

https://theses.gla.ac.uk/

Theses Digitisation:

https://www.gla.ac.uk/myglasgow/research/enlighten/theses/digitisation/

This is a digitised version of the original print thesis.

Copyright and moral rights for this work are retained by the author

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge

This work cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given

Enlighten: Theses <u>https://theses.gla.ac.uk/</u> research-enlighten@glasgow.ac.uk

Characterisation of a plastid DNA replicase in *Plasmodium falciparum*

Sze Inn Fiona Seow

Division of Infection and Immunity Institute of Biomedical and Life Sciences

This thesis is submitted for the degree of Doctor of Philosophy Faculty of Biomedical and Life Sciences University of Glasgow

June 2004

ProQuest Number: 10390651

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



ProQuest 10390651

Published by ProQuest LLC (2017). Copyright of the Dissertation is held by the Author.

All rights reserved. This work is protected against unauthorized copying under Title 17, United States Code Microform Edition © ProQuest LLC.

> ProQuest LLC. 789 East Eisenhower Parkway P.O. Box 1346 Ann Arbor, MI 48106 – 1346



1

. . .

<u>Abstract</u>

Human malaria is caused by the parasite genus, *Plasmodium*, with *P. falciparum* being the most pathogenic species. The disease affects between 300 to 500 million worldwide. Currently, efforts to contain the disease are curbed by the limited range of antimalarials and the rise of resistance to these treatments. The race is now on to gain a thorough understanding of the biology of the parasite that can be use for the development of strategies to combat the wrath of this disease. Since the 22.8 Mb genome sequence of *P. falciparum* has been fully-sequenced, the data has provided new candidate genes that may be valid drug targets. One area that is of great interest is that of the apicoplast of *Plasmodium*. The evolutionary origin of the apicoplast is generally believed to reflect a secondary endosymbiotic event, in which a proto-eukaryote harbouring a photosynthetic cyanobacterium itself became an endosymbiont of a heterotrophic eukaryote. In *Plasmodium*, this organelle appears to be essential in these parasites and enzyme systems specific to apicoplast have been identified as targets for anti-protozoal chemotherapy.

Segments of a gene (*PfPREX*; formerly referred to as *PfPOM1*), encoding a large protein with a typical N-terminal bipartite plastid targeting sequence have been cloned from *Plasmodium falciparum* and expressed in bacteria. Domains with homology to prokaryotic DNA polymerase, DNA primase and DNA helicase, all were found specified in an opening reading frame (ORF). The putative polymerase domain possesses both DNA polymerase activity and 3'-5' exonuclease activity. The putative helicase-primase domain has also been expressed and both DNA primase and helicase activities were detected. RT-PCR indicates that the open reading frame is expressed as a single transcript whilst Western blot analysis

has shown that the protein is cleaved post-translationally. Localisation studies carried out by collaborators (Sato S, NIMR, Mill Hill) with a GFP-reporter revealed that the protein is targeted exclusively to the plastid.

In addition, the evolutionary history of *PfPREX* appears to be more complex and cannot be explained fully using the secondary endosymbiotic theory. The DNA polymerase domain has great homology to prokaryotic DNA polymerase I. The DNA primase and DNA helicase domain, on the other hand, shows striking similarity to proteins with these functions encoded in T-odd bacteriophages. T-odd bacteriophage related RNA polymerase proteins have previously been identified as playing roles in transcription of plastid genes. Now it seems that additional phage-like proteins also play a key role in the replication of the *Plasmodium* plastid genome. The presence of these T-odd genes in *Plasmodium* may be explained with different theories, mainly the non-orthologous gene displacement with the T-odd bacteriophage genes.

Nonetheless, the discovery of *PfPREX* has opened insights to many exciting possibilities and insight into *Plasmodium falciparum*. This gene can provide not only a valid drug target, due to its prokaryotic origin, but also an insight into the complex evolutionary history of *Plasmodium*.

......

Acknowledgements

Firstly, I would like to thank my supervisor, Dr Michael Barrett, for all his support and guidance and most important his confidence in me throughout the project.

I would also like to thank everyone in the North Labs, their company throughout my stay in the lab had been great. The lunches we took together, the chats we had, I am going to miss you all. I would like to especially thank Dorothy who had been tremendously help in the lab. All my good friends that I had made in the lab, thank you for your support.

Finally, I would like to thank my family, my parents who supported my decision to come to UK to do my postgraduate training. My husband, Frederic Waldteufel, who has been always by my side, for his support and most important, his love, which is the reason that 1 could have achieved so much.

Declaration

This thesis and the results presented in it are entirely my own work, except where indicated.

Sze Inn Fiona Seow

TABLE OF CONTENTS

Abstract	ii
Acknowledgements	iv
Declaration	v
Table of Contents	vi
List of Figures	xii
List of Tables	xvii
List of Abbreviations	xix

Chapter I -	- Introduction	1
1.1.	Malaria - an introduction	2
1.2.	Plasmodium - the parasite	3
1.3.	Clinical features of malaria	6
1.3.1.	Falciparum malaria	б
1.3.2.	Treatment of malaria	6
1.4.	Molecular biology of <i>Plasmodium</i>	
		9
1.4.1.	The genome of <i>Plusmodium</i>	9
1.4.2.	Extrachromosomal DNAs	11
1.4.2.1.	Plastid DNA	12
1.4.3.	DNA replication	14
1.4.3.1.	Enzymes involved in DNA replication	19
1.4.3.1.	1. DNA helicases	19
1.4.3.	1.1.1.Molecular biology of DNA helicases	21
1.4.3.	1.1.2. Molecular biology of DnaB-like helicases	24
1.4.3.1.	2. DNA primases	24
1.4.3.	1.2.1.Prokaryotic DNA primases	25
1.4.3.1.	3. DNA polymerases	28
1.4.3.	1.3.1.E. coli DNA polymerase I	31
1.4.3.2.	DNA replication in <i>Plasmodium</i>	34
1.4.4.	Plastid DNA replication	36

.

j

1.5.	Apicoplast	37
1.5.1.	Introduction	37
1.5.2.	Origin and evolution of the apicoplast	38
1.5.3.	Protein targeting to the apicoplast	40
1.5.4.	Functions of the apicoplast	43
1.5.4.1.	Fatty acid synthesis	43
1.5.4.2.	Isoprenoid synthesis	44
1.5.4.3.	Haem synthesis	44
1.5.5.	The apicoplast as a potential drug target	45
1.5.5.1.	Fatty acid synthesis	45
1.5.5.2.	Isoprenoid synthesis	46
1.5.5.3.	DNA replication and transcription	46

Aims of the project

48

1.000

Chapter II	- Materials and Methods	49
2.1.	Materials	50
2.2.	Parasite lines/clones and culture	50
2.3.	Bacterial strains and cultures	51
2.4.	Basic molecular biology techniques	
		51
2.4.1.	Preparation of parasite extracts	51
2.4.2.	Preparation of genomic DNA from P. falciparum	52
2.4.3.	Preparation of total RNA from <i>P. falciparum</i>	52
2.4.4.	Preparation of plasmid DNA from <i>E. coli</i>	53
2.4.5.	Restriction digestion of plasmid DNA	53
2.4.6.	Agarose gel electrophoresis of DNA	53
2.4.7.	Polymerase chain reaction (PCR)	54
2.4.8.	DNA extraction from low-melting point agarose gels	55
2.4.9.	Reverse-transcriptase PCR amplification (RT-PCR)	55
2.4.9.1	First strand cDNA synthesis	55
2.4.9.2.	Reverse transcriptase-PCR	56
2.4.10.	Formaldehyde gel electrophoresis of RNA	56
2.4.11.	Northern hybridisation	57
2.4.11.1.	Labelling of DNA probes	57
2.4.12.	Chemical transformation of E. coli	58

