and the prokaryotic sequences (Figure 5.3). In the prokaryotic PAP1 enzymes this C-terminal region contributes greatly toward interactions between monomeric subunits resulting in a tetrameric arrangement, as described in detail in Section 1.3.10.1. Human and other eukaryotic PAP1 enzymes are known to be monomeric (Section 1.3.9), hence one might not expect to find these specific C-terminal features necessary for quaternary structure formation.

The mutation of Gln180 (Section 5.2.3), caused only a slight reduction in specific activity and turnover. Dramatically, it was shown that removal of nine residues from the C-terminal of *Hsa*PAP1_{6H} did not affect binding capacity and even seemed to cause an increase in specific activity and turnover. Furthermore, the addition of six residues to the C-terminal (the His-tag, Section 3.4), a mutation in its own right, did not affect the catalytic properties of the enzyme (Section 4.2). It seems, therefore, that some residues of the C-terminal region do not have any functional necessity in *Hsa*PAP1 and modification of this region could make the enzyme more catalytically efficient. As outlined in Section 1.3.10.1, this region features in the formation of a quaternary

structure in prokaryotic PAP1 enzymes, which is not observed in eukaryotes (Section 1.3.9).

By co-introduction of diagnostic restriction sites, using silent mutations, each mutant construct could be verified (Figure 5.8) prior to commercial DNA sequencing. However, in light of the discussion in Section 3.8, it must be noted that silent mutations may change mRNA conformation, which could result in altered mRNA stability and/or translation efficiency. Therefore, based on the work carried out here, a possible alternative approach to identify mutations prior to DNA sequencing would be to further develop the gradient PCR screening strategy, as described in Section 3.6.2 (Figure 3.28).

As reported in Section 5.2.4, the *rHsaPAP1*_{6H} mutant enzymes could be successfully expressed at levels comparable to *rHsaPAP1*_{6H} (Figure 5.12), with the exception of *rHsaPAP1*_D117+GGG_{6H} (Table 5.1). No protein band of expected molecular weight 24,277 was detected for that mutant. RT-PCR analysis (Figure 5.13) shows that the amount of mRNA transcript produced was equivalent to that for *rHsaPAP1*_{6H}. It may be that the mutant mRNA structure folded in a way which impeded efficient translation. Alternatively, the addition of the three Gly residues may have dramatically altered the protein structure, resulting in an unstable folding and subsequent degradation of the protein.

The catalytically active mutants were purified (Figure 5.15) and their kinetic parameters determined (Table 5.2; Figures 5.16 to 5.18). As outlined in Section 5.2.1, mutagenesis of residues Phe13, Phe16 and Tyr147 was carried out to determine to what extent these residues contribute to the substrate binding pocket of *HsaPAP1* (Figure 5.7). A similar study of these binding pocket residues in *BamPAP1* was carried out by Ito *et al.* (2001), as reported in Section 1.3.11. This *BamPAP1* study identified the Phe10 residue (corresponding to Phe13 in *HsaPAP1*) as an essential contributor to the binding pocket. In the present study, changing *HsaPAP1* Phe13 to the larger, aromatic Trp resulted in complete loss of activity. Substitution of Phe13 with the aromatic, yet polar Tyr caused a reduction of specific activity and rate of turnover (k_{cat}), in correlation with an increase in K_m . The unintentional co-mutation of Pro175 to Ser (Table 5.1) caused a significant decrease in specific activity and k_{cat} , while not further affecting the binding capacity (K_m). The substitution of Phe16 to Tyr caused a more subtle reduction in specific activity and k_{cat} (compared to Phe13), while increasing the binding capacity, characterised by a drop in K_m . The changing of Tyr147 to Phe, creating the triple Phe

arrangement seen in many prokaryotic PAP1 enzymes (Figure 1.4), did not affect the kinetic properties to any notable extent. The individual substitution of both Phe13 and Phe16 with the smaller, hydrophobic Leu caused dramatic reduction in specific activity and k_{cat} and a corresponding drop in binding capacity (rise in K_m). When both these mutations were imposed simultaneously, activity was lost completely. These results are in correlation with the *Bam*PAP1 study. It is assumed these residues contribute similarly to the substrate recognition mechanism in *Hsa*PAP1. Their similar positioning within the *Hsa*PAP1 3D model as the *Bam*PAP1 residues (Figure 5.21) suggests that the model gives an accurate prediction of the *Hsa*PAP1 substrate binding pocket structure.

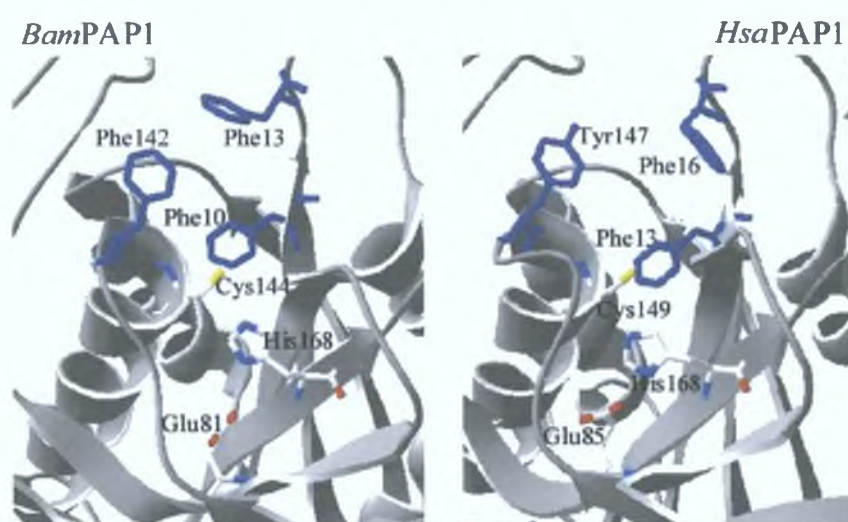


Figure 5.21 Comparison of *Hsa*PAP1 and *Bam*PAP1 binding pockets

The side-chain orientation of the *Bam*PAP1 substrate binding pocket residues Phe10, Phe13 and Phe142, (indicated in blue), compared to the corresponding *Hsa*PAP1 3D model residues Phe13, Phe16 and Tyr147. Catalytic triad residues (Figure 5.4) are represented in CPK colour scheme (C = white, O = red, N = blue and S = yellow). For amino acid information see Appendix D. Generated using DeepView (Section 2.11).

The mutation of Asp97 caused a significant decrease in specific activity and k_{cat} , while not affecting binding capacity greatly. Changing Asn98 to Gln reduced the specific activity and turnover slightly, while the change to Val did not cause any effect. Mutation of Ser115-Ile116-Asp117 to A-Y-F (Section 5.2.2), a motif found in prokaryotes (Figure 1.4), on an extended external loop (Figure 5.7) caused a slight reduction in catalytic efficiency.

The loss of activity caused by the mutation of Ile120 to Gly implicates this residue to be in a structurally significant location. In the *Hsa*PAP1 3D model Ile120 is located at the N-terminal of an external α -helix (Figure 5.7). In the prokaryotic primary sequences this residue is conservatively substituted with Leu (Figure 5.6) and similarly located at the

N-terminal of a corresponding α -helix (data not shown). This lends further support for the *HsaPAP1* 3D model.

5.3.1 Structure-based drug design

The conventional process by which the pharmaceutical industry facilitates the discovery of new drugs involves the screening of large (ever increasing) libraries of chemical compounds in search for those with a particular biological or medicinal relevance (Malhotra and Taqueti, 2000). The anticipation is that a sufficiently large collection of molecules will harbour enough structural diversity to prove effective for any particular biological target of interest. Compounds that are found to produce the desired biological effect (lead compounds) are then tested and optimised for their medicinal properties. Despite that many drugs have been successfully developed using this approach, the time-, money- and labour-intensity can often prove prohibitive unless the reward is sufficiently lucrative. Many previous efforts to minimise the time and resources necessary for such high-throughput screening (HTS) only resulted in compromised accuracy and sensitivity of the screenings.

The identification of lead compounds is the most critical stage in the process of modern drug discovery. Structure-based strategies have become an integral part of this stage. This owes in particular to the volume and pace at which the structures of proteins (many of them important drug targets) and their co-crystals are presently becoming available, together with the growing sophistication of computational bio-informatics tools. Structural analysis and design methodologies are increasingly being used for the determination of new lead compounds and molecular scaffold generation as well as the optimisation of receptor or enzyme affinity. Such approaches show many advantages over traditional HTS, including reduced reagent storage, handling of large compound libraries, lowering of false positives often associated with HTS and the ability to find low-molecular-weight leads even when HTS fails. The structure-based approach to drug design relies on a firm understanding of the molecular recognition between a protein and its binding ligand, as well as knowledge of the structure-activity relationships (Andricopulo and Montanari, 2005).

Here follows a case study of recently published findings by a multi-disciplinary research collaboration (Janssen *et al.*, 2005; Das *et al.*, 2005) demonstrating how extensive knowledge of crystallography, protein structure analysis and molecular modelling plays a vital role in modern drug discovery. In this research effort a new group of highly

potent novel anti-HIV drugs has been developed. These drugs will contribute to evermore effective therapies against the acquired immunodeficiency syndrome (AIDS) caused by this virus. The global pandemic of HIV/AIDS is one of the most significant medical challenges facing humanity today. The death toll to date stands at 23 million, with more than 42 million people presently infected.

Virally encoded enzymes that coordinate the replication of the virus, e.g. reverse transcriptase (RT), protease and integrase, are the primary conventional protein targets of HIV therapy. The therapeutic aim is to use inhibitors against these enzymes in an effort to suppress replication of the virus. The virus has a fast turnover rate, undergoing one replication approximately every two days. This, together with a high rate of mutation due to the lack of an efficient error correction mechanism during replication, results in the rapid emergence of enzyme variants that are resistant to current anti-HIV drugs. Such mutant enzymes, while still capable of carrying out normal HIV functions, owe their inhibitor resistance to subtle structural changes. The treatment of AIDS patients with combinations or “cocktails” of anti-HIV drugs can often lead to multi-drug-resistant variants. These resistant varieties are constantly being selected for by the use of these drugs and therefore dominate the viral population pool. The daunting challenge in HIV therapy is the design of drugs which retain their usefulness in the face of this “moving target” nature of the virus. Successful therapy must anticipate changes in the structure of these protein targets.

HIV-1 RT is a heterodimer, having one subunit containing polymerase and RNase H domains and another that contributes structural support. One widely used class of anti-HIV drugs, lead compounds from a HTS study, are non-nucleoside reverse transcriptase inhibitors (NNRTIs). They specifically inhibit HIV-1 replication, at nanomolar concentration, by binding to HIV-1 RT. The location of the binding pocket for these inhibitors is in close proximity to the polymerase catalytic site. RT undergoes conformational changes in correlation with various stages of its function. Due to this dynamic nature of HIV-1 RT structure the NNRTI binding pocket is transitional and only truly exists when the inhibitor is bound. Upon NNRTI binding further conformational changes are obstructed and the process of polymerisation is inhibited.

The researchers carried out in-depth analysis of the crystal structures of wild-type and mutant HIV-1 RT, in complex with nucleic acid substrates as well as several established NNRTIs. This provided a detailed overview of how substrate and/or inhibitor binding correlated with different functional conformations of HIV-1 RT, as well as giving

insight as to how specific mutations might prevent NNRTIs from binding: While NNRTI binding does not block substrate binding, rather it prevents the chemical process of polymerisation. The NNRTI binding pocket is of a hydrophobic nature, conferred by several aromatic and hydrophobic amino acid side-chains. Binding of NNRTIs into the pocket is mediated by various hydrophobic interactions and hydrogen bonds. NNRTIs exhibit considerable variation in chemical composition and size and in consequence they each exhibit a unique interaction profile with the binding pocket. HIV-1 RT mutations conferring resistance represented specific side-chain substitutions in and around the NNRTI binding pocket, supporting their key role in the relevant NNRTI binding. The structural mechanism of drug resistance is accomplished by altered size, shape, polarity and accessibility of different regions in the binding pocket.

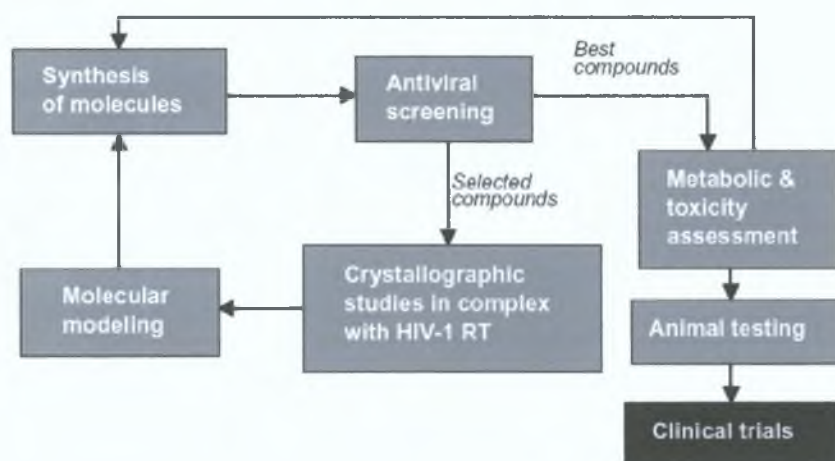


Figure 5.22 Structure-based design strategy for novel anti-HIV drugs

A flow chart highlighting the multi-disciplinary design, optimisation and assessment strategy employed in the discovery of novel HIV-1 RT inhibitors (Taken from Das *et al.*, 2005).

A design strategy was initiated, in ultimate pursuit of a compound that could bind into the NNRTI binding pocket of HIV-1 RT despite various mutational structure changes (Figure 5.22). Derivatives of various NNRTIs were synthesised and investigated for their inhibition profiles. The crystal structures of HIV-1 RT/inhibitor complexes were solved and the ligand binding modes determined. Molecular modelling was carried out, involving the docking of inhibitor compounds into the binding pockets of the various crystal structures, to evaluate and analyse probable binding properties of further derivative compounds. The measured potencies of the inhibitors against mutated HIV-1

RT isolates were correlated with the structure/modelling data and the results were filtered back into the design and development process.

As the cycles of this design strategy gradually progressed, new inhibitors were identified which had altered binding conformation to the compounds from which they were originally derived. These altered binding profiles also resulted in retained effectiveness against key NNRTI-resistant mutants, due to reduction/loss of steric conflict caused by mutant amino acid side-chains. As the process continued, the newer compounds demonstrated potency against a broader spectrum of drug resistant HIV-1 strains. Eventually a new class of NNRTI, the diarylpyrimidine (DAPY) compounds, was identified.

It was noticed from the crystal structures of the DAPY compounds in complex with HIV-1 RT, that their electron densities were weak relative to the first generation of NNRTIs. This was attributed to positional disorder within the binding pocket. This raised speculation that these compounds were able to bind HIV-1 RT in more than one conformation. Such conformational heterogeneity would also explain the low crystal resolution that was obtained for these complexes. This theory was confirmed using HIV-1 RT variants having reduced binding possibilities and therefore less heterogeneous crystals. A much better crystal resolution was achieved. Molecular modelling analysis confirmed that the DAPY compounds are able to undergo conformational changes and repositioning within the binding pocket. Molecular dynamics calculations show that the DAPY compounds exhibit significant torsional flexibility without encountering prohibitive energy barriers.

In summary, this HIV research collaboration has combined several structure-based methodologies, creating a powerful strategy to successfully address the need for sophisticated novel anti-HIV drugs. The ability to bind in multiple conformations permits the DAPY compounds to retain activity against HIV-1 RT despite structural variation, essentially compensating for the effects of resistance mutations. A single one of these drugs is able to perform the same task as a “cocktail” of first generation NNRTIs. The use of these drugs could provide an environment where the manifestation of structural variety does not impart to HIV-1 RT any evolutionary benefit, thereby halting the emergence of drug-resistant variants. As such, this case study is an inspiring demonstration of how structure data can be exploited to serve the process of drug design.

In this thesis, a structure model of *Hsa*PAP1 was established (Section 5.1). The validity of this 3D model is supported by the results of mutational studies presented in Section 5.1.1. The *Hsa*PAP1 model was used to identify residues and domains, whose role in the structure-activity properties of the enzyme were then determined by further mutational and catalytic studies (Section 5.2). Preliminary crystallisation studies on *Hsa*PAP1 were conducted as outlined in Section 4.3. So far, the crystal diffraction patterns to emerge from this work have not been of sufficient quality to allow elucidation of the protein structure. However, this work is ongoing within the research group and it is hoped that the true *Hsa*PAP1 structure will soon be solved.

The involvement of PAP1 in the degradation of biologically active pGlu neuropeptides, such as TRH, has been widely documented (Section 1.3.1). Furthermore, a notable physiological correlation with levels of the pharmacologically active compounds cyclo(His-Pro) and free pGlu have implicated PAP1's involvement in several medical conditions such as Huntington's disease, pyroglutamic acidemia, motor neuron disease and Alzheimer's disease (Section 2.1). To date, substrate specificity studies of PAP1 and identification of inhibitors have focused mainly on using synthetic chemical derivatives of the pGlu substrate moiety, evaluated by biochemical and kinetic analysis (Section 1.3.8). The *in vivo* suitability of those compounds found to have inhibitory potential has yet to be ascertained.

Crystal structures of the various *Hsa*PAP1 mutants created in this project and those that have been proposed as future work (Chapter 6), as well as possible co-structures, will subsequently be sought. If such a set of structures can be accumulated, it will be possible to initiate an extensive analysis and comparison project, including other data such as the prokaryotic PAP1 structures and co-structures reviewed in Section 1.3. The outcome of such an undertaking would be an extremely comprehensive knowledge of *Hsa*PAP1's structural and catalytic profile. This knowledge base would serve as a foundation on which *Hsa*PAP1 inhibitor design projects could be built, studies that could draw on a variety of molecular modelling and analysis tools combined with the type of genetic and biochemical methodologies employed in this thesis. Successful design of new potent *Hsa*PAP1 inhibitors, suitable for *in vivo* application, would represent valuable drugs that can be employed in future pharmacological studies of human PAP1.

6.0 Summary & Recommendations

In this research thesis, the gene for human PAP1 was cloned into a bacterial expression system, enabling high level expression and purification of the recombinant enzyme (Chapter 3). An extensive characterisation of the recombinant PAP1 was undertaken (Chapter 4). The biochemical and kinetic profiles of the recombinant enzyme are comparable, and contribute significantly, to data on mammalian PAP1 reported in the literature (Cummins and O'Connor, 1998).

With the aid of homology modelling, a series of mutants were constructed and subsequently analysed (Chapter 5). The data revealed in this work contributes significant insight to the structural properties of human PAP1, in particular structure-function relationships of the active site region. Together with further data that may be determined by a number of recommended investigations (see below), these findings will open the way to a possible structure-based approach to PAP1 inhibitor design (as discussed in Section 5.3.1).

Some additional studies on recombinant human PAP1 that could provide more clarity on the relationship between the structure and function are hereby outlined:

In a similar way as was used in this study, the predicted 3D model of human PAP1 (Section 5.1) could provide a structural reference for the design of further site-specific mutagenesis work. The next phase should include an in-depth analysis of the catalytic site and hydrophobic binding pocket. In this work, the residues Cys149 and His168 of the assumed catalytic triad were substituted, leading to loss of activity (Section 5.1.1). As shown in the crystal structures of several prokaryotic PAP1 enzymes, the Glu residue of the catalytic triad serves to orientate the His imidazole ring (Section 1.3.10.2). Thus it would be interesting to investigate the effect that a conservative substitution of the Glu85 residue would have in the human PAP1. In addition to this, further residues which are located in hydrophobic pocket should be investigated. An increase or reduction of the hydrophobicity of this region, as well as variation of side-chain size, will likely have a measurable effect on the binding properties and substrate specificity of the enzyme. This could be correlated with a hydrolysis study, using various known PAP1 substrates, as well as investigating the inhibition caused by active site-directed inhibitors of PAP1 (Section 1.3.8).

The nine C-terminal residues of recombinant human PAP1 have no apparent necessity in enzyme function. Indeed their removal seemed to make the enzyme more catalytically efficient (Section 5.3). It would be of interest to determine what effect further deletion at the C-terminal would have on the enzyme. Furthermore, given that

the C-terminal region of the prokaryotic PAP1 enzymes plays a role in quaternary structure formation (Section 1.3.10.1), it would be interesting to construct a hybrid enzyme, having a prokaryotic C-terminal domain. Would such a hybrid protein assume a quaternary arrangement and, if so, what kinetic properties would this multimer have?

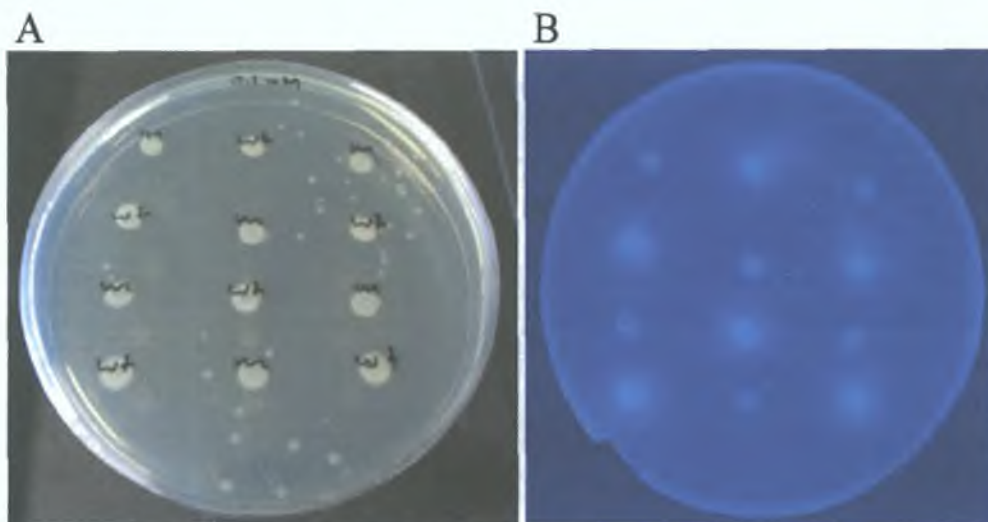


Figure 6.1 Development of a fluorescence plate assay for PAP1 activity

A suitable fluorescence plate assay was developed for the detection of *E. coli* expressing PAP1 activity. *E. coli* XL10-Gold, harbouring expression plasmids for active wild-type (pRV5, Section 3.5) and inactive mutant (pRV5_C149mW, Section 5.1.1) His-tagged recombinant human PAP1 were used as positive and negative controls respectively. Cells from 1 ml of culture (A_{600} of 1.0) were re-suspended in 100 μ l dH₂O. 1 μ l 'spots' were applied to selective LB agar containing 100 μ M IPTG. After overnight incubation at 37°C (A) the plate assay was carried out as described in Section 2.15.4. Fluorescence can be clearly seen (B) for the wild-type (wt) enzyme compared to the mutant (m).

Site-specific mutagenesis relies minimally on knowing the amino acid sequence but in practice becomes powerful as a tool when either structural and/or comparative homology data is available. Given this, one needs premonition of how an amino acid substitution might affect the structure and/or function of an enzyme. A limitation of this site-specific mutagenesis approach is that one can still never have absolute foresight of the outcome. To complement the work carried out in this study, an alternative, random mutagenesis approach, not based on the 3D modelling was considered. Initial studies were carried out on the PAP1 gene using an error-prone PCR approach as described by Nakaniwa *et al.* (2004) which includes using varying Mn^{2+} concentrations. In order to select potentially interesting mutants (using the random mutagenesis approach) one would need to devise a highly sensitive assay for screening large numbers of colonies. A fluorescent plate assay was developed, which can differentiate PAP1 active from inactive colonies. The methodology of this assay is described in Section 2.15.4. Figure 6.1 shows a demonstration of this assay using positive and negative controls for PAP1

activity. One can clearly see the fluorescence from the 'spots' of *E. coli* expressing the active form of human PAP1 compared to an inactive mutant. The assay was also very effective in differentiating between single colonies, unfortunately these images were of a poor quality and are not presented here.

Only preliminary progress was made, regarding the random mutagenesis, by the time this thesis was submitted (data not shown). In summary, specific primers PAPHsK and PAPHsS (Table 2.2) were used to PCR amplify the human PAP1 gene from the plasmid construct pRV5 (Figure 3.8) under conditions of various MnSO_4 concentrations up to 100 μM . The amplified gene was re-inserted to the vector backbone, transformed into *E. coli* XL10-Gold and then the transformation plates were screened using the fluorescent assay. The higher the MnSO_4 concentration used, the higher the percentage of non-fluorescent colonies obtained. These preliminary findings indicate that mutations, which produce an inactive PAP1, can be obtained using this approach. Furthermore it would be worthwhile to use the gene of an inactive PAP1 mutant (or one with reduced activity) as template in the random mutagenesis reaction, in an attempt to regain (or increase) activity.

Finally, this work now provides some useful tools for the continued research of human PAP1 and its *in vivo* role. The clone presents a probe, which can be used in various tissue and array-hybridisation studies and to investigate possible tissue-specific forms and expression of the enzyme. Also, antibodies can be raised against the purified recombinant protein and subsequently used in immuno-histochemical investigations and in applications such as high-resolution tissue microscopy coupled with immuno-cytochemical amplification.

References

-A-

Abe, K., Fukada, K. and Tokui, T. (2004) Marginal Involvement of Pyroglutamyl Aminopeptidase I in Metabolism of Thyrotropin-Releasing Hormone in Rat Brain. *Biological & Pharmaceutical Bulletin* **27**: 1197-1201

Abe, K., Saito, F., Yamada, M. and Tokui, T. (2004) Pyroglutamyl Aminopeptidase I, as a Drug Metabolizing Enzyme, Recognises Xenobiotic Substrates Containing L-2-Oxothiazolidine-4-carboxylic Acid. *Biological & Pharmaceutical Bulletin* **27**: 113-116

Abe, K., Watanabe, N., Kosaka, T., Yamada, M., Tokui, T. and Ikeda, T. (2003) Hydrolysis of Synthetic Substrate, L-Pyroglutamyl p-Nitroanilide is Catalyzed Solely by Pyroglutamyl Aminopeptidase I in Rat Liver Cytosol. *Biological & Pharmaceutical Bulletin* **26**: 1528-1533

Albert, Z. and Szewczuk, A. (1972) Pyrrolidonyl peptidase in some avian and rodent tissues. Histochemical localization and biochemical studies. *ACTA Histochemica* **44**: 98-105

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) Basic Local Alignment Search Tool. *Journal of Molecular Biology* **215**: 403-410

Alur, H.H., Desai, R.P., Mitra, A.K. and Johnston, T.P. (2001) Inhibition of a model protease - pyroglutamate aminopeptidase by a natural oligosaccharide gum from *Hakea gibbosa*. *International Journal of Pharmaceutics* **212**: 171-176

Andricopulo, A.D. and Montanari, C.A. (2005) Structure-activity relationships for the design of small-molecule inhibitors. *Mini Reviews in Medicinal Chemistry* **5**: 585-593

Aoyagi, T., Hatsu, M., Imada, C., Naganawa, H., Okami, Y. and Takeuchi, T. (1992b) Pyrizinostatin: A New Inhibitor Of Pyroglutamyl Peptidase. *The Journal of Antibiotics* **45**: 1795-1796

Aoyagi, T., Hatsu, M., Kojima, F., Hayashi, C., Hamada, M. and Takeuchi, T. (1992a) Benarthin: A New Inhibitor Of Pyroglutamyl Peptidase I. Taxonomy, Fermentation, Isolation and Biological Activities. *The Journal of Antibiotics* **45**: 1079-1083

Armentrout, R.W. and Doolittle, R.F. (1969) Pyrrolidonecarboxyl Peptidase: Stabilization and Purification. *Archives of Biochemistry and Biophysics* **132**: 80-90

Armentrout, R.W.(1969) Pyrrolidonecarboxyl Peptidase from Rat Liver. *Biochimica et Biophysica ACTA* **191**: 756-759

Awadé, A., Cleuziat, Ph., Gonzales, Th. and Robert-Baudouy, J. (1994) Pyrrolidone Carboxyl Peptidase (Pcp): An Enzyme That Removes Pyroglutamic Acid (pGlu) From pGlu-Peptides and pGlu-Proteins. *Proteins: Structure, Function and Genetics* **20**: 34-51

Awadé, A., Cleuziat, Ph., Gonzalès, Th. and Robert-Baudouy, J. (1992) Characterisation of the pcp gene encoding the pyrrolidone carboxyl peptidase of *Bacillus subtilis*. *Federation of European Biochemical Societies* **305**: 67-73

Awadé, A., Gonzalès, Th., Cleuziat, Ph. and Robert-Baudouy, J. (1992) One step purification and characterisation of the pyrrolidone carboxyl peptidase of *Streptococcus pyogenes* over-expressed in *Escherischia coli*. *Federation of European Biochemical Societies* **308**: 70-74

-B-

Bairoch, A. and Apweiler, R. (1996) The SWISS-PROT protein sequence data bank and its new supplement TREMBL. *Nucleic Acids Research* **24**: 21-25

Baneyx, F. (1999) Recombinant protein expression in *Escherichia coli*. *Current Opinion in Biotechnology* **10**: 411-421

Barrett, A.J. and Rawlings, N.D. (2001) Evolutionary Lines of Cysteine Peptidases. *Biological Chemistry* **382**: 727-733

Barrett, A.J., Rawlings, N.D. and Woessner, J.F. (1998) Handbook of Proteolytic Enzymes. Academic Press, London

Bauer, K. and Kleinkauf, H. (1980) Catabolism of Thyroliberin by Rat Adenohypophyseal Tissue Extract. *Journal of Biochemistry* **106**: 107-117

Benson, D.A., Boguski, M., Lipman, D.J. and Ostell, J. (1996) GenBank. *Nucleic Acids Research* **24**: 1-5

- Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N. and Bourne, P.E. (2000) The Protein Data Bank. *Nucleic Acids Research* **28**: 235-242
- Besson, C., Vessillier, S., Gonzalès, T., Saulnier, J. and Wallach, J. (1994) Conductimetric assay of pyroglutamyl peptidase activity. *Analytica Chimica ACTA* **294**: 305-309
- Bharadwaj, D., Roy, M.S., Saha, G. and Hati, R.N. (1992) Pyroglutamate aminopeptidase in rat submaxillary gland. *Indian Journal of Biochemistry* **29**: 442-444
- Birnboim, H.C. and Doly, J. (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Research* **7**: 1513-1523
- Blum, H., Beir, H. and Gross, H.J. (1987) Improved silver staining of plant proteins, RNA and DNA in polyacrylamide gels. *Electrophoresis* **8**: 93-99
- Bradford M.M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry* **72**: 248-254
- Brandl, M., Meyer, M. and Suhnel, J. (1998) *Ab initio* study of a water-mediated uracil cytosine base pair. *Poster*
- Browne, P. and O'Cuinn (1983) An evaluation of the role of a pyroglutamyl peptidase, a post-proline cleaving enzyme and a post-proline dipeptidyl amino peptidase, each purified from the soluble fraction of guinea-pig brain, in the degradation of thyroliberin *in vitro*. *European Journal of Biochemistry* **137**: 75-87
- Brunetti, L., Cacciatore, I., Di Stefano, A., Dupre, S., Giorgi, A., Luisi, G., Michelotto, B., Orlando, G., Pinnen, F., Recinella, L., Sozio, P. and Spirito, A. (2002) Synthesis and biological evaluation of a novel pyroglutamyl-modified TRH analogue. *Il Farmaco* **57**: 479-486
- Bujard, H., Gentz, R., Lanzer, M., Stueber, D., Mueller, M., Ibrahimi, I., Haeuptle, M-T., Dobberstein, B. (1987) A T5 promoter-based transcription-translation system for the analysis of proteins *in vitro* and *in vivo*. *Methods In Enzymology* **155**: 416-433

Busby, Jr., W.H., Youngblood, W.W. and Kizer, J.S. (1982) Studies of Substrate Requirements, Kinetic Properties and Competitive Inhibitors of the Enzymes Catabolizing TRH in Rat Brain. *Brain Research* **242**: 261-270

-C-

Capecchi, J.T. and Loudon, G.M. (1985) Substrate Specificity of Pyroglutamylaminopeptidase 1. *Journal of Medicinal Chemistry* **28**: 140-143

Charli, J.L., Mendez, M., Joseph-Bravo, P. and Wilk, S. (1987) Specific Inhibitors Of Pyroglutamyl Peptidase I And Prolyl Endopeptidase Do Not Change The *In Vitro* Release Of TRH Or Its Content In Rodent Brain. *Neuropeptides* **9**: 373-378

Chomczynski, P. and Sacchi, N. (1987) Single-Step Method of RNA Isolation by Acid Guanidinium Thiocyanate-Phenol-Chloroform Extraction. *Analytical Biochemistry* **162**: 156-159

Clarke, P. (2000) PhD Thesis, Dublin City University, Ireland

Cleuziat, P., Awadé, A. and Robert-Baudouy, J. (1992) Molecular characterisation of pcp, the structural gene encoding the pyrrolidone carboxyl peptidase from *Streptococcus pyogenes*. *Molecular Microbiology* **6**: 2051-2063

Corpet, F. (1988) Multiple Sequence Alignment with Hierarchical clustering. *Nucleic Acids Research* **16**: 10881-10890

Cudney, B. and Patel, S. (1994) Screening and Optimization Strategies for Macromolecular Crystal Growth. *ACTA Crystallographica* **D50**: 414-423

Cummins, P.M. and O'Connor, B. (1996) Bovine Brain Pyroglutamyl Aminopeptidase (Type-1): Purification and Characterisation of a Neuropeptide-Inactivating Peptidase. *The International Journal of Biochemistry and Cell Biology* **28**: 883-893

Cummins, P.M. and O'Connor, B. (1998) Pyroglutamyl peptidase: an overview of the three known enzymatic forms. *Biochimica et Biophysica ACTA* **1429**: 1-17

Dando, P.M., Fortunato, M, Strand, G.B., Smith, T.S. and Barrett, A.J. (2003) Pyroglutamyl-peptidase I: cloning, sequencing, and characterisation of the recombinant human enzyme. *Protein Expression and Purification* **28**: 111-119

Das, K., Lewi, P.J., Hughes, S.H., Arnold, E. (2005) Crystallography and the design of anti-AIDS drugs: conformational flexibility and positional adaptability are important in the design of non-nucleoside HIV-1 reverse transcriptase inhibitors. *Progress in Biophysics and Molecular Biology* **88**: 209-231

De Boer, H.A., Comstock, L.J. and Vasser, M. (1983) The *tac* promoter: A functional hybrid derived from the *trp* and *lac* promoters. *Biochemistry* **80**: 21-25

Dealler, S.F., Campbell, L., Kerr, K.G., McGoldrick, J., Flannigan, K.A. and Hawkey, P.M. (1989) Reliable Five-Minute Method for Identification of *Streptococcus pyogenes*. *European Journal of Clinical Microbiology & Infectious Diseases* **8**: 308-310

DeGandarias, J.M. De, Casis, O., Echevarria, E., Irazusta, J. and Casis, L. (1992) Pyroglutamyl-peptidase I activity in the cortex of the cat brain during development. *International Journal of Developmental Biology* **36**: 335-337

DeGandarias, J.M., Irazusta, J., Fernandez, D., Varona, A. and Casis, L. (1994) Developmental Changes of Pyroglutamate-peptidase I Activity in Several Regions of the Female and Male Rat Brain. *International Journal of Neuroscience* **77**: 53-60

DeGandarias, J.M., Irazusta, J., Gil, J., Varona, A., Ortega, F. and Casis, L. (2000) Subcellular ontogeny of brain pyroglutamyl peptidase I. *Peptides* **21**: 509-517

DeGandarias, J.M., Irazusta, J., Silio, M., Gil, J., Saitua, N. and Casis, L. (1998) Soluble and membrane-bound pyroglutamyl-peptidase I activity in the developing cerebellum and brain cortex. *International Journal of Developmental Biology* **42**: 103-106

Doolittle, R.F. and Armentrout, R.W. (1968) Pyrrolidonyl Peptidase. An Enzyme for Selective Removal of Pyrrolidonecarboxylic Acid Residues from Polypeptides. *Biochemistry* **7**: 516-521

Doran, J.D. and Carey, P.R. (1996) α -Helix Dipoles and Catalysis: Absorption and Raman Spectroscopic Studies of Acyl Cysteine Proteases. *Biochemistry* **35**: 12495-12502

-E-

Exterkate, F.A. (1977) Pyrrolidone Carboxyl Peptidase in *Streptococcus cremoris*: Dependence on an Interaction with Membrane Components. *Journal of Bacteriology*. **129**: 1281-1288

-F-

Faivre-Bauman, A., Loudes, C., Barret, A., Tixier-Vidal, A. and Bauer, K. (1986) Possible Role Of Neuropeptide Degrading Enzymes On Thyroliberin Secretion In Fetal Hypothalamic Cultures Grown In Serum Free Medium. *Neuropeptides* **7**: 125-138

Falkous, G., Shaw, P.J., Ince, P.G. and Mantle, D. Comparison of cytoplasmic and lysosomal proteolytic enzyme levels in spinal cord tissue from motor neurone disease and control cases. *6th International Symposium on ALS/MND*, 1995, Dublin, Ireland

Friedman, T.C., Kline, T.B. and Wilk, Sherwin (1985) 5-Oxoprolinal: Transition-State Aldehyde Inhibitor of Pyroglutamyl-Peptide Hydrolase. *Biochemistry* **24**: 3907-3913

Fujiwara, K. and Tsuru, D. (1978) New Chromogenic and Fluorogenic Substrates for Pyrrolidonyl Peptidase. *Journal of Biochemistry* **83**: 1145-1149

Fujiwara, K., Kitagawa, T. and Tsuru, D. (1981a) Inactivation of Pyroglutamyl Aminopeptidase by L-Pyroglutamyl Chloromethyl Ketone. *Biochimica et Biophysica ACTA* **655**: 10-16

Fujiwara, K., Kobayashi, R. and Tsuru, D. (1979) The Substrate Specificity Of Pyrrolidone Carboxyl Peptidase From *Bacillus Amyloliquefaciens*. *Biochimica et Biophysica ACTA* **570**: 140-148

Fujiwara, K., Matsumoto, E., Kitagawa, T. and Tsuru, D. (1981b) Inactivation of Pyroglutamyl Aminopeptidase by N^α-Carbobenzoxy-L-Pyroglutamyl Chloromethyl Ketone. *Journal of Biochemistry* **90**: 433-437

Fujiwara, K., Matsumoto, E., Kitagawa, T. and Tsuru, D. (1982) N- α -Carbobenzoxy Pyroglutamyl Diazomethyl Ketone As Active-Site-Directed Inhibitor For Pyroglutamyl Peptidase. *Biochimica et Biophysica ACTA* **702**: 149-154

-G-

Gonzalès, T. and and Awadé, A. (1992) Purification and characterisation of recombinant pyrrolidone carboxyl peptidase of *Bacillus subtilis*. *Journal of Chromatography* **584**: 101-107

Gonzalès, T. and Robert-Baudouy, J. (1994) Characterisation of the pcg Gene of *Pseudomonas fluorescens* and of Its Product, Pyrrolidone Carboxyl Peptidase (Pcp). *Journal of Bacteriology* **176**: 2569-2576