2.5.	Molecular cloning techniques for PfPREX	58
2.5.1.	Gateway cloning technology	58
2.5.2.	Generation of expression constructs	59
2.5.2.1.	Generation of entry clones	59
2.5.2.1.1	. Preparation of att-PCR products	60
2.5.2.1.2	2. 'BP' reaction for the generation of the entry clones	60
2.5.2.2.	Generation of destination clones	61
2.5.2.2.1	.'LR' reaction for the generation of destination clones	61
2.6.	Protein analysis	61
2.6.1.	Sodium dodecyl sulphate polyacrylamide gel electrophoresis	
	(SDS-PAGE)	61
2.6.2.	Western blotting	62
2,6.3.	Protein quantification	63
2.7.	Expression and purification of recombinant proteins	63
2.7.1.	Expression and purification of His-tagged recombinant proteins	63
2.7.2.	Expression and purification of GST-tagged recombinant proteins	64
2.7.3.	Production of antibodies	65
2.0	Time At	
2.8.	Functional analysis of the recombinant proteins	00
2.8.1.	DNA primase activity assays	66
2.8.2.	DNA neficase activity assays	0/
2.8.3.	DNA polymerase activity assays	08
2.8.4.	3'-5' exonuclease activity assays	09
2.8.3.	Nucleotide hydrolysis assays	70
2.9.	Subcellular localisation of <i>PfPREX</i> gene products	71
2.9.1.	Indirect immunofluorescence	71
2.9.2.	<i>In situ</i> hybridisation	72
2.9.2.1.	DNA probe preparation	72
2.9.2.2.	In situ hybridisation	72
2.9.2.3.	Fluorescent detection of the <i>in situ</i> hybridisation DNA probe	72
2.9.3.	Fluorescent microscopy	73
2.10.	DNA sequencing	73
2.11.	Software	73

Chapter III -	Identification of a novel gene, PfPREX from
Plasmodium j	alciparum

<u>76</u>

3.1.	Introduction - background to the discovery of the gene, <i>PfPREX</i>	77
3.2. 3.2.1.	Identification of the gene and sequence analysis Identification of a DNA polymerase-like domain in the PfPREX	78
	sequence	78
3.2.2.	The family A group of DNA polymerases	81
3.2.3	3'-5' exonuclease domain	81
3.2.4	The DNA polymerase domain	85
3.3.	Identification of DNA helicase and DNA primase domains	85
3.3.1.	The DNA helicase domain of PfPREX	86
3.3.2.	The DNA primase domain of PfPREX	88
3.3.2.1.	Conserved features of DNA primases	88
3.4.	The hinge domain of PfPREX	92
3.5.	Identification of a transit peptide at the extreme N-terminal domain	92
3.6.	PREX homologues in other Apicomplexans	95
3.6.1.	Homologues in <i>Plasmodium</i>	95
3.6.2.	PREX homologues in other Apicomplexans	97
3.7.	Determination of the arrangement of the different domains withi	n
	PfPREX	99
3.8.	Phylogenetic relationship between PREX homologues in	100
	арконфиланs	108
3.9.	Discussion	111

Chapter IV - Functional analysis and subcellular localisation of the DNA polymerase (*PfPREXpol*) domain encoded by the *PfPREX* gene113

4.1.	Introduction	114
4.2.	Results	117
4.2.1.	Cloning of the PfPREXpol domain of the PfPREX gene	117
4.2.2	Expression and purification of PfPREXpol	121
4.2.3.	Functional characterisation of PfPREXpol	122
4.2.3.1.	DNA polymerase activity	122

.3.	Discussion	145
4.2.6.	Subcellular localisation of the PfPREXpol gene product	142
4.2.5.2.	Comparison of PfPREXpol and E. coli Klenow fragment 3D	structures134
4.2.5.1	A modelled 3D structure of PfPREXpol protein	131
4.2.5.	Homology modelling of PfPREXpol	131
4.2.4.	Expression of PfPREXpol in P. falciparum	128
4.2.3.2.	3'-5' exonuclease activity	127

Discussion 4.3.

Chapter V - Functional analysis and subcellular localisation of the DNA primase/helicase (PfPREXpriheli) domain of the PfPREX gene 148

5.1.	Introduction	149
5.2.	Results	151
5.2.1.	Cloning of the PfPREX priheli domains of the PfPREX gene	151
5.2.2.	Expression and purification of DNA primase/helicase and hinge	
	recombinant proteins	155
5.2.3.	Functional characterisation of PfPREXpriheli	156
5.2.3.1.	DNA primase activity	156
5.2.3.2.	DNA helicase activity	158
5.2.3.2	2.1. Nucleotide hydrolysis Assay	161
5.2.4.	Expression of PfPREX priheli in P. falciparum	162
5.2.5.	Subcellular localisation of PfPREX priheli gene product	167
5.3.	Discussion	169

Chapter VI - The Evolution of *PfPREX*

6.1.	Introduction	173
6.2.	Phylogenetic analysis of DNA polymerases	174
6.2.1.	DNA polymerase families	174
6.2.2.	Alignment of DNA polymerases	175
6.2.3.	Generation of phylogenetic trees for the DNA polymerase domains	180
6.3.	Phylogenetic analysis of the DNA helicase and DNA primase	
	domains of PIPREX	187
6.3.1.	DNA helicases and DNA primases	187
6.3.1.1.	DNA primases	187
6.3.1.2.	DNA helicases	187
6.3.1.3.	DNA Helicase/primase Organisation	188
6.3.2.	Alignment of DNA primases	189
6.3.3.	Alignment of DNA helicases	190
6.3.4.	Generation of phylogenetic trees for the DNA primases	192

6.3.5.	Generation of phylogenetic trees for the DNA helicases	193
6.4.	Discussion	201
<u>Chapter</u>	VII – General Discussion	205
Bibliogr	aphy	216
Appendie	CS	241
Appendix	Α	242
Appendix	B	250
Appendix	C	255
Appendix	D	265

- 14 - 14 - 14

•.

.

. .

1.2.4. ...

2

.

List of Figures

Chapter I – Introduction

.1.	Geographical distribution of the incidence of malaria.	2
.2.	(A) Life cycle of <i>Plasmodium</i> . (B) Photograph of a female anopheline mosquito.	4
.3.	Gene map of the plastid DNA of P. falciparum.	13
.4.	A DNA replication fork	16
.5.	Replication fork showing the leading strand and the lagging strand	18
.6.	The main groups of DNA helicases.	23
.7.	A schematic diagram representing the motifs of DnaB like helicases.	23
.8.	Schematic diagram of the domain structure and the conserved residues in DnaG-like primase.	n 26
.9.	3D structure of <i>E. coli</i> Klenow fragment and <i>Thermus aquaticus Taq</i> polymerase.	33
.10.	Ultrastructure of a ring stage <i>Plasmodium falciparum</i> showing the varior organelles like the plastid and mitochondrion.	us 37
.11.	Secondary endosymbiotic event where the apicomplexan ancestor engulfed a eukaryote that already harboured a plastid	39
.12.	Diagrammatic representation of the chloroplast protein transport system (Tic-Toc)	42
	.1. .2. .3. .4. .5. .6. .7. .8. .9. .10. .11. .12.	 Geographical distribution of the incidence of malaria. (A) Life cycle of <i>Plasmodium</i>. (B) Photograph of a female anopheline mosquito. Gene map of the plastid DNA of <i>P. falciparum</i>. A DNA replication fork Replication fork showing the leading strand and the lagging strand The main groups of DNA helicases. A schematic diagram representing the motifs of DnaB like helicases. Schematic diagram of the domain structure and the conserved residues in DnaG-like primase. 3D structure of <i>E. coli</i> K lenow fragment and <i>Thermus aquaticus Taq</i> polymerase. Ultrastructure of a ring stage <i>Plasmodium falciparum</i> showing the vario organelles like the plastid and mitochondrion. Secondary endosymbiotic event where the apicomplexan ancestor engulfed a eukaryote that already harboured a plastid Diagrammatic representation of the chloroplast protein transport system (Tic-Toc)

Chapter II - Materials and Methods

Figure 2.1.	Schematic representation of the Gateway cloning technology.	59
Figure 2.2.	Helicase [3H] scintillation proximity enzyme (SPA) assay (Amersham).	68

See.

Chapter III - Identification of a novel gene, PfPREX from Plasmodium falciparum

Figure 3.1.	BLAST results with the description consisting of four columns	79
Figure 3.2.	Diagrammatic representation of PIPREX protein with the different functional domains it encodes for	80
Figure 3.3.	Diagrammatic representations of the roles of different members of the fa A group of DNA polymerases	umily 82
Figure 3.4.	Alignments of deduced amino acid sequence of the polA domain of DN. polymerases from family A.	A 83
Figure 3.5.	Pairwise alignment of the amino acid sequences of the Klenow fragment of E coli and PfPREXpol using ClustalX 1.81 with default settings.	t 84
Figure 3.6.	Alignments of deduced amino acid sequences of DNA helicases.	87
Figure 3.7.	Alignments of deduced amino acid sequences of DNA primases.	89
Figure 3.8.	Diagrammatic representation of the 3 key domains in prokaryotic DNA primase	9 1
Figure 3.9.	Result of the PlasmoAP apicoplast targeting peptide prediction	94
Figure 3.10.	Multiple alignment of the PREX homologues using ClustalX 1.81 with default settings.	100
Figure 3.11.	Putative amino acid of PfPREX which is colour coded for the different functional domains.	107
Figure 3.12.	Phylogenetic trees for PREX homologues.	110
<u>Chapter IV</u> (PfPREXpo	- Functional analysis and subcellular localisation of the DNA polym I) domain encoded by the <i>PfPREX</i> gene	erase
Figure 4.1.	Diagrammatic representation of the <i>PfPREX</i> showing the gene specific primers GWpolF1 and GWpolR1	118
Figure 4.2,	PCR amplification products of the polymerase domain of PIPREX	119
Figure 4.3.	The pDEST17GWpol expression construct.	120
Figure 4.4.	Coomassie blue-stained SDS-polyacrylamide gel of expressed and purified PfPREXpol.	121

ļ

•

Figure 4.5.	Characterisation of DNA polymerase activity of PfPREXpol and the eff of different conditions on its activity.	fects 125
Figure 4.6	Time course experiment of 3'-5' exonuclease domain of PfPREXpol.	128
Figure 4.7.	RT-PCR amplification of the PfPREX transcript from <i>P. falciparum</i> 3D7 total RNA.	129
Figure 4.8	Western Blot of parasitised rbc (3D7) and non-parasitised rbc lysates	131
Figure 4.9.	3D structure of Klenow fragment from <i>E. coli</i> and PfPREXpol which w modelled on the Klenow fragment	vas 133
Figure 4.10.	Comparison of the structures of Klenow and PfPREXpol.	135
Figure 4.11.	Superimposed stereodiagrams of Klenow and PfPREXpol	139
Figure 4.13.	Subcellular localisation of PfPREXpol using immunofluorescence technique	144
	-	
Chapter V primase/hel	Functional analysis and subcellular localisation of the DNA icase (PfPREXpriheli) domain of the PfPREX gene	<u>148</u>
Chapter V · primase/hel Figure 5.1.	Functional analysis and subcellular localisation of the DNA icase (PfPREXpriheli) domain of the PfPREX gene Schematic diagram showing the location of the primers used on the gen PfPREX.	<u>148</u> e 152
Chapter V primase/hel Figure 5.1. Figure 5.2.	Functional analysis and subcellular localisation of the DNA icase (PfPREXpriheli) domain of the PfPREX gene Schematic diagram showing the location of the primers used on the gen PfPREX. Cloning of recombinant His6-tagged PfPREXpriheli DNA helicase/primase.	<u>148</u> e 152 153
Chapter V primase/hel Figure 5.1. Figure 5.2. Fi*p+1Xgur	Functional analysis and subcellular localisation of the DNA icase (PfPREXpriheli) domain of the PfPREX genc Schematic diagram showing the location of the primers used on the gen <i>PfPREX</i> . Cloning of recombinant His6-tagged PfPREXpriheli DNA helicase/primase. e 5.3. PCR amplification products of the fragments from the DNA prim DNA helicase and hinge domains of PfPREX.	<u>148</u> e 152 153 nase, 154
Chapter V - primase/hel Figure 5.1. Figure 5.2. Fi*p+1Xgur Figure 5.4.	 Functional analysis and subcellular localisation of the DNA icase (PfPREXpriheli) domain of the PfPREX gene Schematic diagram showing the location of the primers used on the gen PfPREX. Cloning of recombinant His6-tagged PfPREXpriheli DNA helicase/primase. e 5.3. PCR amplification products of the fragments from the DNA prin DNA helicase and hinge domains of PfPREX. Coomassie blue-stained SDS-polyacrylamide gel of nickel affinity purit proteins. 	<u>148</u> e 152 153 nase, 154 fied 155

Figure 5.7. Characterisation of DNA helicase activity of PfPREXpriheli and the effects of enzyme quantities on its activity and its time course 159

Characterisation of the DNA primase activity of PfPREXpriheli and the effects of different conditions on its activity.

158

-

Figure 5.6

Figure 5.8.	Characterisation of DNA helicase activity of PfPREXpriheli and the effects of different conditions on its activity.	
Figure 5.9.	Nucleotide hydrolysis assay of PfPREXpriheli.	
Figure 5.10.	. RT-PCR amplification of the PfPREX transcript from <i>P. falciparum</i> 3D' total RNA.	
Figure 5.11.	Western Blot of parasitised rbc (3D7) and non-parasitised rbc lysates probed with anti-PfPREXpri and anti-PfPREXheli.	
Figure 5.12.	Subcellular localisation of PfPREXpol using immunofluorescence technique.	168
<u>Chapter VI</u>	- The Evolution of <i>PfPREX</i>	
Figure 6.1.	Alignment of deduced amino acid sequences of DNA polymerase domains of DNA polymerases from family A.	179
Figure 6.2.	Phylogenetic trees generated by MEGA derived from the DNA polymerase alignment	183
Figure 6.3.	Organisation of DNA primase/helicase systems in bacteria and phages	188
Figure 6.4.	Alignment of deduced amino acid sequence of DNA primases.	194
Figure 6.5.	Alignment of deduced amino acid sequence of DNA helicases.	196
Figure 6.6.	Phylogenetic trees generated by MEGA derived from the DNA primase alignment	
Figure 6.7.	Phylogenetic trees generated by MEGA derived from the DNA helicase alignment	199
<u>Chapter Vl</u>	I – General Discussion	
Figure 7.1.	GFP containing leader sequences expressed in transfectants in <i>P. falciparum</i> .	210

1 1 1 1 いたい 外線構成 しょうせい たいていたい しょうけい しゅうしゅう いちょう しょうきょう しょう 1470 - 1810 -

<u>Appendix C</u>

Figure C.1.	Nucleotide sequence of <i>PfPREX</i>	255
Figure C.2.	Amino acid sequence of PfPREX	256
Figure C.3.	Amino acid sequence of P. chabaudi PomI sequence_	257
Figure C.4.	Amino acid sequence of PyPREX	257
Figure C.5.	Amino acid sequence of <i>PbPREX</i>	258
Figure C.6.	Partial amino acid sequence of <i>PcPREX</i>	258
Figure C.7.	Amino acid sequence of <i>PkPREX</i>	259
Figure C.8.	Amino acid sequence of <i>PvPREX</i>	259
Figure C.9.	Amino acid sequence of <i>TaPREX</i>	260
Figure C.10	Amino acid sequence of TpPREX	260
Figure C.11	. Multiple alignment of amino acid sequences of <i>Plasmodium</i> PREX	
	homologues	261

ż

÷

2

1

Appendix D

Figure D.1.	. Full alignment of deduced amino acid sequences of DNA polymerase	
	domains of DNA polymerases from family A	265
Figure D.2.	Full alignment of deduced amino acid sequence of DNA primases	271
Figure D.3.	Full alignment of deduced amino acid sequence of DNA helicases	274