Gonzales, T. and Robert-Baudouy, J. (1996) Bacterial aminopeptidases: Properties and functions. *FEMS Microbiology Reviews* **18**: 319-344

Griffiths, E.C. (1985) Thyrotropin Releasing Hormone: Endocrine And Central Effects. *Psychoneuroendocrinology* **10**: 225-235

Guex, N. and Peitsch, M.C. (1997) SWISS-MODEL and the Swiss-Pdb Viewer: An environment for comparative protein modelling. *Electrophoresis* **18**: 2714-2723

-H-

Haitinger, L. (1882) Vorlauge mittheilung über glutaminsäure und pyrrol. *Monatshefte für Chemie* **3**: 228-229

Hanahan, D. (1985) Techniques for transformation of E. coli. In: DNA Cloning: a Practical Approach, pp. 109-135. Edited by D. M. Glover. Oxford: IRL Press

Hannig, G. and Makrides, S.C. (1998) Strategies for optimizing heterologous protein expression in *Escherichia coli*. *Trends in Biotechnology* **16**: 54-60

He, W. and Barrow, C.J. (1999) The A β 3-Pyroglutamyl and 11-Pyroglutamyl Peptides Found in Senile Plaque Have Greater β -Sheet Forming and Aggregation Propensities *in Vitro* than Full-Length A β . *Biochemistry* **38**: 10871-10877

Holmes, D.S. and Quigley, M.(1981) A Rapid Boiling Method for the Preparation of Bacterial Plasmids. *Analytical Biochemistry* **114**: 193-197

-I-

Inoue, H., Nojima, H. and Okayama, H. (1990) High efficiency transformation of *Escherichia coli* with plasmids. *Gene* **96**: 23-28

Ito, K., Inoue, T., Takahashi, T., Huang, H., Esumi, T., Hatakeyama, S., Tanaka, N., Nakamura, K.T. and Yoshimoto, T. (2001) The Mechanism of Substrate Recognition of Pyroglutamyl-peptidase I from *Bacillus amyloliquefaciens* as Determined by X-ray Crystallography and Site-directed Mutagenesis. *The Journal of Biological Chemistry* **276**: 18557-18562

-J-

Jackson, I.M.D. (1982) Thyrotropin-Releasing Hormone. *The New England Journal Of Medicine* **30**: 145-155

Jancarik, J. and Kim, S-H. (1991) Sparse matrix sampling: a screening method for crystallization of proteins. *Journal of Applied Crystallography* **24**: 409-411

Janssen, P.A., Lewi, P.J., Arnold, E., Daeyaert, F., de Jonge, M., Heeres, J., Koymans, L., Vinkers, M., Guillemont, J., Pasquier, E., Kukla, M., Ludovici, D., Andries, K., de Bethune, M.P., Pauwels, R., Das, K., Clark, A.D. Jr, Frenkel, Y.V., Hughes, S.H., Medaer, B., De Knaep, F., Bohets, H., De Clerck, F., Lampo, A., Williams, P., Stoffels, P. (2005) In search of a novel anti-HIV drug: multidisciplinary coordination in the discovery of 4-[[4-[[4-[(1E)-2-cyanoethenyl]-2,6-dimethylphenyl]amino]-2-pyrimidinyl]amino]benzonitrile (R278474, rilpivirine). *Journal of Medicinal Chemistry* **48**: 1901-1909

-K-

Kabashima, T., Li, Y., Kanada, N., Ito, K. and Yoshimoto, T. (2001) Enhancement of the thermal stability of pyroglutamyl peptidase I by introduction of an intersubunit disulfide bond. *Biochimica et Biophysica ACTA* **1547**: 214-220

Kaufhold, A., Lutticken, R. and Schwien, U. (1989) Few-Minutes Test for the Identification of Group A Streptococci and Enterococci with Chromogenic Substrates. *Zentralblatt für Bakteriologie* **272**: 191-195

Kawarabayasi Y, Sawada M, Horikawa H, Haikawa Y, Hino Y, Yamamoto S, Sekine M, Baba S, Kosugi H, Hosoyama A, Nagai Y, Sakai M, Ogura K, Otsuka R, Nakazawa H, Takamiya M, Ohfuku Y, Funahashi T, Tanaka T, Kudoh Y, Yamazaki J, Kushida N, Oguchi A, Aoki K, Kikuchi H. (1998) Complete sequence and gene organization of the genome of a hyper-thermophilic archaebacterium, *Pyrococcus horikoshii* OT3. *DNA Research* **5**: 55-76

Kim J.K., Kim S.J., Lee H.G., Lim J.S., Kim S.J., Cho S.H., Jeong W.H., Choe I.S., Chung T.W., Paik S.G., Choe Y.K. (2001) Molecular Cloning and Characterisation of *Mycobacterium bovis* BCG *pcp* Gene Encoding Pyrrolidone Carboxyl Peptidase. *Molecules and Cells* **12**: 347-352

Kozak, M. (1983) Comparison of Initiation of Protein Synthesis in Procaryotes, Eucaryotes and Organelles. *Microbiological Reviews* **47**: 1-45

Kwiatkowska, J., Torain, B. and Glenner, G.G. (1974) A Pyrrolidonecarboxylate Peptidase from the Particulate Fraction of *Klebsiella cloacae*. *The Journal of Biological Chemistry* **249**: 7729-7736

-L-

Laemmli, U.K. (1970) Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. *Nature* **227**: 680-685

Lauffart, B. and Mantle, D. (1988) Rationalization of Aminopeptidase Activities in Human Skeletal Muscle Soluble Extract. *Biochimica et Biophysica ACTA* **956**: 300-306

Lauffart, B., McDermott, J.R., Biggins, J.A., Gibson, A.M. and Mantle, D. (1989) Purification and characterization of pyroglutamyl aminopeptidase from human cerebral cortex. *Biochemical Society Transactions* **17**: 207-208

Ledent, P., Duez, C., Vanhove, M., Lejeune, A., Fonze, E., Charlier, P., Rhazi-Filali, F., Thamm, I., Guillaume, G., Samyn, B., Devreese, B., Van Beeumen, J., Lamotte-

Brasseur, J. and Frere, J-M. (1997) Unexpected influence of a C-terminal-fused His-tag on the processing of an enzyme and on the kinetic and folding parameters. *FEBS Letters* **413**: 194-196

Le Saux, O., Gonzales, T. and Robert-Baudouy, J. (1996) Mutational Analysis of the Active Site of *Pseudomonas fluorescens* Pyrrolidone Carboxyl Peptidase. *Journal of Bacteriology* **178**: 3308-3313

-M-

Makrides, A.C. (1996) Strategies for Achieving High-Level Expression of Genes in *Escherichia coli*. *Microbiological Reviews* **60**: 512-538

Malhotra, R. and Taqueti, V. (2000) Structure-based Rational Drug Design: Not So Irrational Anymore. *Harvard Science Review* 56-60

Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning, A Laboratory Manual. Cold Spring Harbour Laboratory

Mantle, D., Lauffart, B. and Gibson, A. (1991) Purification and characterisation of leucyl aminopeptidase and pyroglutamyl aminopeptidase from human skeletal muscle. *Clinica Chimica ACTA* **197**: 35-46

Mantle, D., Lauffart, B., McDermot, J. and Gibson, A. (1990) Characterisation of aminopeptidases in human kidney soluble fraction. *Clinica Chimica ACTA* **187**: 105-114

Mantle, D., Lauffart, B., Perry, E.K. and Perry, R.H. (1989) Comparison of major cortical aminopeptidase activity in normal brain and brain from patients with Alzheimer's disease. *Journal of the Neurological Sciences* **89**:227-234

Martini, F., Bossa, F. and Barra, D. (1985) Assay of Pyroglutamyl Aminopeptidase by High-Performance Liquid Chromatography and its Use in Peptide Sequencing. *Peptides* **6**: 103-105

Mather, M.W., Keightley, J.A. and Fee, J.A. (1993) Recovery and Cloning of Genomic DNA Fragments from Dried Agarose Gels. *Methods in Enzymology* **218**: 695-704

Mineyama, R. and Saito, K. (1998) Partial purification and some properties of pyroglutamyl peptidase from *Enterococcus faecalis*. *Microbios* **94**: 47-62

Mitchell, M.J., Conville, P.S. and Gill, V.J. (1987) Rapid Identification of Enterococci by Pyrrolidonyl Aminopeptidase Activity (PyRase). *Diagnostic Microbiology and Infectious Disease* **6**: 283-286

Mozdzanowski, J., Bongers, J. and Anumula, K. (1998) High-Yield Deblocking of Amino Termini of Recombinant Immunoglobulins with Pyroglutamate Aminopeptidase. *Analytical Biochemistry* **260**: 183-187

Mudge, A.W. and Fellows, R.E. (1973) Bovine Pituitary Pyrrolidonecarboxylyl Peptidase. *Endocrinologia* **93**: 1428-1434

Mulczyk, M. and Szewczuk, A. (1970) Pyrrolidonyl Peptidase in Bacteria: A New Colrimetric Test for Differentiation of Enterobacteriaceae. *Journal of General Microbiology* **61**: 9-13

Mullis, K.B. and Faloona, F.A. (1987) Specific Synthesis of DNA *in Vitro* via a Polymerase-Catalyzed Chain Reaction. *Methods in Enzymology* **155**: 335-350

Murzin, A.G. (1993) Can homologous proteins evolve different enzymatic activities? *Trends in Biochemical Science* **18**: 403-405

-N-

Nakamura, Y., Gojobori, T. and Ikemura, T. (2000) Codon usage tabulated from international DNA sequence databases: status for the year 2000. *Nucleic Acids Research* **28**: 292

Nicholas, K.B., Nicholas, H.B. Jr. and Deerfield II, D.W. (1997) GeneDoc: Analysis and Visualization of Genetic Variation. *EMBnet News* **4**: 14

Nilsson, L.O. and Mannervik, B. (2001) Improved heterologous expression of human glutathione transferase A4-4 by random silent mutagenesis of codons in the 5' region. *Biochimica et Biophysica ACTA* **1528**: 101-106

Novy, R., Drott, D., Yaeger, K. and Mierendorf, R. (2001) Overcoming the codon bias of *E. coli* for enhanced protein expression. *Innovations* **12**:1-3

-O-

O'Connor, B. and O'Cuinn, G. (1984) Localization of a narrow-specificity thyroliberin hydrolizing pyroglutamate aminopeptidase in synaptosomal membranes of guinea-pig brain. *European Journal of Biochemistry* **144**: 271-278

O'Cuinn, G., O'Connor, B. and Elmore, M. (1990) Degradation of Thyrotropin-Releasing Hormone and Luteinising Hormone-Releasing Hormone by Enzymes of Brain Tissues. *Journal of Neurochemistry* **54**: 1-13

Odagaki, Y., Hayashi, A., Okada, K., Hirotsu, K., Kabashima, T., Ito, K., Yoshimoto, T., Tsuru, D., Sato, M. and Clardy, J. (1999) The Crystal Structure Of Pyroglutamyl Peptidase I from *Bacillus amyloliquefaciens* reveals a new structure for a cysteine protease. *Structure* **7**: 399-411

Ogasahara, K., Khechinashvili, N., Nakamura, M., Yoshimoto, T. and Yutani, K. (2001) Thermal stability of pyrrolidone carboxyl peptidase from the hyperthermophilic Archaeon, *Pyrococcus furiosus*. *European Journal of Biochemistry* **268**: 3233-3242

Ogasahara, K., Nakamura, M., Nakura, S., Tsunasawa, S., Kato, I., Yoshimoto, T. and Yutani, K. (1998) The Unusually Slow Unfolding Rate Causes the High Stability of Pyrrolidone Carboxyl Peptidase from a Hyperthermophile, *Pyrococcus furiosus*: Equilibrium and Kinetic Studies of Guanidine Hydrochloride-Induced Unfolding and Refolding. *Biochemistry* **37**: 17537-17544

O'Leary, R. and O'Connor, B. (1995) Thyrotropin-Releasing Hormone. *Journal of Neurochemistry* **65**: 953-963

Orlowski, M. and Meister, A. (1971), Enzymology of Pyrrolidone Carboxylic Acid, In: *The Enzymes, IV: Hydrolysis*, Boyer, P.D. (ed.), Academic Press, New York 123-151.

-P-

Padmakumar, V.C. and Varadarajan, R. (2003) A gradient PCR-based screen for use in site-directed mutagenesis. *Analytical Biochemistry* **314**: 310-315

- Patti, J.M., Schneider, A., Garza, N. and Boles, J.O. (1995) Isolation and characterisation of pcp, a gene encoding a pyrrolidone carboxyl peptidase in *Staphylococcus aureus*. *Gene* **166**: 95-99
- Pereira, P.J.B., Bergner, A., Macedo-Ribeiro, S., Huber, R., Matschiner, G., Fritz, H, Sommerhoff, C.P. and Bode, W. (1998) Human β -tryptase is a ring-like tetramer with active sites facing a central pore. *Nature* **392**: 306-311
- Perlman, J.H., Laakkonen, L., Osman, R. and Gershengorn, M.C. (1994b) A Model of the Thyrotropin-releasing Hormone (TRH) Receptor Binding Pocket. *The Journal of Biological Chemistry* **269**: 23383-23386
- Perlman, J.H., Thaw, C.N., Laakkonen, L., Bowers, C.Y., Osman, R. and Gershengorn, M.C. (1994a) Hydrogen Bonding Interaction of Thyrotropin-releasing Hormone (TRH) with Transmembrane Tyrosine 106 of the TRH Receptor. *The Journal of Biological Chemistry* **269**: 1610-1613
- Pierro, M. and Orsatti, M. (1970) On intestinal hydrolysis of pyrrolidonylpeptides. *Bollettino Della Societa Italiana Di Biologia Sperimentale* **45**: 1630
- Podell, D.N. and Abraham, G.N. (1978) A Technique For The Removal Of Pyroglutamic Acid From The Amino Terminus Of Proteins Using Calf Liver Pyroglutamate Amino Peptidase. *Biochemical And Biophysical Research Communications* **81**: 176-185
- Porath, J., Carlsson, J., Olsson, I. And Belfrage, G. (1975) Metal chelate affinity chromatography, a new approach to protein fractionation. *Nature* **258**: 598-599
- Prasad, C. (1987) Activation/Inactivation Of Rat Tissue Pyroglutamate Aminopeptidase By Disulfide Bond-Reducing Agents. *Neuropeptides* **9**: 211-215
- Prasad, C. (1995) Bioactive Cyclic Dipeptides. *Peptides* **16**: 151-164
- Prasad, C. and Peterkofsky, A. (1976) Demonstration of Pyroglutamylpeptidase and Amidase Activities toward Thyrotropin-releasing Hormone in Hamster Hypothalamus Extracts. *The Journal of Biological Chemistry* **251**: 3229-3234

Prasad, C., Masatomo, M., Woodrum, P., Wilber, J.F. and Ewards, R.M. (1983) Developmental Changes in the Distribution of Rat Brain Pyroglutamate Aminopeptidase, A Possible Determinant of Endogenous Cyclo(His-Pro) Concentrations. *Neurochemical Research* **8**: 389-399

-R-

Ramirez, M., Sanchez, B., Arechaga, G., Garcia, S., Lardelli, P., Venzon D. and De Gandarias, J.M. (1991) Diurnal variation and left-right distribution of pyroglutamyl peptidase I activity in the rat brain and retina. *ACTA Endocrinologica* **125**: 570-573

Rawlings, N.D., Tolle, D.P. and Barrett, A.J. (2004) MEROPS: the peptidase database. *Nucleic Acids Research* **32**: D160-D164

Russo, C., Violani, E., Salis, S., Venezia, V., Dolcini, V., Damonte, G., Benatti, U., D'Arrigo, C., Patrone, E., Carlo, P. and Schettini, G. (2002) Pyroglutamate-modified amyloid b-peptides - A β N3(pE) - strongly affect cultured neuron and astrocyte survival. *Journal of Neurochemistry* **82**: 1480-1489

-S-

Sanchez, B., Alba, F., de Dios Luna, J., Martinez, J.M., Prieto, I. And Ramirez, M. (1996) Pyroglutamyl peptidase I levels and their left-right distribution in the rat retina and hypothalamus are influenced by light-dark conditions. *Brain Research* **731**: 254-257

Schauder, B., Schomburg, L., Kohrle, J. and Bauer, K. (1994) Cloning of a cDNA encoding an ectoenzyme that degrades thyrotropin-releasing hormone. *Proceedings of the National Academy of Sciences of the United States of America* **91**: 9534-9538

Schmitmeier, S., Thole, H., Bader, A. and Bauer, K. (2002) Purification and characterization of the thyrotropin-releasing hormone (TRH)-degrading serum enzyme and its identification as a product of liver origin. *European Journal of Biochemistry* **269**: 1278-1286

- Schomburg, L., Turwitt, S., Prescher, G., Lohmann, D., Horsthemke, B. and Bauer, K. (1999) Human TRH-degrading ectoenzyme: cDNA cloning, functional expression, genomic structure and chromosomal assignment. *European Journal of Biochemistry* **265**: 415-422
- Schwede, T., Kopp, J., Guex, N. and Peitsch, M.C. (2003) SWISS-MODEL: an automated protein homology-modelling server. *Nucleic Acids Research* **31**: 3381-3385
- Singleton, M.R. and Littlechild, J.A. (2001) Pyrrolidone Carboxylpeptidase from *Thermococcus litoralis*. *Methods In Enzymology* **330**: 394-403
- Singleton, M.R., Isupov, M.N. and Littlechild, J.A. (1999a) Crystallisation and preliminary X-ray diffraction studies of pyrrolidone carboxyl peptidase from the hyperthermophilic archaeon *Thermococcus litoralis*. *ACTA Crystallographica* **D55**: 702-703
- Singleton, M.R., Isupov, M.N. and Littlechild, J.A. (1999b) X-ray structure of pyrrolidone carboxyl peptidase from the hyperthermophilic archaeon *Thermococcus litoralis*. *Structure* **7**: 237-244
- Singleton, M.R., Taylor, S.J.C., Parrat, J.S. and Littlechild, J.A. (2000) Cloning, expression, and characterization of pyrrolidone carboxyl peptidase from the archaeon *Thermococcus litoralis*. *Extremophiles* **4**: 297-303
- Skerra, A. and Schmidt, T.G.M. (1999) Application of a peptide ligand for streptavidin: the *Strep*-tag. *Biomolecular Engineering* **16**: 79-86
- Smith P.K., Krohn R.I., Hermanson G.T., Mallia A.K., Gartner F.H., Provenzano M.D., Fujimoto E.K., Goeke N.M., Olson B.J. and Klenk D.C. (1985) Measurement of protein using bicinchoninic acid. *Analytical Biochemistry* **150**: 76-85
- Sokabe, M., Kawamura, T., Sakai, N., Yao, M., Watanabe, N. and Tanaka, I. (2002) X-ray crystal structure of pyrrolidone-carboxylate peptidase from hyperthermophilic *Pyrococcus horikoshii*. *Journal of Structural and Functional Genomics* **2**: 145-154

Sullivan, J.J., Muchnick, E.E., Davidson, B.E. and Jago, G.R. (1977) Purification and Properties of the Pyrrolidonecarboxylate Peptidase of *Streptococcus faecium*. *Australian Journal of Biological Science* **30**: 543-552

Szewczuk, A. and Kwiatkowska, J. (1970) Pyrrolidone Peptidase in Animal, Plant and Human Tissues. Occurrence and some Properties of the Enzyme. *European Journal of Biochemistry* **15**: 92-96

Szewczuk, A. and Mulczyk, M. (1969) Pyrrolidonyl Peptidase in Bacteria. The Enzyme from *Bacillus subtilis*. *European Journal of Biochemistry* **6**: 63-67