List of Tables

Chapter I – Introduction

Table 1.1.	Summary of characteristics of the P. falciparum nuclear genome.	10
Table 1.2.	Classification of DNA polymerases.	29
Table 1.3.	Gene content of the 35 kb circular DNA of P. falciparum.	40
Chapter I	I - Materials and Methods	
Table 2.1.	Primers used for all PCR reactions are listed in this table.	54
Table 2.2.	Primer sequences used for the RT-PCR reactions.	56
<u>Chapter I</u>	II - Identification of a novel gene, <i>PfPREX</i> from <i>Plasmodium falciparu</i>	m
Table 3.1.	Similarity table of PREX homologues from <i>Plasmodium</i> and other apicomplexans.	98
Table 3.2.	Table showing the <i>Plasmodium</i> PREX homologues and the presence of a signal sequence and apicoplast targeting sequence.	98
Table 3.3.	Genetic distances between the PREX homologues of six <i>Plasmodium</i> species.	109
Chapter I	V - Functional analysis and subcellular localisation of the DNA polym	erase
(PTPREX)	pol) domain encoded by the PfPREX gene	
Table 4.1.	GO function search using Plasmodb for DNA-directed DNA polymerases <i>Plasmodium falciparum</i> database.	in the 116
Table 4.2.	Characteristics of other DNA polymerase I-like proteins from other source organisms.	es of 124
Table 4.3.	Comparison of key residues on Klenow and PfPREXpol.	142
Chapter V primase/h	7 - Functional analysis and subcellular localisation of the DNA elicase (PfPREXpriheli) domain of the PfPREX gene	
Table 5.1.	Putative genes found using a GO function search through the <i>Plasmodium falciparum</i> genome using DNA helicase and DNA primase function searches in Plasmodb version 4.1.	150

ź

Chapter VI - The Evolution of PfPREX

Table 6.1.	The DNA polymerase sequences used for multiple alignment.	177
Table 6.2.	Various parameters used in ClustalX 1.81 for the alignment of the DNA polymerase, DNA helicase and DNA primase sequences.	178
Table 6.3.	The DNA primase sequences used for multiple alignment.	190
Table 6.4.	The DNA helicase sequences used for multiple alignment.	191

. . . .

-

÷

200

-

List of Abbreviations

ι- Iota
α - Alpha
β - Beta
δ - Delta
ε - Epsilon
η - Eta
γ - Gamma
к - Карра
λ - Lambda
μ - Mu
θ - Theta
ζ-Zeta
μg – Microgram
3D - Three-dimensional
A Adenine
aa — amino acid
ACC – Acetyl CoA carboxylase
ACP – Acyl carrier protein
ADP Adenosine-5'- diphosphate
ALAD - δ -aminolevulinate
dehydratase
ALAS - δ -aminolevulinate synthetase
Amp – Ampicillin
APS Ammonium persulphate
ara $ATP - 1-\beta$ -D-
arabinofuranosyladenine-5'-
triphosphate
ATP - Adenosine-5'-triphosphate

ATPase - Adenosine triphosphatase bp - Base pairs BSA – Bovine serum albumin C - Cytosine cDNA – Complementary DNA Ci – Curie CO₂ – Carbon dioxide CTP - Cytosine-5'-triphosphate DAPI-4,6-diamidino-2-phenylindole dATP - Deoxyadenosine-5' triphosphate dCTP - Deoxycytosine-5'-triphosphate ddH₂O - Double-distilled water ddTTP -- Didoexy-thymidine-5'-triphosphate dGTP - Deoxyguanosine-5'-triphosphate DMSO – Dimethyl sulphoxide DNA - Deoxyribonucleic acid DNase - Deoxyribonuclease dNMP - Deoxynucleoside monophosphate DOXP-1-deoxy-D-xylulose-5-phosphate dsDNA - Double-stranded DNA DTT - 1.4-dithio-DL-threitol dTTP - Deoxythymidine-5'-triphosphate DUE - DNA unwinding element dUTP - Deoxyuridine-5'-triphosphate E. coli – Escherichia coli EDTA - Ethylenediamine tetraacetic acid EtBr - Ethidium bromide FabH - β-ketoacyl-ACP synthase III FabI - Enoyl-ACP-reductase

 $\mathbf{x}\mathbf{i}\mathbf{x}$

SPA – Scintillation proximity assay
SSB – Single-stranded specific DNA
binding protein
ssDNA – Single-stranded DNA
ssurRNA – Small subunit ribosomal
RNA
T – Thymine
TAE – Tris-acetate-EDTA
TBS – TRIS-buffered saline
TCA – Trichloroacetic acid
TE – Tris-EDTA
TEMED - N,N,N',N'tetraethylethylenediamine

 T_m – Melting point temperature TRIS – Tris(hydroxymethyl)aminomethane tRNA – Transfer RNA U – Unit UPGMA – Unweighted pair group method with arithmetric mean UTP – Uridine-S'-triphosphate UV – Ultra-violet V – Volts v/v – concentration weight to volume w/v – concentration volume to volume WHO – World Health Organisation μ M – Micromolar

 ~ 1

FITC - Fluorescein isothiocyanate FSB – First strand buffer g – Gram G – Guanine GFP -- Green fluorescent protein GO - Gene ontology GPI - Glycosylphosphatidylinositol GST - Glutathione-S-transferase GTP - Guanosine-5'-triphosphate $H_2SO_4 - Sulphuric acid$ HB - Hybridisation buffer HEPES – 4-(2-hydroxyethyl)piperazincethanesulphonic acid HGT - Horizontal gene transfer His-Histidine HPMPA - S-9-(3-hydroxy-2phosphoryl)-methoxy-propyladenine IC50 - Inhibitory concentration 50% IDF - Indirect immunofluroscence Ig – Immunoglobulin IR – Inverted repeat kb – Kilobases KCl - Potassium chloride kDa – Kilodaltons Km – Kanamycin LB – Luria-Bertani M-MolarMb - Megabases mCi -- Millicurie ME - Minimum evolution mg – Milligram

 Mg^{2+} - Magnesium ions $MgCl_2 - Magnesium$ chloride ml – Millilitre mM -- Millimolar mmol - Millimole MOPS - 3-(N-Morpholino) propanesulfonic acid MP – Maximum parsimony N₂ - Nitrogen NaCl - Sodium chloride NEM - N-ethylmaleimide ng - Nanogram NJ Neighbour-joining nm - Nanometer NTP - Nucleoside-5'-triphosphate O₂ – Oxygen OD – Optical density ORF - Open reading frame PAGE - Polyacrylamide gel electrophoresis PBS – Phosphate-buffered saline PCR – Polymerase chain reaction pol – Polymerase Poly (dT) – Deoxyribothymidylate homopolymer chain of undefined length rbc – Red blood cell RNA - Ribonucleic acid RNase-Ribonuclease rp – Ribosomal protein rRNA – Ribosomal RNA RT-PCR – Reverse transcriptase PCR SDS - Sodium dodecyl sulphate

Chapter I

Introduction

1.1. Malaria – an introduction

Over 40% of the world's population lives in tropical and subtropical areas (Figure 1.1) where malaria is endemic (WHO, 1998). In a poem written 102 years ago by Ronald Ross, malaria was described as the 'million murdering death'. This remains true today as it still affects about 300-500 million people and kills 1.5 to 2.7 million a year (WHO, 1998). Most of these deaths are children; every 40 seconds, a child dies of malaria (Bremen, 2001). Despite the efforts of different organisations, especially the World Health Organisation (WHO) and their control programmes, the disease remains a threat to public health and economic development of the endemic countries.





The rapid spread of resistance of both the parasite and the mosquito vector to currently available anti-malarials and insecticides respectively is making the control of malaria difficult. In addition to this, other factors come into play, all of which contribute to the ever-increasing disease burden. Population movements into malarious regions, deforestation and global warming are just a handful of factors that contribute to the resurgence of malaria. Therefore, there is an immediate need to find new innovative ways to combat this disease.

1.2. *Plasmodium* – the parasite

Malaria is caused by protozoan parasites of the genus *Plasmodium*. There are more than 50 species of *Plasmodium* known but only four, *P. vivax*, *P. falciparum*, *P. malariae* and *P. ovale*, cause malaria in humans. The most pathogenic species is *P. falciparum* as it causes the most severe form of malaria.

The life cycle of *P. falciparum* (and other *Plasmodium* species) involves several stages in both human and mosquito hosts (Figure 1.2a). The vector responsible for transmission of the disease is the female anopheline mosquito (Figure 1.2b). Following the bite of an infected mosquito, sporozoites injected from the insect's salivary glands enter the bloodstream. They disappear from the bloodstream and invade the liver. The hepatocytes are invaded and after an incubation period of 8 to 12 days, asexual replication of the parasite initiates, generating many daughter merozoites. Eventually these merozoites are released from infected ruptured hepatocytes into the blood circulation. This is where the erythrocytic cycle (blood stage) begins. The merozoites invade other uninfected erythrocytes.

During the erythrocytic cycle, the merozoite grows into a ring form within the erythrocyte. This is followed by the development of early trophozoites and onto



(B)



Figure 1.2. (A) Life cycle of *Plasmodium*. (B) Photograph of a female anopheline mosquito. (Source: Parasite Image Library, http://www.dpd.cdc.gov/dpdx/HTML/Image Library.htm)

mature trophozoites. These undergo multiple nuclear fissions without cell division to become schizonts which in turn give rise to daughter merozoites. Rupture of the infected erythrocytes releases merozoites into the circulation to initiate further cycles of asexual replication. This one cycle takes place over an approximately 48 hour period in *P. falciparum* but the timing is different in other species.

Some parasites, at the ring stage, develop into male and female gametocytes which may be taken up by a feeding mosquito where sexual reproduction occurs. Within the mosquito, the male gametocytes undergo exflagellation to become gametes. Fertilisation occurs in the midgut of the mosquito where the male and female gametes fuse together to form a diploid zygote. After 12 to 24 hours, the zygote elongates to form an ookinete, which in turns develops into an oocyst. Sporogony in the oocyst generates sporozoites which are released when the oocyst ruptures. The developmental time in the mosquito varies according to ambient temperature but is typically in the order of 7 to 14 days.

DNA replication of the nuclear genome of *Plasmodium* takes place at various points during its life cycle. There are at least five points within the life cycle that DNA synthesis takes place: - (1) during gametogenesis, (2) after fertilisation – meiosis-associated synthesis, (3) sporogony, (4) pre-erythrocytic schizogony in hepatocytes and (5) erythrocytic schizogony. It has been shown that DNA synthesis initiates during the carly trophozoite stage of asexual development and continues late into schizogony (Inselburgh & Banyal, 1984), while for the plastid DNA, its replication occurs just before the initiation of schizogony (Williamson *et al*, 1996) but presumably

also at other times prior to cell division to ensure all progeny retain a plastid through the life cycle.

1.3. Clinical features of malaria

1.3.1. Falciparum malaria

Falciparum malaria is the most severe form of the disease (Warrell, 1992). There are complications of falciparum malaria that, when left untreated, can be life threatening. The most notorious form of severe malaria is cerebral malaria, in which the patient becomes comatose before dying. Characteristics of cerebral malaria are the impairment of consciousness and convulsions in the patient. The mortality of cerebral malaria is about 15 to 20 % if good ospital care is provided, while in developing countries, the mortality rate can be as high as 95%.

Other features of severe falciparum malaria include anaemia, which is common in pregnant women, and hepatic dysfunction. The latter is more common in adults where the liver is enlarged and tender. Hypoglycacmia is also a common feature of falciparum malaria where the patient feels anxious, breathless and may suffer from seizures. Another complication of the disease and the most dreaded is pulmonary oedema, which is commonly associated as a terminal event.

1.3.2. Treatment of malaria

There is a pressing need for effective, safe, practicable therapeutic treatment for malaria as the number of cases of malaria around the world is increasing. The available range of standard anti-malarial drugs is narrow. There are four classes of compounds used as antimalarials: - 4-aminoquinolines, amino-alcohols, artemisinin derivatives, and antifolates. Not only are there a limited number of drugs to use against

malaria, there is also a lack of affordable new drugs coupled with the spread of drug resistant *Plasmodium*. Therefore, there is a strong need to develop new therapeutic treatments by searching for new drugs and also re-accessing old, forgotten drugs.

The amino alcohols include quinine (QN) and mefloquine and are effective against blood stage parasites. Quinine is used mainly in cases of severe malaria and also in combination with antibiotics like doxycycline. It is an erythrocytic drug and has no effect on the exo-erythrocytic phase or the gametocytic phase (Rimchala *et al* 1996). Mefloquine, which is structurally related to quinine, is used in both prophylaxis and therapy for complicated malaria. Mefloquine works by interfering with the transportation of haemoglobin products and other substances from the host cell to the parasite's food vacuole (Hellgren *et al*, 1997).

The 4-aminoquinolines, which include chloroquine, are effective against blood stages of the parasite. Chloroquine has been used as the front line drug in chemotherapy for decades as it is cheap, safe and practical for outpatient use. It is a very potent schizonticidal drug and is also effective against erythrocytic stage of all the four *Plasmodium* species. It has no effect on sporozoites, hypnozoites or gametocytes. It is a weak base, and concentrates itself in the parasite's food vacuole. It works by upsetting the transportation of toxic haem digestion waste from the parasite, which accumulates in the parasite and brings upon its demise. However, over usage and mismanagement of chloroquine have contributed to the extensive spread of resistance to this drug (Warhust, 2001).

Another group of drugs are the antifolates which affect the synthesis and utilisation of folate. These include pyrimethamine (2,4,diaminopyrimidine), a dihydrofolate reductase inhibitor and proguanil. They act by inhibiting the dihydrofolate reductase necessary for synthesis of tetrahydrofolate, a precursor in the parasite DNA synthesis. Sulphonamides (e.g. sulphadoxine) and sulphones (dapsone) act by competing for enzyme, dihydropteroate synthetase with para-aminobenzoic acid and therefore inhibit folate synthesis. They act on crythrocytic *P.falciparum*, but not sporozoites. Pyrimethamine is used together with sulphadoxine, in a combination to combat chloroquine-resistant strains of *P. falciparum*. Unfortunately, resistance to this drug combination has risen already (Sibley *et al*, 2001; Warhurst, 2002). Pyrimethamine and other antifolate drugs can also be combined with other sulpha drugs, and with Atovaquone (Malarone) and mefloquine.

The last group - Artemisinin and its derivatives kill all erythrocytic stages of the malaria parasites. It is also reported to be effective towards gametocytes (White, 1997). They are effective against both uncomplicated and severe malaria (Meshnik, 2002). It produces ultra-structural changes to the growing trophozoite parasite. A whorl is produced in the food vacuole and the parasite's mitochondria proliferate. This reduces the parasite's survival (Hien & White 1993).

There are other new drug combinations available against malaria like atovaquone-proguanil, artemether-lumefantrine and pyronaridine but more work is needed to fully validate their efficacy. In addition, antibiotics now are also roped in to combat malaria. Antibiotics like tetracycline and doxycycline are often combined with pyrimethamine or quinine for 100% cure. Clindamycin is also another candidate

antibiotic on clinical trials for use as an antimalarial drug (Lell & Kremsner, 2002). Multiple strategies need to be adopted to try to tackle not only the treatment but also the prevention of malaria. With the availability of both *P. falciparum* and *Anopheles gambiae* genomes, it is hoped that this new information will facilitate identification of novel drug targets. More information on the basis of resistance can also be used to facilitate the development of combinations in future to reduce the likelihood of the development of drug resistance.

1.4. Molecular biology of *Plasmodium*

1.4.1. The Genome of *Plasmodium*

The *P. falciparum* 3D7 nuclear genome is composed of 22.8 Mb of DNA distributed among 14 chromosomes ranging in size from approximately 0.6 to 3.3 Mb. The genome has been fully sequenced (Gardner *et al*, 2002) and is found to be 80.6% A-T rich. (*E. coli* is approximately 50% A-T and the human nuclear DNA contains 63% A-T.) Such an A-T rich genome is reflected in a skewed codon usage in *P. falciparum* (Saul & Battistutta, 1988; Weber, 1987). Most of the third positions of codons in *P. falciparum* contain either an A or a T where this position can vary without changing the amino acid.

The *P. falciparum* genome is predicted to encode 5,268 gene products which include proteins, tRNA and rRNA, (Gardner *et al*, 2002) with a mean gene length of approximately 2.3 kb. Other characteristics of the *P. falciparum* genome are summarised in Table 1.1 (Adapted from Gardner *et al*, 2002).

Feature	P. falciparum
Size of genome	22.8 Mb
AT content (%)	80.6
No. of genes	5268
Mean gene length excluding introns	2.3 kb
Gene density (kb per gene)	4.3
% coding	52.6
Genes with introns (%)	53.9

Table 1.1. Summary of characteristics of the P. falciparum nuclear genome.