-T-

Tanaka, H., Chinami, M., Mizushima, T., Ogasahara, Ota, M., Tsukihara, T. and Yutani, K. (2001) X-Ray Crystalline Structures of Pyrrolidone Carboxyl Peptidase from a Hyperthermophile, *Pyrococcus furiosus*, and Its Cys-Free Mutant. *Journal of Biochemistry* **130**: 107-118

Tsunasawa, S., Nakura, S., Tanigawa, T. and Kato, I. (1998) Pyrrolidone Carboxyl Peptidase from the Hyperthermophilic Archaeon *Pyrococcus furiosus*: Cloning and Overexpression in *Escherichia coli* of the Gene, and its Application to Protein Sequence Analysis. *Journal of Biochemistry* **124**: 778-783

Tsuru, D., Fujiwara, K. and Kado, K. (1978) Purification and Characterisation of L-Pyrrolidonecarboxylate Peptidase from *Bacillus amyloliquefaciens*. *Journal of Biochemistry* **84**: 467-476

Tsuru, D., Nakamura, K., Yoshimoto, T. and Fujiwara, K. (1984) Pyroglutamyl-Peptidase from *Bacillus amyloliquefaciens*. An improved purification method and some properties of the enzyme. *Biochimica et Biophysica ACTA* **791**: 117-122

Tsuru, D., Sakabe, K., Yoshimoto, T. and Fujiwara, K. (1982) Pyroglutamyl Peptidase from Chicken Liver: Purification and some Properties. *Journal of Pharmacobio-Dynamics* **5**: 859-868

-U-

Uliana, J.A. and Doolittle, R.F. (1969) Pyrrolidonecarboxyl Peptidase: Studies on the Specificity of the Enzyme. *Archives of Biochemistry and Biophysics* **131**: 561-565

-V-

Valdivia, A., Irazusta, J., Fernandez, D., Mugica, J., Ochoa, C. and Casis, L. (2004) Pyroglutamyl peptidase I and prolyl endopeptidase in human semen: increased activity in necrozoospermia. *Regulatory Peptides* **122**: 79-84

-W-

Wilk, S., Friedman, T.C. and Kline, T.B. (1985) Pyroglutamyl Diazomethyl Ketone: Potent Inhibitor of Mammalian Pyroglutamyl Peptide Hydrolase. *Biochemical And Biophysical Research Communications* **130**: 662-668

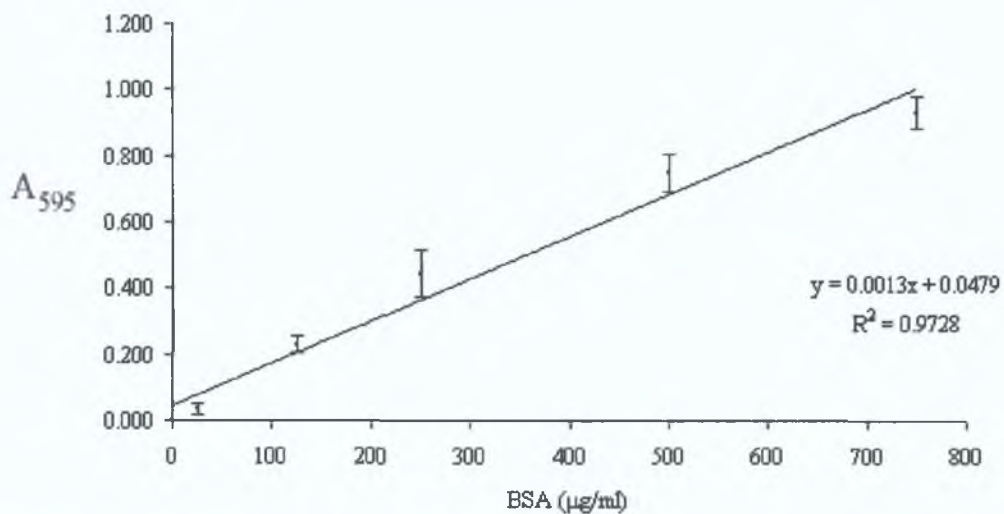
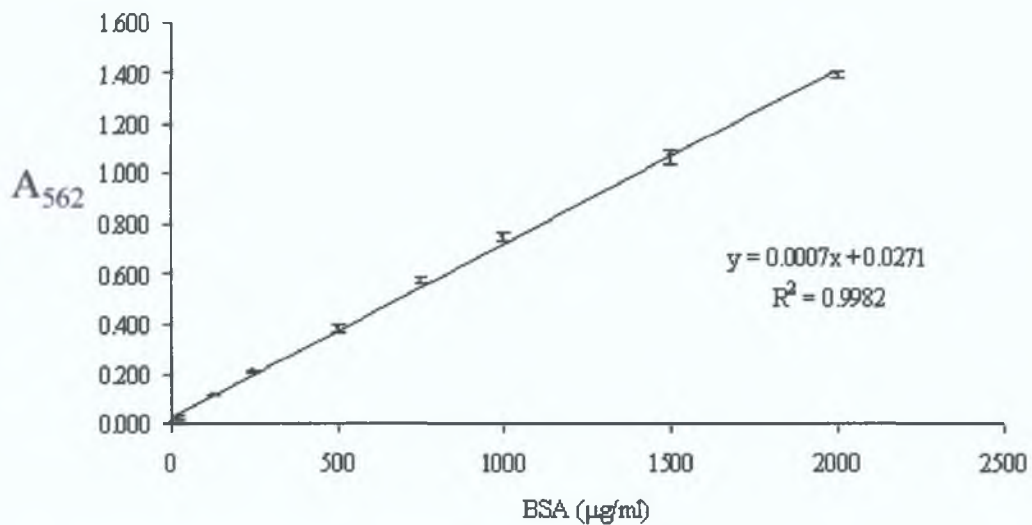
-Y-

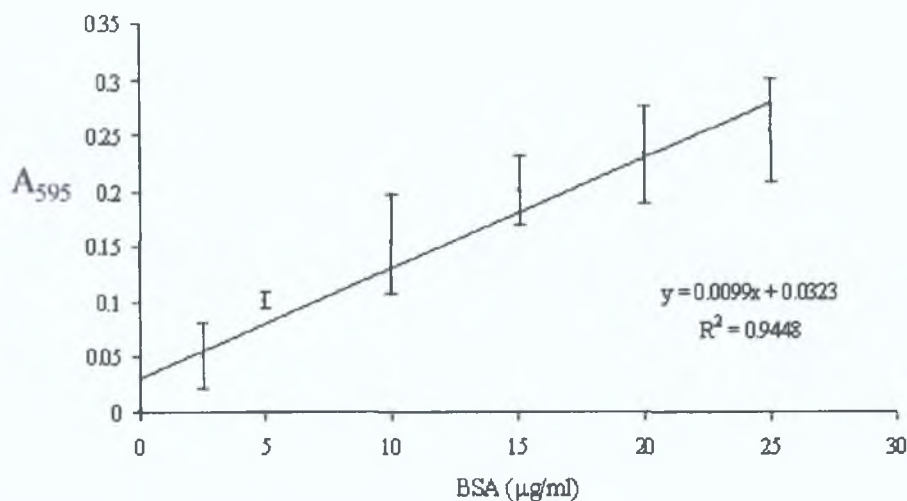
Yoshimoto, T., Shimoda, T., Kitazono, A., Kabashima, T., Ito, K. and Tsuru, D. (1993) Pyroglutamyl Peptidase Gene from *Bacillus amyloliquefaciens*: Cloning, Sequencing, and Crystallisation of the Expressed Enzyme. *Journal of Biochemistry* **113**: 67-73

-Z-

Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research* **31**: 3406-3415

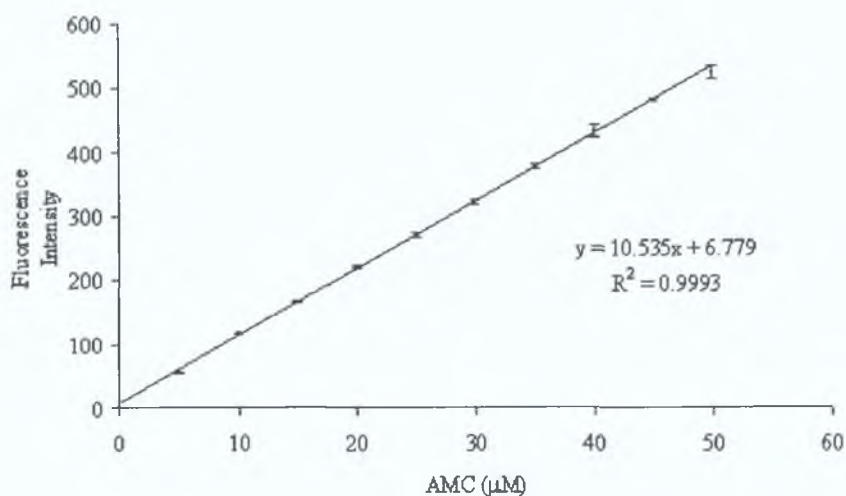
Appendix A
Standard Curves





BSA standard curve for qualitative Coomassie assay

Standard curve of BSA (μg/ml) versus absorbance at 595 nm for the determination of protein concentration by the qualitative Coomassie assay described in Section 2.14.3.



AMC standard curve

Standard curve of AMC (μM) versus fluorescence intensity as described in Section 2.15.1, for the determination of AMC released by PAP1 activity.

Appendix B
Sequence Data

>pRV1_forward read, MWG (top strand)
GACTCACTATAGGGCGAATTGGGCCCTCTAGATGCATGCTCGAGCGGCCGCCAGTGTGATGG
ATATCTGCAGAATTCGGCTAACAGAAGCAGGTCCGAGGCACAGCCCGATCCCGCCATGGAG
CAGCCGAGGAAGGCGGTGGTAGTGACGGGATTTGGCCCTTTTGGGGAACACACCGTGAACG
CCAGTTGGATTGCAGTTCAGGAGCTAGAAAAAGCTAGGCCTTGGCGACAGCGTGGACCTGCA
TGTGTACGAGATTCCGGTTGAGTACCAAACAGTCCAGAGACTCATCCCCGCCCTGTGGGAGA
AGCACAGTCCACAGCTGGTGGTGCATGTGGGGGTGTCAGGCATGGCGACCACAGTCACACT
GGAGAAATGTGGACACAACAAGGGCTACAAGGGGCTGGACAACCTGCCGCTTTTGGCCCGGC
TCCCAGTGCTGCGTGGAGGACGGGCCTGAAAGCATTGACTCCATCATCGACATGGATGCTGT
GTGCAAGCGAGTCACCACGTTGGGCCTGGATGTGTGCGGTGACCATCTCGCAGGATGCCGGCA
GATATCTCTGCGACTTTACCTACTACACCTCTTTGTACCAGAGTCACGGTCGATCAGCCTTCG
TCCACGTGCCCCCACTGGGGAAGCCGTACAACGCGGACCAGCTGGGCAGGGCACTGAGAGC
CATCATTGAGGAGATGTTGGACC

>pRV1_reverse read, MWG (top strand)
TGGTAGTGACGGGATTTGGCCCTTTTGGGGAACACACCGTGAACGCCAGTTGGATTGCAGTT
CAGGAGCTAGAAAAGCTAGGCCTTGGCGACAGCGTGGACCTGCATGTGTACGAGATTCCGG
TTGAGTACCAAACAGTCCAGAGACTCATCCCCGCCCTGTGGGAGAAGCACAGTCCACAGCT
GGTGGTGCATGTGGGGGTGTCAGGCATGGCGACCACAGTCACACTGGAGAAATGTGGACAC
ACAAGGGCTACAAGGGGCTGGACAACCTGCCGCTTTTGGCCCGGCTCCCAGTGCTGCGTGGA
GGACGGGCCTGAAAGCATTGACTCCATCATCGACATGGATGCTGTGTGCAAGCGAGTCACC
ACGTTGGGCCTGGATGTGTGCGGTGACCATCTCGCAGGATGCCGGCAGATATCTCTGCGACTT
TACCTACTACACCTCTTTGTACCAGAGTCACGGTCGATCAGCCTTCGTCCACGTGCCCCCACT
GGGGAAGCCGTACAACGCGGACCAGCTGGGCAGGGCACTGAGAGCCATCATTGAGGAGATG
TTGGACCTCCTGGAGCAGTCAGAGGGCAAAATCAACTATTGCCACAAACACTGAGGGACGC
TCAGGTCTCCTAAGACCTCATCCTGAAGCCGAATTCCAGCACACTGGCGGCCGTTACTAGTG
GATCCGAGCTCGGTACCAAGCTT

>pRV3_forward read, MWG (top strand)
AATCATAAAAAATTTATTTGCTTTGTGAGCGGATAACAATTTACACAGAATTCATTAAAGA
GGAGAAATTAACCATGGAGCAGCCGAGGAAGGCGGTGGTAGTGACGGGATTTGGCCCTTTT
GGGGAACACACCGTGAACGCCAGTTGGATTGCAGTTCAGGAGCTAGAAAAGCTAGGCCTTG
GCGACAGCGTGGACCTGCATGTGTACGAGATTCCGGTTGAGTACCAAACAGTCCAGAGACT
CATCCCCGCCCTGTGGGAGAAGCACAGTCCACAGCTGGTGGTGCATGTGGGGGTGTCAGGC
ATGGCGACCACAGTCACACTGGAGAAATGTGGACACAACAAGGGCTACAAGGGGCTGGACA
ACTGCCGCTTTTGGCCCGGCTCCCAGTGCTGCGTGGAGGACGGGCCTGAAAGCATTGACTCC
ATCATCGACATGGATGCTGTGTGCAAGCGAGTCACCACGTTGGGCCTGGATGTGTGCGGTGAC
CATCTCGCAGGATGCCGGCAGATATCTCTGCGACTTTACCTACTACACCTCTTTGTACCAGAG
TCACGGTCGATCAGCCTTCGTCCACGTGCCCCCACTGGGGAAGCCGTACAACGCGGACCAGC
TGGGCAGGGCACTGAGAGCCATCATTGA

>pRV3_reverse read, MWG (top strand)
AGCCGAGGAAGGCGGTGGTAGTGACGGAATTTGGCCCTTTTGGGGAACACACCGTGAACGC
CAGTTGGATTGCAGTTCAGGAGCTAGAAAAGCTAGGCCTTGGCGACAGCGTGGACCTGCAT
GTGTACGAGATTCCGGTTGAGTACCAAACAGTCCAGAGACTCATCCCCGCCCTGTGGGAGAA
GCACAGTCCACAGCTGGTGGTGCATGTGGGGGTGTCAGGCATGGCGACCACAGTCACACTG
GAGAAATGTGGACACAACAAGGGCTACAAGGGGCTGGACAACCTGCCGCTTTTGGCCCGGCT
CCCAGTGCTGCGTGGAGGACGGGCCTGAAAGCATTGACTCCATCATCGACATGGATGCTGTG
TGCAAGCGAGTCACCACGTTGGGCCTGGATGTGTGCGGTGACCATCTCGCAGGATGCCGGCAG
ATATCTCTGCGACTTTACCTACTACACCTCTTTGTACCAGAGTCACGGTCGATCAGCCTTCGT
CCACAGTGCCCCCACTGGGGAAGCCAGTACAACGCGGACCAGCTGGGCAGGGCACTGAGAG
CCATCATTGAGGAGATGTTGGACCTCCTGGAGCAGTCAGAGGGCAAAATCAACTATTGCCAC
AAACACGGATCTCATCACCATCACCATCACTAAGCTTAATTAGC

>pRV8_forward read, Qiagen (top strand)

GAGAAATCATAAAAAATTTATTTGCTTTGTGAGCGGATAACAATTATAATAGATTCAATTGT
GAGCGGATAACAATTTACACAGAAATTCATTAAAGAGGAGAAATTAACCATGGAGCAGCCG
AGGAAGGCGGTGGTAGTGACGGGATTTGGCCCTTTTGGGGAACACACCGTGAAACGCCAGTT
GGATTGCAGTTTCAGGAGCTAGAAAAGCTAGGCCTTGGCGACAGCGTGGACCTGCATGTGTA
CGAGATTCCGGTTGAGTACCAAACAGTCCAGAGACTCATCCCCGCCCTGTGGGAGAAGCAC
AGTCCACAGCTGGTGGTGCATGTGGGGGTGTCAGGCATGGCGACCACAGTCACACTGGAGA
AATGTGGACACAACAAGGGCTACAAGGGGCTGGACAACCTGCCGCTTTTGGCCCGGCTCCCA
GTGCTGCGTGGAGGACGGGCCTGAAAGCATTGACTCCATCATCGACATGGATGCTGTGTGCA
AGCGAGTCACCACGTTGGGCCTGGATGTGTGCGGTGACCATCTCGCAGGATGCCGGCAGATAT
CTCTGCGACTTTACCTACTACACCTCTTTGTACCAGAGTCACGGTCGATCAGCCTTCGTCCAC
GTGCCCCCACTGGGGAAGCCGTACAACGCGGACCAGCTGGGCAGGGCACTGAGAGCCATCA
TTGAGGAGATGTTGGACCTCCTGGAGCAGTCAGAGGGCAAAATCAACTATTGCCACAAACA
CGGATCTCATCACCATCACCATCACTAAGCTTAATTAGCTGAGCTTGGACTCCTGTTGATAG
ATCCAGTAATGACCTCAGAACTTCATCTGGATTTGTTCA

>pRV8_reverse read, Qiagen (top strand)

AGCAGCCGAGGAAGGCGGTGGTAGTGACGGGATTTGGCCCTTTTGGGGAACACACCGTGAA
CGCCAGTTGGATTGCAGTTCAGGAGCTAGAAAAGCTAGGCCTTGGCGACAGCGTGGACCTG
CATGTGTACGAGATTCCGGTTGAGTACCAAACAGTCCAGAGACTCATCCCCGCCCTGTGGGA
GAAGCACAGTCCACAGCTGGTGGTGCATGTGGGGGTGTCAGGCATGGCGACCACAGTCACA
CTGGAGAAATGTGGACACAACAAGGGCTACAAGGGGCTGGACAACCTGCCGCTTTTGGCCCG
GCTCCCACTGCTGCGTGGAGGACGGGCCTGAAAGCATTGACTCCATCATCGACATGGATGCT
GTGTGCAAGCGAGTCACCACGTTGGGCCTGGATGTGTGCGGTGACCATCTCGCAGGATGCCGG
CAGATATCTCTGCGACTTTACCTACTACACCTCTTTGTACCAGAGTCACGGTCGATCAGCCTT
CGTCCACGTGCCCCCACTGGGGAAGCCGTACAACGCGGACCAGCTGGGCAGGGCACTGAGA
GCCATCATTGAGGAGATGTTGGACCTCCTGGAGCAGTCAGAGGGCAAAATCAACTATTGCC
ACAAACACGGATCTCATCACCATCACCATCACTAAGCTTAATTAGACTGAGCTTG

>pRV5_forward read, Qiagen (top strand)

TCGGCTCGTATAATGTGTGGAATTGTGAGCGGATAACAATTTACACAGGAAACAGAATTCA
TTAAAGAGGAGAAATTAACCATGGAGCAGCCGAGGAAGGCGGTGGTAGTGACGGGATTTGG
CCCTTTTGGGGAACACACCGTGAAACGCCAGTTGGATTGCAGTTCAGGAGCTAGAAAAGCTA
GGCCTTGGCGACAGCGTGGACCTGCATGTGTACGAGATTCCGGTTGAGTACCAAACAGTCCA
GAGACTCATCCCCGCCCTGTGGGAGAAGCACAGTCCACAGCTGGTGGTGCATGTGGGGGTG
TCAGGCATGGCGACCACAGTCACACTGGAGAAATGTGGACACAACAAGGGCTACAAGGGGC
TGGACAACCTGCCGCTTTTGGCCCGGCTCCCACTGCTGCGTGGAGGACGGGCCTGAAAGCATT
GACTCCATCATCGACATGGATGCTGTGTGCAAGCGAGTCACCACGTTGGGCCTGGATGTGTC
GGTGACCATCTCGCAGGATGCCGGCAGATATCTCTGCGACTTTACCTACTACACCTCTTTGTA
CCAGAGTCACGGTCGATCAGCCTTCGTCCACGTGCCCCCACTGGGGAAGCCGTACAACGCGG
ACCAGCTGGGCAGGGCACTTGAGAGCCATCATTGAGGAGATGTTGGACCTCCTGGAGCAGT
CAGAGGGCAAAATCAAC