An unusual feature of *Plasmodium* genome is the presence of genes combined to give bifunctional proteins. Many of these proteins are found in other organisms as single functional entities. For example, there is the bifunctional dihydrofolate reductase-thymidylate synthase enzyme (DHFR-TS) (Bzik *et al*, 1987) while both DHFR and TS exist as distinct, monofunctional enzymes in bacteriophages, bacteria and mammals (Blakey & Benkovic, 1984). These enzymes are involved in the folate metabolic pathway. There are at least another two bifunctional enzymes in *P. falciparum* that are also involved in the same pathway. The first is the dihydropteroate synthetase-dihydro-6-hydroxymethylpterin pyrophosphokinase (Triglia & Cowman, 1994) and the second is the dihydrofolate synthetase-folylpolyglutamate synthetase (DHFS-FPGS) (Salcedo *et al*, 2001). Other bifunctional enzymes include the ornithine decarboxylase-S-adenosylmethionine decarboxylase (ODC-AdoMetDC) (Müller *et al*, 2000) which is involved in polyamine biosynthesis, and the glucose-6-phosphate dehydrogenase-6-phosphogluconolactonase (G6PD-6PGL) (Clarke *et al*, 2001), which is involved in the first two steps of the pentose phosphate pathway. All of the examples given above are involved in the metabolic pathways, other examples include two large integral membrane proteins which possess guanylyl cyclase and adenyl cyclase activities (Carucci *et al*, 2000).

With all these examples of bifunctional proteins found in P. falciparum, one might wonder about the evolution of the genes encoding these proteins. Apparently, it seems that gene fusion in P. falciparum is relatively common. It has been speculated that there is biological advantage for bifunctional proteins in metabolic pathways, as this arrangement optimises substrate channelling and the formation of products with minimum regulation involved (Ivanetich & Santi, 1990).

1.4.2. Extrachromosomal DNAs

Other than the nuclear genome, *Plasmodium* possesses two distinct extrachromosomal DNAs or episomes. One of these is a multi-copy 6 kb element, also known as the mitochondrial DNA (Feagin, 1992). The other is a 35 kb circular episome (Gardner *et al*, 1988) that has been determined to be similar to the plastid genome of plants but was once thought to be of mitochondrial nature (Feagin *et al*, 1991). This non-photosynthetic 'chloroplast'-like genome was isolated in *P. falciparum* (Gardner *et al*, 1988) and then fully sequenced (Wilson *et al*, 1996).

The mitochondrial DNA was first discovered as tandem repeats in *P. yoelii* (Vaidya & Arasu, 1987). It is present in multi-copies within *Plasmodium*, ranging from 15 per cell in *P. gallinaceum* (Joseph *et al*, 1989) to 20 per cell in *P. falciparum* (Preiser

et al, 1996) and 150 per cell in *P. yoelii* (Vaidya & Arasu, 1987). The mitochondrial genome of *P. falciparum* has been shown to carry genes for some characteristically mitochondrial proteins like cytochrome b and subunit I of cytochrome oxidase (Vaidya *et al*, 1989). It is the smallest known mitochondrial genome.

1.4.2.1. Plastid DNA

Unlike the mitochondrial DNA, the plastid DNA is only found as either one or two copies in *P. falciparum* (Wilson & Williamson, 1997). When first discovered, some scientists mistook it to be associated with the mitochondrial DNA, being part of a multipartite mitochondrial system (Feagin *et al*, 1991). The plastid DNA has been fully sequenced (Wilson *et al*, 1996) (Figure 1.3) and on closer examination, this 35 kb circle bears no homology to mitochondrial DNA. Instead, its ribosomal RNA (rRNA) genes are arranged in a chloroplast-like inverted repeats format (Gardner *et al*, 1991a, b). The inverted repeat (IR) covers about a third of the plastid DNA and encodes the duplicated large and small subunit rRNA genes (Gardner *et al*, 1994). The total length of the unit repeat is approximately 5.1 kb. The two repeats are designated IR_A and IR_B – using the nomenclature previously established for chloroplast DNA (Gardner *et al*, 1993). DNA replication is believed to initiate within these repeats (Singh *et al*, 2003).


Figure 1.3. Gene map of the plastid DNA of *P. falciparum*. Colour-coded genes:- yellow- RNA polymerases, light blue – unknown ORFs, orange – putative ribosomal proteins, dark blue – large subunit rRNA, red – small subunit rRNA. (Source: Wilson *et al*, 1996).

A prominent feature of the plastid is found along the IR_B region where a string of 15 putative ribosomal protein (rp) genes is found. These genes are found along a 7 kb stretch where they are found to be important in the initial assembly of the 30S ribosomal subunit (Wittmann, 1983). The rpoB and rpoC genes that are found on the IR_A region have helped to ascertain the plastid origin of the 35 kb circle (Gardner *et al*, 1991a, b). Other plastid genes include the *tufA* which codes for the protein synthesis factor, Tu and downstream to it is an open reading frame that appears to code for a gene homologous to the *clpC* belonging to a molecular chaperone family.

1.4.3. DNA replication

In all cellular organisms from bacteria to humans (but excluding a variety of viral species), genetic information is locked within a double helix formed by two antiparallel DNA stands. Thus, all replication processes involve the melting of these two stands of DNA followed by the polymerisation of each complementary strand on the resulting single-stranded templates (Kornberg & Baker, 1992). The level of complexity involved in the enzymatic machinery required for DNA replication is considerably greater than might have been expected. The enzymes involved in DNA replication will be discussed in detail later. The processes that are common to most DNA replication pathways are detailed below.

DNA replication usually starts at the origin of DNA replication, a particular sequence which increases the efficiency of initiation of DNA replication. It provides a locus for the assembly of a multi-protein complex which includes various enzymes, proteins and accessory factors required for DNA synthesis to take place. The initiation of DNA replication involves three sequential steps. The first step involves specific proteins binding to the origin of replication. In *E. coli*, it is the DnaA protein that binds the replication origin (Ori C). OriC is highly conserved among Enterobacteria (Zyskind *et al*, 1983) and has several distinctive features. These include the presence of DnaA boxes which are repetitive 9mers with sequences 5'-TTATC/ACAC/AA-3' in the form of two inverted repeats. The positions of these DnaA boxes are highly conserved too (Zyskind *et al*, 1983). These are responsible for binding to the DnaA protein (Fuller & Kornberg, 1983; Matsui *et al*, 1985). Located to the left of the DnaA boxes, are a series of AT- rich 13mers. These are arranged in a series of three and they are involved in the process of localized denaturation required to initiate DNA replication.

Я

The DnaA protein serves at least two functions. The first is to initiate DNA unwinding using a DNA helicase while the second function is to guide other replication proteins to the origin. The DnaA protein itself is an ATPase that requires the presence of ATP (which is hydrolysed to release ADP) in order to bind to OriC. Therefore, initiation of DNA replication starts with the DnaA protein binding to the DnaA box sequences found at OriC, forming an initial open complex. Next, the unwinding of the 13mers takes place, leading to the second step of DNA replication, which is the DNA unwinding of the two complementary DNA strands. The enzyme involved in this step is a DNA helicase. DnaA is responsible for guiding the DNA helicase to the open complex. In *E. coli*, the DNA helicase responsible for the replication is called DnaB.

DNA unwinding usually begins at an easily unwound DNA sequence, referred to as the DNA unwinding element (DUE). A DUE is identified by *cis*-acting mutations that increase the stability of the double helix (Umek & Kowalski, 1990). These DUEs, it seems, do not have consensus sequences. It was observed that the binding of proteins to OriC, results in DNA unwinding in the DUEs in *E. coli* (Kowalski & Eddy, 1989) before DnaB can start the unwinding of the double helix.

The initial opening of the duplex allows the establishment of a replication fork. The most important step in this process is the loading of a DNA helicase onto the DNA strands. In *E. coli*, one molecule of DnaB, a hexamer of identical subunits, clamps around each of the two single-stranded DNA in the open complex formed between OriC and DnaA. This is now called the pre-priming complex. Along with the DNA helicase, DNA unwinding is also facilitated by single-stranded specific DNA binding proteins (SSB), which bind to single-stranded DNA fibres and prevent them from reannealing; and topoisomerase I, which releases torsional stress generated by the unwinding of the DNA double helix (Figure 1.4). The DNA helicase, using energy provided by NTP hydrolysis, 'translocates' along the DNA strand, melting the hydrogen bonds that keep the DNA strand in association with its complementary strand.