>pRV5_reverse read, Qiagen (top strand)

ATTCATTAAAGAGGAGAAATTAACCATGGAGCAGCCGAGGAAGGCGGTGGTAGTGACGGGA
TTTGGCCCTTTTGGGGAACACACCGTGAAACGCCAGTTGGATTGCAGTTCAGGAGCTAGAAAA
GCTAGGCCTTGGCGACAGCGTGGACCTGCATGTGTACGAGATTCCGGTTGAGTACCAAACAG
TCCAGAGACTCATCCCCGCCCTGTGGGAGAAGCACAGTCCACAGCTGGTGGTGCATGTGGG
GGTGTGAGGCATGGCGACCACAGTCACACTGGAGAAATGTGGACACAACAAGGGCTACAAG
GGGCTGGACAACCTGCCGCTTTTGGCCCGGCTCCCACTGCTGCGTGGAGGACGGGCCTGAAAG
CATTGACTCCATCATCGACATGGATGCTGTGTGCAAGCGAGTCACCACGTTGGGCCTGGATG
TGTCGGTGACCATCTCGCAGGATGCCGGCAGATATCTCTGCGACTTTACCTACTACACCTCTT
TGTACCAGAGTCACGGTCGATCAGCCTTCGTCCACGTGCCCCCACTGGGGAAGCCGTACAAC
GCGGACCAGCTGGGCAGGGCACTGAGAGCCATCATTGAGGAGATGTTGGACCTCCTGGAGC
AGTCAGAGGGCAAAATCAACTATTGCCACAAACACGGATCTCATCACCATCACCATCACTAA
GCTTGGCTGTTTTG

>pRV5_Δ6xHis_reverse read, Qiagen (top strand)

AATCACTGCATATTTAGGTCGCTCAAGGCGCACTCCCTTCTGGATAATGTTTATGCGCCGAC
ATCATAACGGTTCTGGCAAATAATCTGAAATGAGCTGTTGACAATTAATCATCGGCTCGTAT
AATGTGTGGAATTGTGAGCGGATAACAATTTACACAGGAAACAGAATTCATTAAGAGGA
GAAATTAACCATGGAGCAGCCGAGGAAGGCGGTGGTAGTGACGGGATTTGGCCCTTTTGGG
GAACACACCGTGAACGCCAGTTGGATTGCAGTTCAGGAGCTAGAAAAGCTAGGCCTTGGCG
ACAGCGTGGACCTGCATGTGTACGAGATTCCGGTTGAGTACCAAACAGTCCAGAGACTCATC
CCCGCCCTGTGGGAGAAGCACAGTCCACAGCTGGTGGTGCATGTGGGGGTGTCAGGCATGG
CGACCACAGTCACACTGGAGAAATGTGGACACAACAAGGGCTACAAGGGGCTGGACAACCTG
CCGCTTTTGGCCCGGCTCCAGTGCTGCGTGGAGGACGGGCCTGAAAGCATTGACTCCATCA
TCGACATGGATGCTGTGTGCAAGCGAGTCACCACGTTGGGCCTGGATGTGTGCGTGACCATC
TCGCAGGATGCCGGCAGATATCTCTGCGACTTTACCTACTACACCTCTTTGTACCAGAGTCAC
GGTCGATCAGCCTTCGTCCACGTGCCCCCACTGGGGAAGCCGTACAACGCGGACCAGCTGG
GCAGGGCACTGAGAGCCATCATTGAGGAGATGTTGGACCTCCTGGAGCAGTCAGAGGGCAA
AATCAACTATTGCCACAAACACTAAGGATCCGTCGACCTGCAGCCAAGCTTCTGTTT

>pRV5_R5^SK6^S_forward read, MWG (top strand)

CGCGTAAAGCGGTGGTAGTGACGGGATTTGGCCCTTTTGGGGAACACACCGTGAACGCCAG
TTGGATTGCAGTTCAGGAGCTAGAAAAGCTAGGCCTTGGCGACAGCGTGGACCTGCATGTGT
ACGAGATTCCGGTTGAGTACCAAACAGTCCAGAGACTCATCCCCGCCCTGTGGGAGAAGCA
CAGTCCACAGCTGGTGGTGCATGTGGGGGTGTCAGGCATGGCGACCACAGTCACACTGGAG
AAATGTGGACACAACAAGGGCTACAAGGGGCTGGACAACCTGCCGCTTTTGGCCCGGCTCCC
AGTGCTGCGTGGAGGACGGGCCTGAAAGCATTGACTCCATCATCGACATGGATGCTGTGTGC
AAGCGAGTCACCACGTTGGGCCTGGATGTGTGCGGTGACCATCTCGCAGGATGCCGGCAGAT
ATCTCTGCGACTTTACCTACTACACCTCTTTGTACCAGAGTCACGGTCGATCAGCCTTCGTCC
ACGTGCCCCCACTGGGGAAGCCGTACAA

>pRV5_R5^Sa_forward read, MWG (top strand)

TTAAACCATGGAGCAGCCGCGCAAGGCGGTGGTAGTGACGGGATTTGGCCCTTTTGGGGAAC
ACACCGTGAACGCCAGTTGGATTGCAGTTCAGGAGCTAGAAAAGCTAGGCCTTGGCGACAG
CGTGGACCTGCATGTGTACGAGATTCCGGTTGAGTACCAAACAGTCCAGAGACTCATCCCCG
CCCTGTGGGAGAAGCACAGTCCACAGCTGGTGGTGCATGTGGGGGTGTCAGGCATGGCGAC
CACAGTCACACTGGAGAAATGTGGACACAACAAGGGCTACAAGGGGCTGGACAACCTGCCG
TTTTGCCCCGGCTCCAGTGCTGCGTGGAGGACGGGCCTGAAAGCATTGACTCCATCATCGA
CATGGATGCTGTGTGCAAGCGAGTCACCACGTTGGGCCTGGATGTGTGCGTGACCATCTCGC
AGGATGCCGGCAGATATCTCTGCGACTTTACCTACTACACCTCTTTGTACCAGAGTCACGGT
CGATCAGCCTTCGTCCACGTGCCCCCACTGGGGAAGCCGTACAACGCGGACCAGCTGGGCA
GGGCACTGAGAGCCATCATTGAGGAGATGTTGGACCTCCTGGAGCAGTCAGAGGGCAAAAT
CAACTATTGCCACAAACACGGATCTCATCACCATCACCATCACTAAGCTTGGCTGTTTTGGC
GGATGAGAGAAGATTTTACGCCTGATACAGATTAATCAGAACGCAGAAGCGGTCTGATAA
AACAGAATTTGCCTGGCGGCAGTAGCGCGGTGGTCCCACCTGACCCCATGCCGAACCTCAGA
AGTGAACGCCGTAGCGCCGATGGTAGTGTGGGGTCTCCCCCATGCGAGAGTAGGGAACTG
CCAGGCATCAAATAAAAC

>pRV5_R5^Sa_reverse read, MWG (top strand)

TCGTGTCGCTCAAGGCGCACTCCCCGTTCTGGATAATGTTTTTTGCGCCGACATCATAACGGT
TCTGGCAAATATTCTGAAATGAGCTGTTGACAATTAATCATCGGCTCGTATAATGTGTGGAA
TTGTGAGCGGATAACAATTTACACAGGAAACAGAATTCATTAAGAGGAGAAATTAACCA
TGGAGCAGCCGCGCAAGGCGGTGGTAGTGACGGGATTTGGCCCTTTTGGGGAACACACCGT
GAACGCCAGTTGGATTGCAGTTCAGGAGCTAGAAAAGCTAGGCCTTGGCGACAGCGTGGAC
CTGCATGTGTACGAGATTCCGGTTGAGTACCAAACAGTCCAGAGACTCATCCCCGCCCTGTG
GGAGAAGCACAGTCCACAGCTGGTGGTGCATGTGGGGGTGTCAGGCATGGCGACCACAGTC
ACACTGGAGAAATGTGGACACAACAAGGGCTACAAGGGGCTGGACAACCTGCCGCTTTTGCC
CCGGCTCCCACTGCTGCGTGGAGGACGGGCCTGAAAGCATTGACTCCATCATCGACATGGAT
GCTGTGTGCAAGCGAGTCACCACGTTGGGCCTGGATGTGTGCGGTGACCATCTCGCAGGATGC
CGGCAGATATCTCTGCGACTTTACCTACTACACCTCTTTGTACCAGAGTCACGGTCGATCAGC
CTTCGTCCACGTGCCCCCACTGGGGAAGCCGTACAACGCGGACCAGCTGGGCAGGGCACTG
AGAGCCATCATTGAGGAGATGTTGGACCTCCTGGAGCAGTCAGAGGGCAAAATCAACTATT
GCCACAAACACGGATCTCATCACCATCACCATCACTAAGCTTGGCTGTTTGGCG

>pRV5_R5^b_forward read, MWG (top strand)
TTAACCATGGAGCAGCCGCGTAAGGCGGTGGTAGTGACGGGATTTGGCCCTTTTGGGGAAC
ACACCGTGAACGCCAGTTGGATTGCAGTTCAGGAGCTAGAAAAGCTAGGCCTTGGCGACAG
CGTGGACCTGCATGTGTACGAGATTCCGGTTGAGTACCAAACAGTCCAGAGAGTCATCCCCG
CCCTGTGGGAGAAGCACAGTCCACAGCTGGTGGTGCATGTGGGGGTGTCAGGCATGGCGAC
CACAGTCACACTGGAGAAATGTGGACACAACAAGGGCTACAAGGGGCTGGACAACCTGCCG
TTTTGCCCCGGCTCCCAGTGCTGCGTGGAGGACGGGCCTGAAAGCATTGACTCCATCATCGA
CATGGATGCTGTGTGCAAGCGAGTCACCACGTTGGGCCTGGATGTGTGCGTGACCATCTCGC
AGGATGCCGGCAGATATCTCTGCGACTTTACCTACTACACCTCTTTGTACCAGAGTCACGGT
CGATCAGCCTTCGTCCACGTGCCCCCACTGGGGAAGCCGTACAACGCGGACCAGCTGGGCA
GGGCACTGAGAGCCATCATTGAGGAGATGTTGGACCTCCTGGAGCAGTCAGAGGGCAAAAT
CAACTATTGCCACAAACACGGATCTCATCACCATCACCATCACTAAGCTTGGCTGTTTTGGC
GGATGAGAGAAGATTTTACGCCTGATACAGATTAAATCAGAACGCAGAAGCGGTCTGATAA
AACAGAATTTGCCTGGCGGCAGTAGCGCGGTGGTCCCACCTGACCCCATGCCGAACCTCAGA
AGTGAAACGCCGTAGCGCCGATGGTAGTGTGGGGTCTCCCCATGCGAGAGTAGGGAACCTGC
CAGGCATCAATAAAACGAAAA

>pRV5_C149mW_forward read, Lark Technologies (top strand)
GNAATGAGCTGTTGACAATTAATCATCGGCTCGTATAATGTGTGGAATTGTGAGCGGATAAC
AATTTACACAGGAAACAGAATTCATTAAGAGGAGAAATTAACCATGGAGCAGCCGAGGA
AGGCGGTGGTAGTGACGGGATTTGGCCCTTTTGGGGAACACACCGTGAACGCCAGTTGGATT
GCAGTTCAGGAGCTAGAAAAGCTAGGCCTTGGCGACAGCGTGGACCTGCATGTGTACGAGA
TTCCGGTTGAGTACCAAACAGTCCAGAGACTCATCCCCGCCCTGTGGGAGAAGCACAGTCCA
CAGCTGGTGGTGCATGTGGGGGTGTCAGGCATGGCGACCACAGTCACACTGGAGAAATGTG
GACACAACAAGGGCTACAAGGGGCTGGACAACCTGCCGCTTTTGGCCCCGGCTCCCAGTGCTGC
GTGGAGGACGGGCCTGAAAGCATTGACTCCATCATCGACATGGATGCTGTGTGCAAGCGAG
TCACCACGTTGGGCCTGGATGTGTGCGGTGACCATCTCGCAGGATGCCGGCAGATACCTCTGG
GACTTTACCTACTACACCTCTTTGTACCAGAGTCACGGTCGATCAGCCTTCGTCCACGTGCCC
CCACTGGGGAAGCCGTACAACGCGGACCAGCTGGGCAGGGCACTGAGAGCCATCATTGAGG
AGATGTTGGACCTCCTGGAGCAGTCAGAGGGCAAAATCAAATATTGCCACAAACACGGAT
TTCATCACCATCACCATCACTAAANCTNGGCTGTTTTNGGCGGATGAAANAAAAATTTTAGC
CCTGATTGAGAATTAATCAGAAANCCAGAAACGGGTCTGANAAAAANAAANTTTGCCTGG
GNGGNANNAACCCCGGGGGGTCCCCCNTGANCCCTTGCCNAAANTNAAAAATGAAACC
CCCCNCCCCCNATNGTATTTGNGGGG

>pRV5_C149mW_reverse read, Lark Technologies (top strand)
NAANNAACCNNGGTTTTAATNTCCGGGGGNCAAATTTTTGGGTGANGAANCNNNGNAAATT
GGGGAAGGGTGGGCGGGGGNAAANCNGCNAAATTNNGGTNGNCNNAGGGCCCTCCNGTTTG
GNAAAATTTTTGGCCCNACATAAAACGGTTTGGCAAAATTTTAAANAACNGTNGNNANTT
ATTATCGGCTCGTAAATTTGNGNANTNTAGCGGATAACNATTTACACAGGAACAGAATTC
ATAAAGAGGAGAAATTANCCATGANACGCCAGGAAGGCGGTGGTAGTGACGGGATTTGNC
CNTTTTGGGAACACACCGTGAACGCCAGTTGGANTGCAGTTCAGGAGCTAGAAAAGCTAGG
CCTTGGCGACAGCGTGGACCTGCATGTGTACGAGATTCCGGTTGAGTACCAAACAGTCCAGA
GACTCATCCCCGCCCTGTGGGAGAAGCACAGTCCACAGCTGGTGGTGCATGTGGGGGTGTC
GGCATGGCGACCACAGTCACACTGGAGAAATGTGGACACAACAAGGGCTACAAGGGGCTGG
ACAACCTGCCGCTTTTGGCCCCGGCTCCCAGTGCTGCGTGGAGGACGGGCCTGAAAGCATTGAC
TCCATCATCGACATGGATGCTGTGTGCAAGCGAGTCACCACGTTGGGCCTGGATGTGTGCGT
GACCATCTCGCAGGATGCCGGCAGATACCTCTGGGACTTTACCTACTACACCTCTTTGTACC
AGAGTCACGGTCGATCAGCCTTNGTCCACGTGCCCCCACTGGGGAAGCCGTACAACGCGGA
CCAGCTGGGCAGGGCACTGAGAGCCATCATTGAGGAGATGTTGGACCTCCTGGAGCAGTCA
GAGGGCAAAATCAACTATTGCCACAAACACGGATCTCATCACCATCACCATCACTAAGCTTG
GCTGTTNNGC

>pRV5_H168mD_forward read, Lark Technologies (top strand)

GNAATGAGCTGTTGACAATTAATCATCGGCTCGTATAATGTGTGGAATTGTGAGCGGATAAAC
AATTTACACAGGAAACAGAATTCATTAAAGAGGAGAAATTAACCATGGAGCAGCCGAGGA
AGAGCGGTGGTAGTGACGGGATTTGGCCCTTTTGGGGAACACACCGTGAACGCCAGTTGGA
TTGCAGTTCAGGAGCTAGAAAAGCTAGGCCTTGGCGACAGCGTGACCTGCATGTGTACGA
GATTCCGGTTGAGTACCAAACAGTCCAGAGACTCATCCCCGCCCTGTGGGAGAAGCACAGTC
CACAGCTGGTGGTGCATGTGGGGGTGTCAGGCATGGCGACCACAGTCACACTGGAGAAATG
TGGACACAACAAGGGCTACAAGGGGCTGGACAACCTGCCGCTTTTGCCCCGGCTCCCAGTGCT
GCGTGGAGGACGGGCCTGAAAGCATTGACTCCATCATCGACATGGATGCTGTGTGCAAGCG
AGTCACCACGTTGGGCCTGGATGTGTGCGGTGACCATCTCGCANGATGCCGGCAGATATCTCT
GCGACTTTACCTACTACACCTCTTTGTACCAGAGTCACGGTCGATCGGNCNTTCGTCGACGT
GCCCCACTGGGGAAGCCGTACACGCGGACCAGCTGGGCAGGGCACTGANANCATCATGGA
GGAAATTTGGNNCTNCTGGAGCATTACAGAGGNCAAANTCAATTTTGGCCNAAAAANCGGNTT
TTNTTNCNTCNCNTTNTTNAAGTTNGNNTGTTTNGGNGGANNAAAAAANTTTTNNCCCTGT
TCNAAANTNAANNNAANNCNCNAAANCGGNTTTNAAAAAANTTTTCTTGGGGGNNA
NACCCGGGGGNCCCCCTNGNCCCTNGGCCAAANTAAAAAANNNCCCCCNCCCCCNNG
GGNGTGGGGGGGGGNNCCCCCNCCCANAAAAAN

>pRV5_H168mD_reverse read, Lark Technologies (top strand)

AAANGAAANNNGGNTTTNNANNNGGGCGGGGNCCCNANNTTTTNGGGGTNNGGANNCCN
TNTGNAAAATTTGGNTANGNTTNNCCAAGGTTTTAAACCNNGGAAAAANTNGTTGGNNCA
ANGGGGCANTCCNNTTNGNAAANNTTTTTTGGCCCNAAATTNAAACGGNTTTNGCAAAAA
NTTTNAAATAAGCNNTTGNCAANTAATCATCGGCTCGTAAAAGNNGGNAATNNNNANCGGAT
ANCAATTTCCCCCAGGAANCAGAATTCATTAAAGAGGAGAAATTANCCATGGAGCAGCCNA
GNAAGGCGGTGGTAGTGACGGGATTTGNCCNTTTNGGGAACACACCGTGAACGCCAGTTGG
ATTGCAGTTCAGGAGCTAGAAAAGCTAGGCCTTGGCGACAGCGTGACCTGCATGTGTACG
AGATTCCGGTTGAGTACCAAACAGTCCAGAGACTCATCCCCGCCCTGTGGGAGAAGCACAG
TCCACAGCTGGTGGTGCATGTGGGGGTGTCAGGCATGGCGACCACAGTCACACTGGAGAAA
TGTGGACACAACAAGGGCTACAAGGGGCTGGACAACCTGCCGCTTTTGCCCCGGCTCCCAGTG
CTGCGTGGAGGACGGGCCTGAAAGCATTGACTCCATCATCGACATGGATGCTGTGTGCAAGC
GAGTCACCACGTTGGGCCTGGATGTGTGCGGTGACCATCTCGCAGGATGCCGGCAGATATCTC
TGCGACTTTACCTACTACACCTCTTTGTACCAGAGTCACGGTCGATCGGCCTTNGTCGACGTG
CCCCACTGGGGAAGCCGTACAACGCGGACCAGCTGGGCAGGGCACTGAGAGCCATCATTG
AGGAGATGTTGGACCTCCCTGGAGCAGTCAGAGGGCAAATCAACTATTGCCACAAACACG
GATCTCATCACCATCACCATCACTAAGNCTTGGCTGTTNTGGCGGNATAAAA