Figure 1.4. A DNA replication fork whereby DNA unwinding is facilitated by a DNA helicase, a topoisomerase (DNA gyrase) and single-strand binding proteins (SSB). DNA primase then make short RNA primers that are used by DNA polymerase to begin synthesising complementary strand.

The final step to the initiation of DNA replication is the commencement of DNA synthesis. This involves several replication proteins, including DNA primase and DNA polymerase. The unwound DNA strands need to be primed by a DNA primase so that the DNA polymerase can elongate the primer strands. In most eukaryotic cells and some prokaryotic cells, DNA synthesis is generally primed by short RNA chains. The DNA primase known as DnaG, is responsible for the synthesis of this primer in *E. coli*.

The primase is usually recruited to a segment of single-stranded DNA by first binding to the DnaB hexamer. This complex is called a primosome. Once the DNA primase synthesises the short primer RNA molecules complementary to the DNA, it dissociates from the primosome complex.

Now that most of the enzymes are assembled, a replication fork is formed; and the elongation of the new DNA fibres begins. The DNA polymerase responsible for replication in *E. coli* is the DNA polymerase III. DNA replication proceeds by a semicontinuous mechanism where one DNA strand (the leading strand) is elongated continuously from a single primer, whereas the other DNA strand (the lagging strand) is constructed from many short DNA chains elongated from multiple primers. Polymerisation of the leading strand occurs in the same direction as the replication fork movement while the lagging strand is synthesised in the opposite direction (Figure 1.5). The leading strand is synthesized in the 5' - 3' direction continuously, while the lagging strand synthesis is more complicated. Since DNA polymerase can only add nucleotides to the 3' end of a primer or DNA strand, the elongation of the lagging strand is also in the *p+4X5'-3' direction, although the overall direction of the lagging strand is from its 3'end towards the 5' end. For this to occur, short fragments called Okazaki fragments synthesised in the 5'-3' direction, are made discontinuously from many RNA primers after which the fragments are joined by DNA ligase.



Figure 1.5. Replication fork showing the leading strand and the lagging strand and their direction of strand elongation.

There have been two models proposed for the movement of DNA replication – the factory model versus the tracking model. The tracking model proposes that the DNA replisome moves along the DNA template while the factory model proposes that the DNA replisome is in a fixed position and only the unreplicated DNA is pulled through the stationary replisome (Dingman, 1974; Pardoll *et al*, 1980). The tracking model seems unlikely, as DNA replication based on free moving molecules would result in chaotic entanglement of the newly synthesised DNA strands. The factory model on the other hand, allows the two newly replicated DNA strands to move bidirectionally, away from the unreplicated DNA, thus preventing entanglement. Additionally, both the DNA polymerase and DNA helicase are 'powerful' motors, and as such may be able to provide the force responsible for the movement of the DNA strands. Experimental evidence has been obtained to support the factory model (Lemon & Grossman, 1998)

1.4.3.1. Enzymes involved in DNA Replication

From prokaryotes to eukaryotes, the double stranded nature of DNA provides an elegant structure to ensure the accurate duplication and repair of DNA. This is due to the complementary base pairing of the DNA, which also poses a problem as the hydrogen bonds that hold the double helix together must be disrupted. In order to achieve this, a full repertoire of proteins works together. The main proteins involved are DNA helicase, DNA primase and DNA polymerase. Their functions and roles are detailed below.

1.4.3.1.1. DNA helicases

All key processes that are essential for the maintenance and faithful transmission of DNA ranging from DNA replication, repair and recombination requires the unwinding of the DNA double helix to provide single-stranded DNA as a template. DNA helicases are the enzymes responsible for the catalysis of the unwinding of the double helix. They are molecular motors that couple the energy of nucleoside-5'-triphosphate (NTP) hydrolysis to the transient disruption of hydrogen bonds formed between the Watson-Crick base pairs of the DNA double helix. Hence, helicases are also NTPases.

The first DNA helicase isolated on the basis of its nucleic acid unwinding activity was in 1976 (Abdel Monem & Hoffman-Berling, 1976). It was identified in *E. coli*, and was designated as DNA helicase I. DNA helicase I is encoded by the *tra*I gene of the F episome (Abdel-Monem *et al*, 1983) and is required to unwind the DNA that is transferred during conjugation. Since then, multiple helicases have been identified from prokaryotes (Matson & Kaiser-Rogers, 1990) and eukaryotes (Thommes & Hubscher,

1992). Some of these are putative DNA helicases that have been identified from genome sequence projects while a handful with helicase activities have been isolated and characterised. The omnipresence of these DNA helicases in many sources suggests that these enzymes are ubiquitous in nature. Moreover, it seems that different DNA helicases have specific functions resulting in the presence of multiple DNA helicases in a single organism. For example, in *E. coli*, there are at least 11 distinct DNA helicases (Matson, 1991), eight in calf thymus and four in yeast (Thommes & Hubscher, 1992).

The basic biochemical reaction catalysed by DNA helicase is the hydrolysis of NTP and it is based on this activity that DNA helicases are often characterised. The DNA helicase assays follow the conversion of double-stranded DNA into single strands. The most common assay employs a DNA substrate that is labelled and incubated with Mg²⁺ and an NTP. The DNA substrate is usually a ³²P-labelled oligonucleotide annealed to a larger single-stranded DNA molecular such as M13 viral DNA. The products of the reaction are resolved either on an acrylamide or agarose gel. The radioactive substrate and product molecules differ greatly in size so that the quantity of the substrate that has been rendered single stranded by the helicase can be evaluated.

Many DNA helicases require a region of single-stranded DNA adjacent to the duplex region to serve as an initial binding site. This site also serves to facilitate the formation of an initiation complex. Based on the requirement for a single-stranded overhang, different DNA helicases show different preferences for the polarity of this overhang. Some DNA helicases show a strong preference to utilise a 3' terminated single-stranded DNA. These are the 3'-5' DNA helicases. Others prefer a 5' overhang. These are the 5'-3' DNA helicase group. There are of course exceptions to the case such as the *E. coli* RecBCD protein, a recombination helicase which does not require a single-stranded DNA overhang but prefers to unwind blunt-ended DNA (Taylor & Smith, 1985). Interestingly, RecB alone without RecC and D subunits, prefers to unwind a 3' single-stranded DNA overhang (Boehmer & Emmerson, 1992).

1.4.3.1.1.1.Molecular biology of DNA helicases

DNA Helicases are divided into five main superfamilies and families based on the presence and composition of conserved amino acid residues (Helicase signatures) (Figure 1.6). Comparison of the DNA helicase sequences has revealed various categories of sequence homology. Three vast superfamilies and two smaller families have been identified (Gorbalenya & Koonin, 1993). The superfamilies 1 and 2 (SF1 & SF2) are the largest groups, containing the majority of DNA helicases including examples from viral, prokaryotic and eukaryotic organisms. The members from both families are identified by the presence of six conserved amino acid motifs (I-VI) (Gorbalenya & Koonin, 1993). A seventh motif, termed Ia, is found only in a subset of helicases within the superfamilies and has a poorly defined consensus sequence.

Superfamily 3 (SF3) is much smaller and the members of this superfamily possess only three conserved motifs (Gorbalenya & Koonin, 1993). These DNA helicases are generally found in small DNA and RNA viruses (Gorbalenya *et al*, 1990). Family 4 is a smaller family group which comprises of DNA helicases found in the bacterial and bacteriophage systems. These helicases are always associated physically with DNA primases (Ilyina *et al*, 1992). This group is also known as the DnaB-like helicases, because the type-member of this group is the DnaB protein from *E. coli* other family 4 members share homology to the DnaB helicase. They have three distinct

÷

conserved motifs (Ilyina *et al*, 1992), in addition to the Walker A and B sequences. The last family of helicases are the DNA-RNA helicases that could not be grouped with the other 4 families (Gorbalcnya & Koonin, 1993). An example is the bacterial transcription termination factor, Rho (Brennan & Dombroski, 1997).