>pRV5_F13mW_forward read, MWG (top strand)

CAGCCGAGGAAGGCGGTGGTAGTGACGGGATGGGGCCCTTTTGGGGAACACACCGTGAACG
CCAGTTGGATTGCAGTTCAGGAGCTAGAAAAGCTAGGCCTTGGCGACAGCGTGACCTGCA
TGTGTACGAGATTCCGGTTGAGTACCAAACAGTCCAGAGACTCATCCCCGCCCTGTGGGAGA
AGCACAGTCCACAGCTGGTGGTGCATGTGGGGGTGTCAGGCATGGCGACCACAGTCACACT
GGAGAAATGTGGACACAACAAGGGCTACAAGGGGCTGGACAACCTGCCGCTTTTGCCCCGGC
TCCCAGTGCTGCGTGGAGGACGGGCCTGAAAGCATTGACTCCATCATCGACATGGATGCTGT
GTGCAAGCGAGTCACCACGTTGGGCCTGGATGTGTGCGGTGACCATCTCGCAGGATGCCGGCA
GATATCTCTGCGACTTTACCTACTACACCTCTTTGTACCAGAGTCACGGTCGATCAGCCTTCG
TCCACGTGCCCCCACTGGGGAAGCCGTACAACGCGGACCAGCTGGGCAGGGCACTGAGAGC
CATCATTGAGGAGATGTTGGACCTCCTGGAGCAGTCAGAGGGCAAATCAACTATTGCCAC
AAACACGGATCTCATCACCATCACCATCACTAAGCTTGGCTGTTTTGGCGGATGAGAGAAGA
TTTTACGCTGATACAGATTAAATCAGAACGCAGAAGCGGTCTGATAAAACAGAATTTGCCT
GGCGGCAGTAGCGCGGTGGTCCCACCTGAC

>pRV5_F13mY.P175mS_forward read, MWG (top strand)
GGAGCAGCCGAGGAAGGCGGTGGTAGTGACGGGATATGGCCCTTTTGGCGAACACACCGTG
AACGCCAGTTGGATTGCAGTTCAGGAGCTAGAAAAGCTAGGCCTTGGCGACAGCGTGGACC
TGCATGTGTACGAGATTCCGGTTGAGTACCAAACAGTCCAGAGACTCATCCCCGCCCTGTGG
GAGAAACACAGTCCACAGCTGGTGGTGCATGTGGGGGTGTCAGGCATGGCGACCACAGTCA
CACTGGAGAAATGTGGACACAACAAGGGCTACAAGGGGCTGGACAAC TGCCGCTTTTGCCC
CGGCTCCCAGTGCTGCGTGGAGGACGGGCCTGAAAGCATTGACTCCATCATCGACATGGATG
CTGTGTGCAAGCGAGTCACCACGTTGGGCCTGGATGTGTGCGGTGACCATCTCGCAGGATGCC
GGCAGATATCTCTGCGACTTTACCTACTACACCTCTTTGTACCAGAGTCACGGTCGATCAGCC
TTCGTCCACGTGCCCCCACTGGGGAAGTCGTACAACGCGGACCAGCTGGGCAGGGCACTGA
GAGCCATCATTGAGGAGATGTTGGACCTCCTGGAGCAGTCAGAGGGCAAATCAACTATTG
CCACAAACACGGATCTCATCACCATCACCATCACTAAGCTTGGCTGTTTTGGCGGATGAGAG
AAGATTTTCAGCCTGATACAGATTAATCAGAACGCAGAAGCGGTCTGATAAAACAGAATT
TGCTTGGCGGACAGTAGCGCGGTGGTCCCACCTGACCCC

>pRV5_F13mY.P175mS_forward read, MWG (top strand) repeat
AGCCGAGGAAGGCGGTGGTAGTGACGGGATATGGCCCTTTTGGCGAACACACCGTGAAACGC
CAGTTGGATTGCAGTTCAGGAGCTAGAAAAGCTAGGCCTTGGCGACAGCGTGGACCTGCAT
GTGTACGAGATTCCGGTTGAGTACCAAACAGTCCAGAGACTCATCCCCGCCCTGTGGGAGAA
GCACAGTCCACAGCTGGTGGTGCATGTGGGGGTGTCAGGCATGGCGACCACAGTCACACTG
GAGAAATGTGGACACAACAAGGGCTACAAGGGGCTGGACAAC TGCCGCTTTTGCCCCGGCT
CCCAGTGCTGCGTGGAGGACGGGCCTGAAAGCATTGACTCCATCATCGACATGGATGCTGTG
TGCAAGCGAGTCACCACGTTGGGCCTGGATGTGTGCGGTGACCATCTCGCAGGATGCCGGCAG
ATATCTCTGCGACTTTACCTACTACACCTCTTTGTACCAGAGTCACGGTCGATCAGCCTTCGT
CCACGTGCCCCCACTGGGGAAGTCGTACAACGCGGACCAGCTGGGCAGGGCACTGAGAGCC
ATCATTGAGGAGATGTTGGACCTCCTGGAGCAGTCAGAGGGCAAATCAACTATTGCCACA
AACACGGATCTCATCACCATCACCATCACTAAGCTTGGCTGTTTTGGCGGATGAGAGAAGAT
TTTCAGCCTGATACAGATTAATCAGAACGCAGAAGCGGTCT

>pRV5_F13mY_forward read, MWG (top strand)
GCGGTGGTAGTGACGGGATATGGCCCTTTTGGCGAACACACCGTGAAACGCCAGTTGGATTGC
AGTTCAGGAGCTAGAAAAGCTAGGCCTTGGCGACAGCGTGGACCTGCATGTGTACGAGATT
CCGGTTGAGTACCAAACAGTCCAGAGACTCATCCCCGCCCTGTGGGAGAAAGCACAGTCCAC
AGCTGGTGGTGCATGTGGGGGTGTCAGGCATGGCGACCACAGTCACACTGGAGAAATGTGG
ACACAACAAGGGCTACAAGGGGCTGGACAAC TGCCGCTTTTGCCCCGGCTCCCAGTGCTGCG
TGGAGGACGGGCCTGAAAGTATTGACTCCATCATCGACATGGATGCTGTGTGCAAGCGAGTC
ACCACGTTGGGCCTGGATGTGTGCGGTGACCATCTCGCAGGATGCCGGCAGATATCTCTGCGA
CTTTACCTACTACACCTCTTTGTACCAGAGTCACGGTCGATCAGCCTTCGTCCACGTGCCCC
ACTGGGGAAGCCGT

>pRV5_F13mL_forward read, MWG (top strand)
TGGATTGCAGTTCATGAGCTAGAAAAGCTAGGCCTTGGCGACAGCGTGGACCTGCATGTGTGA
CGAGATTCCGGTTGAGTACCAAACAGTCCAGAGACTCATCCCCGCCCTGTGGGAGAAAGCAC
AGTCCACAGCTGGTGGTGCATGTGGGGGTGTCAGGCATGGCGACCACAGTCACACTGGAGA
AATGTGGACACAACAAGGGCTACAAGGGGCTGGACAAC TGCCGCTTTTGCCCCGGCTCCCA
GTGCTGCGTGGAGGACGGGCCTGAAAGCATTGACTCCATCATCGACATGGATGCTGTGTGCA
AGCGAGTCACCACGTTGGGCCTGGATGTGTGCGGTGACCATCT

>pRV5_F13mL.F16mL_forward read, Qiagen (top strand)
CAATTCTGAATGAGCTGTTGACAATTAATCATCGGCTCGTATAATGTGTGGAATTGTGAGCG
GATAACAATTTACACAGGAAACAGAATTCATTAAAGAGGAGAAATTAACCATGGAGCAGC
CGAGGAAGGCGGTGGTAGTGACGGGATTGGGCCCTTTGGGGGAACACACCGTGAAACGCCAG
TTGGATTGCAGTTCAGGAGCTAGAAAAGCTAGGCCTTGCGGACAGCGTGACCTGCATGTGT
ACGAGATTCCGTTGAGTACCAAACAGTCCAGAGACTCATCCCCGCCCTGTGGGAGAAGCA
CAGTCCACAGCTGGTGGTGCATGTGGGGGTGTCAGGCATGGCGACCACAGTCACACTGGAG
AAATGTGGACACAACAAGGGCTACAAGGGGCTGGACAACCTGCCGCTTTTGCCCCGGCTCCC
AGTGCTGCGTGGAGGACGGGCCTGAAAGCATTGACTCCATCATCGACATGGATGCTGTGTGC
AAGCGAGTCACCACGTTGGGCCTGGATGTGTGCGGTGACCATCTCGCAGGATGCCGGCAGAT
ATCTCTGCGACTTTACCTACTACACCTCTTTGTACCAGAGTCACGGTCGATCAGCCTTCGTCC
ACGTGCCCCCACTGGGGAAGCCGTACAACGCGGACCAGCTGGGCAGGGCACTGAGAGCCAT
CATTGAGGAGATGTTGGACCTCCTGGAGCAGTCAGAGGGCAAAATCAACTATTGCCACAAA
CACGGATCTCATCACCATCACCATCACTAAGCTTGGCTGTTTTGGCGGATGAGAGAAGTTTT
CAGCCTGATACAGATTAATCAGAACGCAGAAGCGGTCTGGATAAACAGATTTGCCTGGCGT

>pRV5_F16mL_forward read, MWG (top strand)
TGACGGGATTTGGCCCTCTGGGCGAACACACCGTGAACGCCAGTTGGATTGCAGTTCAGGAG
CTAGAAAAGCTAGGCCTTGCGGACAGCGTGGACCTGCATGTGTACGAGATTCCGGTTGAGTA
CCAAACAGTCCAGAGACTCATCCCCGCCCTGTGGGAGAAGCACAGTCCACAGCTGGTGGTG
CATGTGGGGGTGTCAGGCATGGCGACCACAGTCACACTGGAGAAATGTGGACACAACAAGG
GCTACAAGGGGCTGGACAACCTGCCGCTTTTGCCCCGGCTCCCACTGCTGCGTGGAGGACGGG
CCTGAAAGCATTGACTCCATCATCGACATGGATGCTGTGTGCAAGCGAGTCACCACGTTGGG
CCTGGATGTGTGCGGTGACCATCTCGCAGGATGCCGGCAGATATCTCTGCGACTTTACCTACT
ACACCTCTTTGTACCAGAGTCACGGTCGATCAGCCTTCGTCCACGTGCCCCCACTGGGGAAG
CCGTACAACGCGGACCAGCTGGGCAGGGCACTGAGAGCCATCATTGAGGAGATGTTGGACC
TCCTGGAGCAGTCAGAGGGCAAAATCAACTATTGCCACAAACACGGATCTCATCACCATCAC
CATCACTAAGCTTGGCTGTTTTGGCGGATGAGAGAAGATTTTCAGCCTGATACAGATTAAAT
CAGAACGCAGAAGCGGTCTGATAAAACAGAATTTGCCTGGCGGCAGTAGCGCGGTGGTCCC
ACCTGACCCCATGCCGAACCTCAGAAGTGAAACGCCGTAGCGCCGATGGT

>pRV5_F16mL_reverse read, MWG (top strand)
GCAGGTTCGTAAATCACTGCATAATTCGTGTGCGCTCAAGGCGCACTCCCGTTCTGGATAATGT
TTTTTGCGCCGACATCATAACGGTTCTGGCAAATATTCTGAAATGAGCTGTTGACAATTAATC
ATCGGCTCGTATAATGTGTGGAATTGTGAGCGGATAACAATTTACACAGGAAACAGAATTC
ATTAAAGAGGAGAAATTAACCATGGAGCAGCCGAGGAAGGCGGTGGTAGTGACGGGATTTG
GCCCTCTGGGCGAACACACCGTGAACGCCAGTTGGATTGCAGTTCAGGAGCTAGAAAAGCT
AGGCCTTGCGGACAGCGTGGACCTGCATGTGTACGAGATTCCGGTTGAGTACCAAACAGTCC
AGAGACTCATCCCCGCCCTGTGGGAGAAGCACAGTCCACAGCTGGTGGTGCATGTGGGGGT
GTCAGGCATGGCGACCACAGTCACACTGGAGAAATGTGGACACAACAAGGGCTACAAGGGG
CTGGACAACCTGCCGCTTTTGCCCCGGCTCCCACTGCTGCGTGGAGGACGGGCCCTGAAAGCAT
TGACTCCATCATCGACATGGATGCTGTGTGCAAGCGAGTCACCACGTTGGGCCTGGATGTGT
CGGTGACCATCTCGCAGGATGCCGGCAGATATCTCTGCGACTTTACCTACTACACCTCTTTGT
ACCAGAGTCACGGTCGATCAGCCTTCGTCCACGTGCCCCCACTGGGGAAGCCGTACAACGCG
GACCAGCTGGGCAGGGCACTGAGAGCCATCATTGAGGAGATGTTGGACCTCCTGGAGCAGT
CAGAGGGCAAAATCAACTATTGCCACAAACACGGATCTCATCACCATCACCATCACTAAGCT
TGGCTGTTTGGCG

>pRV5_F16mY_forward read, MWG (top strand)
CGCCAGTTGGATTGCAGTTCAGGAGCTAGAAAAGCTAGGCCTTGGCGACAGCGTGGACCTG
CATGTGTACGAGATTCCGGTTGAGTACCAAACAGTCCAGAGACTCATCCCCGCCCTGTGGGA
GAAGCACAGTCCACAGCTGGTGGTGCATGTGGGGGTGTCAGGCATGGCGACCACAGTCACA
CTGGAGAAATGTGGACACAACAAGGGCTACAAGGGGCTGGACAACCTGCCGCTTTTGCCCCG
GCTCCCACTGCTGCGTGGAGGACGGGCCTGAAAGCATTGACTCCATCATCGACATGGATGCT
GTGTGCAAGCGAGTCACCACGTTGGGCCTGGATGTGTGCGGTGACCATCTCGCAGGATGCCGG
CAGATATCTCTGCGACTTTACCTACTACACCTCTTTGTACCAGAGTACACGGTCGATCAGCCTT
CGTCCACGTGCCCCCACTGGGGAAGCCGTACAACGCGGACCAGCTGGGCAGGGCACTGAGA
GCCATCATTGAGGAGATGTTGGACCTCCTGGAGCAGTCAGAGGGCAAAATCAACTATTGCC
ACAAACACGGATCTCATCACCATCACCATCACTAAGCTTGGCTGTTTGGCGGATGAGAGAA
GATTTTCAGCCTGATACAGATTAAATCAGAACGCAGAAGCGGTCTGATAAAACAGAATTTGC
CT

>pRV5_F16mY_reverse read, MWG (top strand)
TTTTTGCGCCGACATCATAACGGTTCTGGCAAATATTCTGAAATGAGCTGTTGACAATTAATC
ATCGGCTCGTATAATGTGTGGAATTGTGAGCGGATAACAATTTACACAGGAAACAGAATTC
ATTAAGAGGAGAAATTAACCATGGAGCAGCCGAGGAAGGCGGTGGTAGTGACGGGATTTG
GCCCTTATGGCGAACACACCGTGAACGCCAGTTGGATTGCAGTTCAGGAGCTAGAAAAGCT
AGGCCTTGGCGACAGCGTGGACCTGCATGTGTACGAGATTCCGGTTGAGTACCAAACAGTCC
AGAGACTCATCCCCGCCCTGTGGGAGAAGCACAGTCCACAGCTGGTGGTGCATGTGGGGGT
GTCAGGCATGGCGACCACAGTCACACTGGAGAAATGTGGACACAACAAGGGCTACAAGGGG
CTGGACAACCTGCCGCTTTTGCCCCGGCTCCCACTGCTGCGTGGAGGACGGGCCTGAAAGCAT
TGACTCCATCATCGACATGGATGCTGTGTGCAAGCGAGTCACCACGTTGGGCCTGGATGTGT
CGGTGACCATCTCGCAGGATGCCGGCAGATATCTCTGCGACTTTACCTACTACACCTCTTTGT
ACCAGAGTCACGGTCGATCAGCCTTCGTCCACGTGCCCCCACTGGGGAAGCCGTACAACGCG
GACCAGCTGGGCAGGGCACTGAGAGCCATCATTGA

>pRV5_Y147mF_forward read, MWG (top strand)
AGCCGAGGAAGGCGGTGGTAGTGACGGGATTTGGCCCTTTTGGGGAACACACCGTGAACGC
CAGTTGGATTGCAGTTCAGGAGCTAGAAAAGCTAGGCCTTGGCGACAGCGTGGACCTGCAT
GTGTACGAGATTCCGGTTGAGTACCAAACAGTCCAGAGACTCATCCCCGCCCTGTGGGAGAA
GCACAGTCCACAGCTGGTGGTGCATGTGGGGGTGTCAGGCATGGCGACCACAGTCACACTG
GAGAAATGTGGACACAACAAGGGCTACAAGGGGCTGGACAACCTGCCGCTTTTGCCCCGGCT
CCCAGTGCTGCGTGGAGGACGGGCCTGAAAGCATTGACTCCATCATCGACATGGATGCTGTG
TGCAAGCGAGTCACCACGTTGGGCCTGGATGTGTGCGGTGACCATCTCGCAGGATGCCGGCAG
ATTTCTCTGCGACTTTACCTACTACACCTCTTTGTACCAGAGTCACGGTCGATCAGCCTTCGT
CCACGTGCCCCCACTGGGGAAGCCGTACAACGCGGACCAGCTGGGCAGGGCACTGAGAGCC
ATCATTGAGGAGATGTTGGACCTCCTGGAGCAGTCAGAGGGCAAAATCAACTATTGCCACA
AACACGGATCTCATCACCATCACCATCACTAAGCTTGGCTGTTTGGCGGATGAGAGAAGAT
TTTCAGCCTGATACAGATTAAATCAGAACGCAGAAGCGGTCTGATAAAACAGAATTTGCCTG
GCGGCAGTAGCGCGGTGGTCCCACCTGACCCCATGCCGAACCTCAGAAGTGAAACGCCGTAG
CGCCGATGGTAGTGTGGGGTCTCCCCATGCGAGAGTAGGGAACCTGCCAGGCATCAAATAAA
A

>pRV5_Y147mF_reverse read, MWG (top strand)
TGCATAAATTCTGTGCTCAAGGCGCACTCCCGTTCTGGATAATGTTTTTTCGCGCCGACATC
ATAACGGTTCTGGCAAATATTCTGAAATGAGCTGTTGACAATTAATCATCGGCTCGTATAAT
GTGTGGAATTGTGAGCGGATAACAATTTACACAGGAAACAGAATTCATTAAAGAGGAGAA
ATTAACCATGGAGCAGCCGAGGAAGGCGGTGGTAGTGACGGGATTTGGCCCTTTTGGGGAA
CACACCGTGAACGCCAGTTGGATTGCAGTTCAGGAGCTAGAAAAGCTAGGCCTTGGCGACA
GCGTGGACCTGCATGTGTACGAGATTCCGGTTGAGTACCAAACAGTCCAGAGACTCATCCCC
GCCCTGTGGGAGAAGCACAGTCCACAGCTGGTGGTGCATGTGGGGGTGTCAGGCATGGCGA
CCACAGTCACACTGGAGAAATGTGGACACAACAAGGGCTACAAGGGGCTGGACAACCTGCCG
CTTTTGCCCCGGCTCCCAGTGCTGCGTGGAGGACGGGCCTGAAAGCATTGACTCCATCATCG
ACATGGATGCTGTGTGCAAGCGAGTCACCACGTTGGGCCTGGATGTGTGCGGTGACCATCTCG
CAGGATGCCGGCAGATTTCTCTGCGACTTTACCTACTACACCTCTTTGTACCAGAGTCACGGT
CGATCAGCCTTCGTCCACGTGCCCCCACTGGGGAAGCCGTACAACGCGGACCAGCTGGGCA
GGGCACTGAGAGCCATCATTGAGGAGATGTTGGACCTCCTGGAGCAGTCAGAGGGCAAAAT
CAACTATTGCCACAAACACGGATCTCATCACCATCACCATCACTAAGCTTGGCTGTTTGGCG

>pRV5_D97mE.N98C99mM_forward read, MWG (top strand)

GAGCAGCCGAGGAAGGCGGTGGTAGTGACGGGATTTGGCCCTTTTGGGGAACACACCGTGA
ACGCCAGTTGGATTGCAGTTCAGGAGCTAGAAAAGCTAGGCCTTGGCGACAGCGTGGACCT
GCATGTGTACGAGATTCCGGTTGAGTACCAAACAGTCCAGAGACTCATCCCCGCCCTGTGGG
AGAAGCACAGTCCACAGCTGGTGGTGCATGTGGGGGTGTCAGGCATGGCGACCACAGTCAC
ACTGGAGAAATGTGGACACAACAAGGGCTACAAGGGGCTGGAAATGCGCTTTTGGCCCGGA
TCCCAGTGCTGCGTGGAGGACGGGCCTGAAAGCATTGACTCCATCATCGACATGGATGCTGT
GTGCAAGCGAGTCACCACGTTGGGCCTGGATGTGTGCGGTGACCATCTCGCAGGATGCCGGCA
GATATCTCTGCGACTTTACCTACTACACCTCTTTGTACCAGAGTCACGGTCGATCAGCCTTCG
TCCACGTGCCCCCACTGGGGAAGCCGTACAACGCGGACCAGCTGGGCAGGGCACTGAGAGC
CATCATTGAGGAGATGTTGGACCTCCTGGAGCAGTCAGAGGGGCAAAATCAACTATTGCCAC
AAACACGGATCTCATCACCATCACCATCACTAAGCTTGGCTGTTTTGGCGGATGAGAGAAGA
TTTTTCAGCCTGATACAGATTAAATCAGAACGCAGAAGCGGTCTGATAAAAACAGAATTTGCC
TGGCGGCAGTAGCGCGGTGGTCCCCACCTGACCCCATGCCGAAACTC

>pRV5_D97mE_forward read, MWG (top strand)

AGCCGAGGAAGGCGGTGGTAGTGACGGGATTTGGCCCTTTTGGGGAACACACCGTGAACGC
CAGTTGGATTGCAGTTCAGGAGCTAGAAAAGCTAGGCCTTGGCGACAGCGTGGACCTGCAT
GTGTACGAGATTCCGGTTGAGTACCAAACAGTCCAGAGACTCATCCCCGCCCTGTGGGAGAA
GCACAGTCCACAGCTGGTGGTGCATGTGGGGGTGTCAGGCATGGCGACCACAGTCACACTG
GAGAAATGTGGACACAACAAGGGCTACAAGGGGCTGGAGAACTGCCGCTTTTGGCCCGGAT
CCCAGTGCTGCGTGGAGGACGGGCCTGAAAGCATTGACTCCATCGACATGGATGCTGTG
TGCAAGCGAGTCACCACGTTGGGCCTGGATGTGTGCGGTGACCATCTCGCAGGATGCCGGCAG
ATATCTCTGCGACTTTACCTACTACACCTCTTTGTACCAGAGTCACGGTCGATCAGCCTTCGT
CCACGTGCCCCCACTGGGGAAGCCGTACAACGCGGACCAGCTGGGCAGGGCACTGAGAGCC
ATCATTGAGGAGATGTTGGACCTCCTGGAGCAGTCAGAGGGCAAAATCAACTATTGCCACA
AACACGGATCTCATCACCATCACCATCACTAAGCTTGGCTGTTTTGGCGGATGAGAGAAGAT
TTTCAGCCTGATACAGATTAAATCAGAACGCAGAAGCGGTCTGATAAAAACAGAATTTGCCTG
GCGGCAGTAGCGCGGTGGTCCCACCTGACCCCATGCCGAACTCAGAAGTGAAACGCCGTAG
CGCCGATGGTAGTGTGGGGTCTCCCCATGCGAGAGTAGGGGAACTGCCAGGCATCAAATAA
ACGAAAGGCTCAGTCGAAAGACTGGGCCTTTTCGTTTTATCTGTTGGTTGTCGGTGAACGCT
CTCCTGAGTAGGACAAATCCGCCCCGAGCGGATTTGAACG

>pRV5_D97mE_reverse read, MWG (top strand)

GCCGACATCATAACGGTTCCTGGCAAAATTTCTGAAATGAGCTGTTGACAATTAATCATCGGC
TCGTATAATGTGTGGAATTGTGAGCGGATAACAATTTACACAGGAAACAGAATTCATTTAA
GAGGAGAAATTAACCATGGAGCAGCCGAGGAAGGCGGTGGTAGTGACGGGATTTGGCCCTT
TTGGGGAACACACCGTGAACGCCAGTTGGATTGCAGTTCAGGAGCTAGAAAAGCTAGGCCT
TGGCGACAGCGTGGACCTGCATGTGTACGAGATTCCGGTTGAGTACCAAACAGTCCAGAGA
CTCATCCCCGCCCTGTGGGAGAAGCACAGTCCACAGCTGGTGGTGCATGTGGGGGTGTCAGG
CATGGCGACCACAGTCACACTGGAGAAATGTGGACACAACAAGGGCTACAAGGGGCTGGAG
AACTGCCGCTTTTGGCCCGGATCCCAGTGCTGCGTGGAGGACGGGCCTGAAAGCATTGACTC
CATCATCGACATGGATGCTGTGTGCAAGCGAGTCACCACGTTGGGCCTGGATGTGTGCGTGA
CCATCTCGCAGGATGCCGGCAGATATCTCTGCGACTTTACCTACTACACCTCTTTGTACCAGA
GTCACGGTCGATCAGCCTTCGTCCACGTGCCCCCACTGGGGAAGCCGTACAACGCGGACCAG
CTGGGCAGGGCACTGAGAGCCATCATTGAGAGAGATGTTGGACCTCCTGG

>pRV5_N98mQ_forward read, MWG (top strand)

AGCAGCCGAGGAAGGCGGTGGTAGTGACGGGATTTGGCCCTTTTGGGGAACACACCGTGAA
CGCCAGTTGGATTGCAGTTCAGGAGCTAGAAAAGCTAGGCCTTGGCGACAGCGTGGACCTG
CATGTGTACGAGATTCCGGTTGAGTACCAAACAGTCCAGAGACTCATCCCCGCCCTGTGGGA
GAAGCACAGTCCACAGCTGGTGGTGCATGTGGGGGTGTCAGGCATGGCGACCACAGTCACA
CTGGAGAAATGTGGACACAACAAGGGCTACAAGGGGCTGGACCAGTGCCGCTTTTGGCCCG
GATCCCAGTGCTGCGTGGAGGACGGGCTGAAAGCATTGACTCCATCATCGACATGGATGCT
GTGTGCAAGCGAGTCACCACGTTGGGCTGGATGTGTCGGTGACCATCTCGCAGGATGCCGG
CAGATATCTCTGCGACTTTACCTACTACACCTCTTTGTACCAGAGTCACGGTCGATCAGCCTT
CGTCCACGTGCCCCCACTGGGGAAGCCGTACAACGCGGACCAGCTGGGCAGGGCACTGAGA
GCCATCATTGAGGAGATGTTGGACCTCCTGGAGCAGTCAGAGGGGCAAAATCAACTATTGCC
ACAAACACGGATCTCATCACCATCACCATCACTAAGCTTGGCTGTTTTGGCGGATGAGAGAA
GATTTTCAGCCTGATACAGATTAAATCAGAACGCAGAAGCGGTCTGATAAAACAGAATTTGC
CTGGCGGCAGTAGCGCGGTGGTCCCACCTGACCCCATGCCGAACCTCAGAAGTGAAACGCCG
TAGCGCCGATGGTAGTGTGGGGTCTCCCCATGCGAGAGTAGGGAAGTCCAGGCATCAAAT
AAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTTTTATCTGTGGTTGTGCGGTGAACGC
TCTCCTGAGTAGGACAA

>pRV5_N98mQ_reverse read, MWG (top strand)

TGGATAATGTTTTTTCGCCCGACATCATAACGGTTCTGGCAAATATTCTGAAATGAGCTGTTG
ACAATTAATCATCGGCTCGTATAATGTGTGGAATTGTGAGCGGATAACAATTTACACAGGA
AACAGAATTATTAAAGAGGAGAAATTAACCATGGAGCAGCCGAGGAAGGCGGTGGTAGTG
ACGGGATTTGGCCCTTTTGGGGAACACACCGTGAACGCCAGTTGGATTGCAGTTCAGGAGCT
AGAAAAGCTAGGCCTTGGCGACAGCGTGGACCTGCATGTGTACGAGATTCCGGTTGAGTAC
CAAACAGTCCAGAGACTCATCCCCGCCCTGTGGGAGAAGCACAGTCCACAGCTGGTGGTGC
ATGTGGGGGTGTCAGGCATGGCGACCACAGTCACACTGGAGAAATGTGGACACAACAAGGG
CTACAAGGGGCTGGACCAGTGCCGCTTTTGGCCCGGATCCCAGTGCTGCGTGGAGGACGGGC
CTGAAAGCATTGACTCCATCATCGACATGGATGCTGTGTGCAAGCGAGTCACCACGTTGGGC
CTGGATGTGTGCGTGACCATCTCGCAGGATGCCGGCAGATATCTCTGCGACTTTACCTACTA
CACCTCTTTGTACCAGAGTCACGGTCGATCAGCCTTCGTCCACGTGCCCCCACTGGGGAAGC
CGTACAACGCGGACCAGCTGGGCAGGGCACTGAGAGCCATCATTGAGGAGATGTTGGACCT
CCTGGAGCAGTCAGAGGGCAAAATCAACTATTGCCACAAACACGGATCTCATCACCATCAC
CATCACTAAGCTTGGCTGTTTGGCG

>pRV5_N98mV_forward read, MWG (top strand)

TAACCATGGGAGCAGCCGAGGAAGGCGGTGGTAGTGACGGGATTTGGCCCTTTTGGGGAAC
ACACCGGTGAACGCCAGTTGGATTGCAGTTCAGGAGCTAGAAAAGCTAGGCCTTGGCGACA
GCGTGGACCTGCATGTGTACGAGATTCCGGTTGAGTACCAAACAGTCCAGAGACTCATCCCC
GCCCTGTGGGAGAAGCACAGTCCACAGCTGGTGGTGCATGTGGGGGTGTCAGGCATGGCGA
CCACAGTCACACTGGAGAAATGTGGACACAACAAGGGCTACAAGGGGCTGGACGTCTGCCG
CTTTTGGCCCGGATCCCAGTGCTGCGTGGAGGACGGGCCTGAAAGCATTGACTCCATCATCG
ACATGGATGCTGTGTGCAAGCGAGTCACCACGTTGGGCCTGGATGTGTGCGGTGACCATCTCG
CAGGATGCCGGCAGATATCTCTGCGACTTTACCTACTACACCTCTTTGTACCAGAGTCACGG
TCGATCAGCCTTCGTCCACGTGCCCCCACTGGGGAAGCCGTACAACGCGGACCAGCTGGGCA
GGGCACTGAGAGCCATCATTGAGGAGATGTTGGACCTCCTGGAGCAGTCAGAGGGCAAAAT
CAACTATTGCCACAAACACGGATCTCATCACCATCACCATCACTAAGCTTGGCTGTTTTGGC
GGATGAGAGAAGATTTTCAGCCTGATACAGATTAAATCAGAACGCAGAAGCGGTCTGATAA
AACAGAATTTGCCTGGCGGCAGTAGCGCGGTGGTCCCACCTGACCCCATGCCGAACCTCAGA
AGTGAAACGCCGTAGCGCCGATGGTAGTGTGGGGTCTCCCCATGCGAG

>pRV5_N98mV_reverse read, MWG (top strand)

TGCATAATTCGTGTCGCTCAAGGCGCACTCCCGTTCTGGATAATGTTTTTGGCCCGACATCA
TAACGGTTCTGGCAAATATTCTGAAATGAGCTGTTGACAATTAATCATCGGCTCGTATAATG
TGTGGAATTGTGAGCGGATAACAATTTACACAGGAAACAGAATTCATTAAGAGGAGAGAA
TTAACCATGGAGCAGCCGAGGAAGGCGGTGGTAGTGACGGGATTTGGCCCTTTTGGGGAAC
ACACCGTGAACGCCAGTTGGATTGCAGTTCAGGAGCTAGAAAAGCTAGGCCTTGGCGACAG
CGTGGACCTGCATGTGTACGAGATTCCGTTGAGTACCAAACAGTCCAGAGACTCATCCCCG
CCCTGTGGGAGAAGCACAGTCCACAGCTGGTGGTGCATGTGGGGGTGTCAGGCATGGCGAC
CACAGTCACACTGGAGAAATGTGGACACAACAAGGGCTACAAGGGGCTGGACGTCTGCCGC
TTTTGCCCGGATCCAGTGCTGCGTGGAGGACGGGCCTGAAAGCATTGACTCCATCATCGA
CATGGATGCTGTGTGCAAGCGAGTCACCACGTTGGGCCTGGATGTGTGCGGTGACCATCTCGC
AGGATGCCGGCAGATATCTCTGCGACTTTACCTACTACACCTCTTTGTACCAGAGTCACGGT
CGATCAGCCTTCGTCCACGTGCCCCCACTGGGGAAGCCGTACAACGCGGACCAGCTGGGCA
GGGCACTGAGAGCCATCATTGAGGAGATGTTGGACCTCCTGGAGCAGTCAGAGGGCAAAAT
CAACTATTGCCACAAACACGGATCTCATCACCATCACCATCACTAAGCTTGGCTG

>pRV5_SID115mAYF_forward read, MWG (top strand)

ATTAACCATGGAGCAGCCGAGGAAGGCGGTGGTAGTGACGGGATTTGGCCCTTTTGGGGAA
CACACCGTGAACGCCAGTTGGATTGCAGTTCAGGAGCTAGAAAAGCTAGGCCTTGGCGACA
GCGTGGACCTGCATGTGTACGAGATTCCGTTGAGTACCAAACAGTCCAGAGACTCATCCCC
GCCCTGTGGGAGAAGCACAGTCCACAGCTGGTGGTGCATGTGGGGGTGTCAGGCATGGCGA
CCACAGTCACACTGGAGAAATGTGGACACAACAAGGGCTACAAGGGGCTGGACAACACTGCCG
CTTTTGGCCCGGCTCCAGTGCTGCGTGGAGGACGGGCCTGAGGCCTATTTCTCCATCATCG
ACATGGATGCTGTGTGCAAGCGAGTCACCACGTTGGGCCTGGATGTGTGCGGTGACCATCTCG
CAGGATGCCGGCAGATATCTCTGCGACTTTACCTACTACACCTCTTTGTACCAGAGTCACGG
TCGATCAGCCTTCGTCCACGTGCCCCCACTGGGGAAGCCGTACAACGCGGACCAGCTGGGCA
GGGCACTGAGAGCCATCATTGAGGAGATGTTGGACCTCCTGGAGCAGTCAGAGGGCAAAAT
CAACTATTGCCACAAACACGGATCTCATCACCATCACCATCACTAAGCTTGGCTGTTTTGGC
GGATGAGAGAAGATTTTCAAGCTGATACAGATTAAATCAGAACGCAGAAGCGGTCTGATAA
AACAGAATTTGCCTGGCGGCAGTAGCGCGGTGGTCCCACCTGACCCCATGCCGAACCTCAGA
AGTGAAACGCCGTAGCGCCGATGGTAGTGTGGGGTCTCCCCATGCGAGAGTAGGGAACTGC
CAGGCATCAAATAA

>pRV5_D117+GGG_forward read, MWG (top strand)

GCCGAGGAAGGCGGTGGTAGTGACGGGATTTGGCCCTTTTGGGGAACACACCGTGAACGCC
AGTTGGATTGCAGTTCAGGAGCTAGAAAAGCTAGGCCTTGGCGACAGCGTGGACCTGCATG
TGTACGAGATTCCGTTGAGTACCAAACAGTCCAGAGACTCATCCCCGCCCTGTGGGAGAAG
CACAGTCCACAGCTGGTGGTGCATGTGGGGGTGTCAGGCATGGCGACCACAGTCACACTGG
AGAAATGTGGACACAACAAGGGCTACAAGGGGCTGGACAACCTGCCGCTTTTGGCCCGGCTC
CCAGTGCTGCGTGGAGGACGGGCCTGAAAGCATTGACGGCGGCGGATCCATCATCGACATG
GATGCTGTGTGCAAGCGAGTCACCACGTTGGGCCTGGATGTGTGCGGTGACCATCTCGCAGGA
TGCCGGCAGATATCTCTGCGACTTTACCTACTACACCTCTTTGTACCAGAGTCACGGTCTGATC
AGCCTTCGTCCACGTGCCCCCACTGGGGAAGCCGTACAACGCGGACCAGCTGGGCAGGGCA
CTGAGAGCCATCATTGAGGAGATGTTGGACCTCCTGGAGCAGTC

>pRV5_I120mG_forward read, MWG (top strand)

CAGCCGAGGAAGGCGGTGGTAGTGACGGGATTTGGCCCTTTTGGGGAACACACCGTGAACG
CCAGTTGGATTGCAGTTCAGGAGCTAGAAAAGCTAGGCCTTGGCGACAGCGTGGACCTGCA
TGTGTACGAGATTCCGTTGAGTACCAAACAGTCCAGAGACTCATCCCCGCCCTGTGGGAGA
AGCACAGTCCACAGCTGGTGGTGCATGTGGGGGTGTCAGGCATGGCGACCACAGTCACACT
GGAGAAATGTGGACACAACAAGGGCTACAAGGGGCTGGACAACCTGCCGCTTTTGGCCCGGC
TCCCAGTGCTGCGTGGAGGACGGGCCTGAAAGCATTGACTCGATCGGCGACATGGATGCTGT
GTGCAAGCGAGTCACCACGTTGGGCCTGGATGTGTGCGGTGACCATCTCGCAGGATGCCGGCA
GATATCTCTGCGACTTTACCTACTACACCTCTTTGTACCAGAGTCACGGTCGATCAGCCTTCG
TCCACGTGCCCCCACTGGGGAAGCCGTACAACGCGGACCAGCTGGGCAGGGCACTGAGAGC
CATCATTGAGGAGATGTTGGACCTCCTGGAGCAGTCAGAGGGGCAAAATCAACTATTGCCAC
AAACACGGATCTCATCACCATCACCATCACTAAGCTTGGCTGTTTTGGCGGATGAGAGAAGA
TTTTCAGCCTGATACAGATTAAATCAGAACGCAGAAGCGGTCTGATAAAACAGAATTTGCCT
GGCGGCAGTAGCGCGGTGGTCCCACCTGACCCCATGCCGA

>pRV5_Q180mE.N177mY_forward read, MWG (top strand)

TTAACCATGGAGCAGCCGAGGAAGGCGGTGGTAGTGACGGGATTTGGCCCTTTTGGGGAAC
ACACCGTGAACGCCAGTTGGATTGCAGTTCAGGAGCTAGAAAAGCTAGGCCTTGGCGACAG
CGTGGACCTGCATGTGTACGAGATTCCGGTTGAGTACCAAACAGTCCAGAGACTCATCCCCG
CCCTGTGGGAGAAGCACAGTCCACAGCTGGTGGTGCATGTGGGGGTGTCAGGCATGGCGAC
CACAGTCACACTGGAGAAAATGTGGACACAACAAGGGCTACAAGGGGCTGGACAACTGCCGCT
TTTTGCCCCGGCTCCCAGTGCTGCGTGGAGGACGGGCCTGAAAGCATTGACTCCATCATCGA
CATGGATGCTGTGTGCAAGCGAGTCACCACGTTGGGCCTGGATGTGTCCGTGACCATCTCGC
AGGATGCCGGCAGATATCTCTGCGACTTTACCTACTACACCTCTTTGTACCAGAGTCACGGT
CGATCAGCCTTCGTCCACGTGCCCCCACTGGGGAAGCCGTAACGCGGACGAGCTCGGCAG
GGCACTGAGAGCCATCATTGAGGAGATGTTGGACCTCCTGGAGCAGTCAGAGGGCAAAATC
AACTATTGCCACAAACACGGATCTCATCACCATCACCATCACTAAGCTTGGCTGTTTTGGCG
GATGAGAGAAGATTTTCAGCCTGATACAGATTAAATCAGAACGCAGAAGCGGTCTGATAAA
ACAGAATTTGCCTGGCGGCAGTAGCGCGGTGGTCCACCTGACCCCATGCCGAACTCAGAA
GTGAAACGCCGTAGCGCCGATGGTAGTGTGGGGTCTCCCCATGCGAGAGTAGGGAACTGCC
AGGCATCAAA

>pRV5_Q180mE_forward read, MWG (top strand)

TTAACCATGGAGCAGCCGAGGAAGGCGGTGGTAGTGACGGGATTTGGCCCTTTTGGGGAAC
ACACCGTGAACGCCAGTTGGATTGCAGTTCAGGAGCTAGAAAAGCTAGGCCTTGGCGACAG
CGTGGACCTGCATGTGTACGAGATTCCGGTTGAGTACCAAACAGTCCAGAGACTCATCCCCG
CCCTGTGGGAGAAGCACAGTCCACAGCTGGTGGTGCATGTGGGGGTGTCAGGCATGGCGAC
CACAGTCACACTGGAGAAAATGTGGACACAACAAGGGCTACAAGGGGCTGGACAACTGCCGCT
TTTTGCCCCGGCTCCCAGTGCTGCGTGGAGGACGGGCCTGAAAGCATTGACTCCATCATCGA
CATGGATGCTGTGTGCAAGCGAGTCACCACGTTGGGCCTGGATGTGTCCGTGACCATCTCGC
AGGATGCCGGCAGATATCTCTGCGACTTTACCTACTACACCTCTTTGTACCAGAGTCACGGT
CGATCAGCCTTCGTCCACGTGCCCCCACTGGGGAAGCCGTACAACGCGGACGAGCTCGGCA
GGCACTGAGAGCCATCATTGAGGAGATGT

>pRV5_Q180mE_reverse read, MWG (top strand)

GTGCATGTGGGGGTGTCAGGCATGGCGACCACAGTCACACTGGAGAAAATGTGGACACAACA
AGGGCTACAAGGGGCTGGACAACTGCCGCTTTTGGCCCCGGCTCCCAGTGCTGCGTGGAGGAC
GGGCCTGAAAGCATTGACTCCATCATCGACATGGATGCTGTGTGCAAGCGAGTCACCACGTT
GGGCCTGGATGTGTCCGTGACCATCTCGCAGGATGCCGGCAGATATCTCTGCGACTTTACCT
ACTACACCTCTTTGTACCAGAGTCACGGTGCATCAGCCTTCGTCCACGTGCCCCCACTGGGG
AAGCCGTACAACGCGGACGAGCTCGGCAGGGCACTGAGAGCCATCATTGAGGAGATGTTGG
ACCTCCTGGAGCAGTCAGAGGGCAAAATCACTATTGCCACAAACACGGATCTCATCACCAT
CACCATCACTAAGCTTGGCTGTTTGGCG

>pRV5_Δ3'_forward read, MWG (top strand)

TAACCATGGAGCAGCCGAGGAAGGCGGTGGTAGTGACGGGATTTGGCCCTTTTGGGGAACA
CACCGTGAACGCCAGTTGGATTGCAGTTCAGGAGCTAGAAAAGCTAGGCCTTGGCGACAGC
GTGGACCTGCATGTGTACGAGATTCCGGTTGAGTACCAAACAGTCCAGAGACTCATCCCCG
CCTGTGGGAGAAGCACAGTCCACAGCTGGTGGTGCATGTGGGGGTGTCAGGCATGGCGACC
ACAGTCACACTGGAGAAAATGTGGACACAACAAGGGCTACAAGGGGCTGGACAACTGCCGCT
TTTTGCCCCGGCTCCCAGTGCTGCGTGGAGGACGGGCCTGAAAGCATTGACTCCATCATCGAC
ATGGATGCTGTGTGCAAGCGAGTCACCACGTTGGGCCTGGATGTGTCCGTGACCATCTCGCA
GGATGCCGGCAGATATCTCTGCGACTTTACCTACTACACCTCTTTGTACCAGAGTCACGGTC
GATCAGCCTTCGTCCACGTGC

>pRV5_Δ3'_reverse read, MWG (top strand)

GACCTGCATGTGTACGAGATTCCGGTTGAGTACCAAACAGTCCAGAGACTCATCCCCGCCCT
GTGGGAGAAGCACAGTCCACAGCTGGTGGTGCATGTGGGGGTGTCAGGCATGGCGACCACA
GTCACACTGGAGAAAATGTGGACACAACAAGGGCTACAAGGGGCTGGACAACTGCCGCTTTT
GCCCCGGCTCCCAGTGCTGCGTGGAGGACGGGCCTGAAAGCATTGACTCCATCATCGACATG
GATGCTGTGTGCAAGCGAGTCACCACGTTGGGCCTGGATGTGTCCGTGACCATCTCGCAGGA
TGCCGGCAGATATCTCTGCGACTTTACCTACTACACCTCTTTGTACCAGAGTCACGGTCGATC
AGCCTTCGTCCACGTGCCCCCACTGGGGAAGCCGTACAACGCGGACGAGCTGGGCAGGGCA
CTGAGAGCCATCATTGAGGAGATGTTGGACCTCCTGGAGCAGTCAGAGGGATCTCATCACC
TCACCATCACTAAGCTTGGCTGTTTGGCG

>pRV5_Δ3'Δ6H_forward read, MWG (top strand)
TAAGAGGAGAATTAACCATGGAGCAGCCGAGGAAGGCGGTGGTAGTGACGGGATTTGGCCC
TTTTGGGGAACACACCGTGAACGCCAGTTGGATTGCAGTTCAGGAGCTAGAAAAGCTAGGC
CTTGCGGACAGCGTGGACCTGCATGTGTACGAGATTCCGGTTGAGTACCAAACAGTCCAGAG
ACTCATCCCCGCCCTGTGGGAGAAGCACAGTCCACAGCTGGTGGTGCATGTGGGGGTGTCAG
GCATGGCGACCACAGTCACACTGGAGAAATGTGGACACAACAAGGGCTACAAGGGGCTGGA
CAACTGCCGCTTTTGCCCCGGCTCCCAGTGCTGCGTGGAGGACGGGCCTGAAAGCATTGACT
CCATCATCGACATGGATGCTGTGTGCAAGCGAGTCACCACGTTGGGCCTGGATGTGTCGGTG
ACCATCTCGCAGGATGCCGGCAGATATCTCTGCGACTTTACCTACTACACCTCTTTGTACCAG
AGTCACGGTCGATCAGCCTTCGTCCACGTGCCCCACTGGGGAAGCCGTACAACGCGGACCA
GCTGGGCAGGGCACTGAGAGCCATCATTGAGGAGATGTTGGACCTCCTGGAGCAGTCAGAG
TAAGGATCCGTCGACCTGCAGCCAAGCTTGGCTGTTTTGGCGGATGAGAGAAGATTTTCAGC
CTGATACAGATTAAATCAGAACGCAGAAAGCGGTCTGATAAAACAGAATTTGCCTGGCGGCA
GTAGCGCGGTGGTCCCACCTGACCCCATGCCGAAGTCAGAAAGTAAAACGCCGTAGCGCCGA
TGGTAGTGTGGGG

>pRV5_Δ3'Δ6H_reverse read, MWG (top strand)
ATCGGAAGCTGTGGTATGGCTGTGCAGGTCGTAAATCACTGCATAATTCGTGTCGCTCAAGG
CGCACTCCCGTTCTGGATAATGTTTTTTCGCGCCGACATCATAACGGTTCTGGCAAATATTCTG
AAATGAGCTGTTGACAATTAATCATCGGCTCGTATAATGTGTGGAATTGTGAGCGGATAACA
ATTTACACAGGAAACAGAATTCATTAAAGAGGAGAAATTAACCATGGAGCAGCCGAGGAA
GGCGGTGGTAGTGACGGGATTTGGCCCTTTTGGGGAACACACCGTGAACGCCAGTTGGATTG
CAGTTCAGGAGCTAGAAAAGCTAGGCCTTGGCGACAGCGTGGACCTGCATGTGTACGAGAT
TCCGGTTGAGTACCAAACAGTCCAGAGACTCATCCCCGCCCTGTGGGAGAAGCACAGTCCAC
AGCTGGTGGTGCATGTGGGGGTGTCAGGCATGGCGACCACAGTCACACTGGAGAAATGTGG
ACACAACAAGGGCTACAAGGGGCTGGACAACCTGCCGCTTTTGCCCCGGCTCCCAGTGCTGCG
TGGAGGACGGGCCTGAAAGCATTGACTCCATCATCGACATGGATGCTGTGTGCAAGCGAGT
CACCACGTTGGGCCTGGATGTGTCGGTGACCATCTCGCAGGATGCCGGCAGATATCTCTGCG
ACTTTACCTACTACACCTCTTTGTACCAGAGTCACGGTCGATCAGCCTTCGTCCACGTGCCCC
CACTGGGGAAGCCGTACAACGCGGACCAGCTGGGCAGGGCACTGAGAGCCATCATTGAGGA
GATGTTGGACCTCCTGGAGCAGTCAGAGTAAGGATCCGTCGACCTGCAGCCAAGCTTGGCTG
TTTGCCG

Appendix C
Activity & Kinetics

Quantification of PAP1 activity

One unit of activity is defined as the amount of enzyme, which releases 1 nanomole of AMC per minute at 37°C (unit = nmoles.min⁻¹)

Derivation:

AMC released (X) is defined as fluorescent intensity (Fi) divided by the slope of the appropriate AMC standard curve (Appendix A).

$$\begin{aligned}\frac{F_i}{m} &= X \mu\text{M} \\ &= \mu\text{moles.L}^{-1}\end{aligned}$$

The reaction volume is 500x10⁻⁶ L and duration is 15 min

$$\therefore \text{AMC released} = \frac{X(500 \times 10^{-6})}{15} \mu\text{moles.L}^{-1}.\text{L.min}^{-1}$$

Reaction uses 100x10⁻⁶ L enzyme

$$\begin{aligned}\therefore \text{AMC released by enzyme} &= \frac{X(500 \times 10^{-6})}{15(100 \times 10^{-6})} \mu\text{moles.min}^{-1}.\text{L}^{-1} \\ &= \frac{X(500 \times 10^{-6})(1000)}{15(100 \times 10^{-6})(1000)} \text{nmoles.min}^{-1}.\text{ml}^{-1} \\ &= \frac{X}{3} \text{units.ml}^{-1} \\ &= \frac{F_i}{3m} \text{units.ml}^{-1}\end{aligned}$$

From AMC standard curve, m = 10.535

$\therefore \text{Enzyme activity} = \frac{F_i}{31.605} \text{units.ml}^{-1}$

Determination of K_m and V_{max}

The Michaelis constant (K_m) of an enzyme, for a particular substrate, is defined as the substrate concentration which gives rise to half the maximal enzyme velocity (V_{max}). This constant can be determined by measuring reaction velocity at various concentrations of substrate, giving rise to the Michaelis-Menten hyperbola curve. In the case of PAP1, activity is determined by measuring fluorescent intensity (F_i) resulting at various concentrations of the substrate pGlu-AMC, as described in Section 2.19.3. A plot of F_i versus [pGlu-AMC] yields the Michaelis-Menten curve. Once an enzyme-catalysed reaction follows normal Michaelis-Menten kinetics, data can be applied to a choice of kinetic models:

Lineweaver-Burk: Plot of $1/F_i$ versus $1/[pGlu-AMC]$. The intercept of the line on the x-axis gives $-1/K_m$ and the intercept of the line on the y-axis gives $1/V_{max}$.

Eadie-Hofstee: Plot of F_i versus $F_i/[pGlu-AMC]$. The slope is $-K_m$ and the intercept on the y-axis represents V_{max} .

Hanes-Woolf: Plot of $[pGlu-AMC]/F_i$ versus $[pGlu-AMC]$. The intercept on the x-axis gives $-K_m$, and the slope is $1/V_{max}$.

Determination of k_{cat}

The turnover constant (k_{cat}) of an enzyme can be determined if the V_{max} and total enzyme (E_t) are known:

$$k_{cat} = \frac{V_{max}}{E_t} \quad \frac{nmoles.min^{-1}.ml^{-1}}{nmoles.ml^{-1}} = min^{-1} (s^{-1})$$

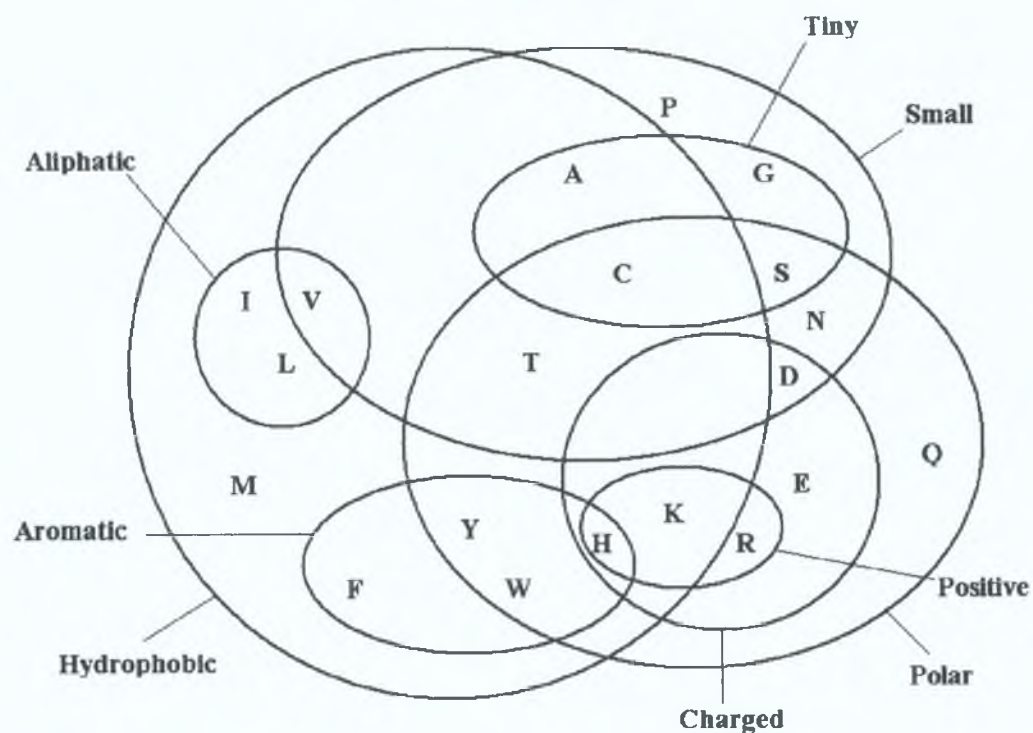
In the case of PAP1, V_{max} ($nmoles.min^{-1}.ml^{-1}$ or $units.ml^{-1}$) is determined experimentally as described above. E_t ($nmoles.ml^{-1}$) can be calculated from the molecular weight ($g.mole^{-1}$) and amount of enzyme used ($g.ml^{-1}$). The molecular weight of His-tagged recombinant human PAP1 (rHsaPAP1_{6H}) is 24,105 $g.mole^{-1}$, as deduced from the amino acid sequence.

Determination of K_i

The inhibition constant (K_i), or dissociation constant for the enzyme-inhibitor complex, can be determined by measuring the apparent K_m (K_m^{app}) observed in the presence of an inhibitor at a specific concentration ($[I]$), as described in Section 2.19.4 and applied to the following equation:

$$K_i = \frac{K_m [I]}{K_m^{app} - K_m}$$

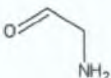
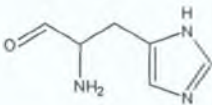
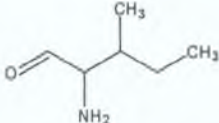
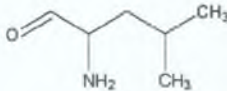
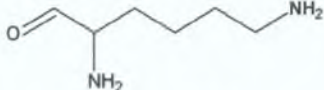
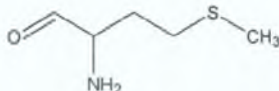
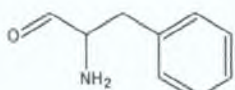
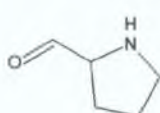
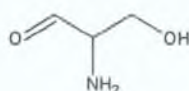
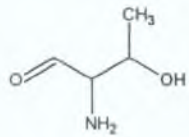
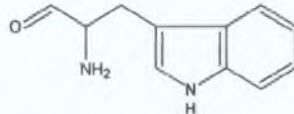
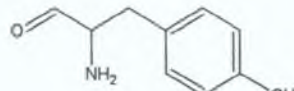
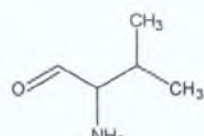
Appendix D
Amino Acids



Amino acid properties

Venn diagram presenting properties of the 20 amino acids

Alanine	Ala	A	<chem>CC(N)C(=O)O</chem>
Arginine	Arg	R	<chem>CCCNC(=[NH2+])N</chem>
Asparagine	Asn	N	<chem>NC(=O)CC(N)C(=O)O</chem>
Aspartic acid	Asp	D	<chem>OC(=O)CC(N)C(=O)O</chem>
Cysteine	Cys	C	<chem>SCCC(N)C(=O)O</chem>
Glutamic acid	Glu	E	<chem>OC(=O)CCC(N)C(=O)O</chem>
Glutamine	Gln	Q	<chem>NC(=O)CCC(N)C(=O)O</chem>

Glycine	Gly	G	
Histidine	His	H	
Isoleucine	Ile	I	
Leucine	Leu	L	
Lysine	Lys	K	
Methionine	Met	M	
Phenylalanine	Phe	F	
Proline	Pro	P	
Serine	Ser	S	
Threonine	Thr	T	
Tryptophan	Trp	W	
Tyrosine	Tyr	Y	
Valine	Val	V	

		Second Position				
	U	C	A	G		
First Position	U	UUU Phe	UCU Ser	UAU Tyr	UGU Cys	U
		UUC Phe	UCC Ser	UAC Tyr	UGC Cys	
		UUA Leu	UCA Ser	UAA Stop	UGA Stop	
		UUG Leu	UCG Ser	UAG Stop	UGG Trp	
	C	CUU Leu	CCU Pro	CAU His	CGU Arg	C
		CUC Leu	CCC Pro	CAC His	CGC Arg	
		CUA Leu	CCA Pro	CAA Gln	CGA Arg	
		CUG Leu	CCG Pro	CAG Gln	CGG Arg	
	A	AUU Ile	ACU Thr	AAU Asn	AGU Ser	A
		AUC Ile	ACC Thr	AAC Asn	AGC Ser	
		AUA Ile	ACA Thr	AAA Lys	AGA Arg	
		AUG Met	ACG Thr	AAG Lys	AGG Arg	
	G	GUU Val	GCU Ala	GAU Asp	GGU Gly	G
		GUC Val	GCC Ala	GAC Asp	GGC Gly	
		GUA Val	GCA Ala	GAA Glu	GGA Gly	
		GUG Val	GCG Ala	GAG Glu	GGG Gly	
		Third Position				

Genetic code

Amino acid codons applicable to both *H. sapiens* and *E. coli*.