The two motifs that are present in all known DNA helicases were first described as the Walker A and B sequences (Motif I and II respectively). The Walker motifs are found in all NTP-binding proteins (Walker et al, 1982). The residues in the Walker A and B sequences are believed to interact with Mg²¹NTP/Mg²¹NDP. The Walker A motif has a consensus sequence of GxxxxGKT (Walker et al, 1982) (Figure 1.7) where the last three residues form an essential part (GKT/S) of the motif. It was suggested that the amino group of the lysine interacts with the phosphates of Mg²⁺NTP/Mg²⁺NDP while the hydroxyl group of the threonine/serine complexes with the Mg²⁺ ion (Walker et al 1982). The Walker B motif has a consensus sequence of DExx where the carboxyl group of the aspartate functions to co-ordinate the Mg^{2+} ion of Mg²⁺NTP/Mg²⁺NDP while the glutamate is suggested to act as a catalytic base during NTP hydrolysis. Through X-ray crystal structure studies, it was noted that other residues within the Walker B motif also have a role to play, extending the consensus sequence of the Walker B motif to DE(Y/F/A)QD (Figure 1.7) where the aspartate in the fifth position is observed to form a salt bridge with the glutamine residue found in motif VI (Caruthers et al, 2000; Korolev et al, 1997). This is only observed in helicases from SF1 (Soultanas et al, 1999; Velankar et al, 1999). Among helicases from SF2, the Walker B consensus sequence is depicted as - DEx(D/H).



Figure 1.6. The main groups of DNA helicases. There are four groups of helicases (SF1-SF3, F4). (Source: adapted from Gorbalenya & Koonin, 1993).



Figure 1.7. A schematic diagram representing the motifs of DnaB like helicases. Open boxes represent the conserved motifs, and the letters inside the boxes are the consensus amino acid sequences of each motif. Labels above the open boxes are the names assigned to the motifs. The consensus sequences were derived from Ilyina *et al*, 1992. Single-letter amino acid abbreviations were used. The '+' represents a hydrophobic residue, an 'o' represents a hydrophilic residue and an 'x' represents a residue that is not restricted to being hydrophobic or hydrophilic. (Adapted from Hall & Matson, 1999).

ŝ

1.4.3.1.1.2. Molecular biology of DnaB-like helicases

The members of this family consist of bacterial and bacteriophage helicases. The DNA helicase in PfPREX shows the strongest homology to this group of DNA helicases. Therefore, this section gives more details of this group of DNA helicases.

The key representative type-member of this family is DnaB from *E. coli*. Ilyina *et al* (1992) carried out an alignment of DnaB-related helicases and found that they possess five distinct motifs of which two are the Walker A and B sequences (H1 and H2 respectively). Other examples of family members are T7 gene 4 helicase and T4 gene 41 helicase. They belong to the group of hexameric helicases whereby six helicase subunits assemble into a hexameric ring with a central channel to unwind doublestranded DNA. The process involves one strand passing through the channel while the other passes outside the ring (West, 1996). The other three motifs (H1a, H3 and H4) also play key roles too in the unwinding of DNA. Motif H1a contains a glutarnate that has been suggested to play a role in catalysis (Sawaya *et al*, 1999). Motif H3 has been suggested to participate in the interaction of Mg²⁺ in NTP/NDP hydrolysis and contributes to the binding of oligonucleotides (Caruthers & McKay, 2002). Finally motif H4 has a role linked with DNA binding (Gorbalenya & Koonin, 1993; Washington *et al*, 1996).

1.4.3.1.2. DNA primases

DNA primases are another essential component of the DNA replication machinery. Since all DNA polymerases are unable to initiate polymerisation *de novo*, they require a primer molecule to provide a 3'-OH to which the first nucleotide is added (Kornberg & Baker, 1992). DNA primase synthesises short RNA molecules which then serve as primers for both the initiation of leading stand replication at the origin of replication and of Okazaki fragments on the lagging strand (Kornberg & Baker, 1992). DNA primases have been identified in bacteria, viruses and eukaryotes. With the recent explosion in genome sequencing, a multitude of proteins have been discovered that show homology to known DNA primases.

DNA primases can be classified into two groups (Frick & Richardson, 2001). The first group contains DNA primases from bacteria and bacteriophages, while the second is made up of eukaryotic DNA primases. The two groups differ in their structure and their relationship with other proteins in the replication complex.

1.4.3.1.2.1. Prokaryotic DNA primases

Prokaryotic DNA primases include proteins from bacteria and bacteriophages. The prototypical bacterial primase is the enzyme, DnaG from *E. coli* (Kornberg & Baker, 1992). All other known bacterial primases, as well as the bacteriophage primases, are homologous to DnaG. They all possess similar functional characteristics. Proteolytic digestions of DnaG from *E. coli* and phage T7 primase have shown that the prokaryotic DNA primases are composed of three regions (Marians, 2000; Washington *et al*, 1996), an N-terminal zinc-binding domain, an oligonucleotide synthesis site and either a helicase or a region that interacts with a helicase.

The zinc-binding domain contains the "zinc motif" which is defined as two pairs of residues, made up by either cysteines or histidines or both (Motif 1) (Figure 1.8). The configuration of the zinc motif allows for the formation of a zinc finger which is essential for primase activity. This zinc motif is present in many primases from both

prokaryotic and eukaryotic sources (Ilyina *et al*, 1992; Mendelmann *et al*, 1994; Toh, 1986). The position of the zinc-binding domain varies between different classes of primases. In prokaryotic primases, the zinc-binding domain is located at the N-terminus of the polypeptide while it is found at the C-terminus of primases of eukaryotic viruses. For standard eukaryotic primases, the domain is located in the central catalytic domain (Figure 1.8).



Figure 1.8. Schematic diagram of the domain structure and the conserved residues in DnaG-like primase. The six consensus sequences identified by Ilyina *et al*, 1992 are numbered at the top of the boxes.

The importance of the zinc-binding domain to the primase activity is demonstrated in the phage T7 gene 4 protein which codes for both a DNA primase and helicase (Mendelmann *et al*, 1994). When the zinc motif is disrupted in the protein, priming activity is lost. The same loss of activity is observed with the phage P4 primase when the two cysteine residues in the zinc motif are replaced by glycine residues (Ziegelin *et al*, 1995).

The central domain or catalytic domain contains the majority of the conserved motifs (Salas, 1991). This domain is also known as the RNA synthesis domain, which is responsible for synthesis of the primer strand. It contains five conserved motifs (Motifs 3-6). Motif 2 is the only motif for which no functional or

structural role has yet been assigned. Motifs 3-6 contain residues that appear to play roles in NTP binding and oligonucleotide synthesis. Additional charged residues in the proximity of motif 3 are shared among several prokaryotic and eukaryotic RNA polymerases (Versalovic & Lupski, 1993). Experiments using E. coli DnaG confirmed that this region plays a role in RNA synthesis. Motifs 4 to 6 may play roles in the coordination of NTP and Mg²⁺ ions. Motifs 4 and 5 contain conserved negatively charged residues that are preceded by runs of hydrophobic residues. Within this arrangement, the residues are predicted to form beta strands which resemble motifs involved in Mg²⁺-mediated NTP binding observed in ATPases (Fry et al, 1986, Gorbalenya & Koonin, 1989). Motif 4 has the consensus sequence of EGxxD where the glutamate residue is predicted to be essential for primer synthesis (Strack et al. 1992). Aravind *et al* (1998) predicted that the same glutamate is needed to facilitate nucleotide polymerisation. Motif 5 has a consensus sequence of DxDxxGxxA where the DxD sequence acts to coordinate Mg²⁺ (Aravind et al, 1998). The same DxD sequence can also be located on motif 6, where it is conserved in all DnaG-like bacterial primases (Szafranski *et al*, 1997) but only the first aspartate is present in phage primases.

Parts of the catalytic domain seem to have structural relationships with otherwise unrelated proteins such as some DNA topoisomerases and several nucleases (Aravind *et al*, 1998). This shared region is named the TOPRIM domain (for topoisomerase/primase). The TOPRIM domains in primases are smaller than those in topoisomerases (~80 residues as compared to ~120 residues in topoisomerase) (Podobnik *et al*, 2000). Motifs 4 and 5 are found within the TOPRIM domain and the DxD dyad is central to this region.

The C-terminal domains of prokaryotic primases are not evolutionary conserved. This region is required for the association with a helicase or itself is a helicase. Therefore, other than where the C-terminus is a helicase, conserved motifs are not generally present within this domain. Those that possess a helicase at the C-terminus include bacteriophage T7 and P4 primases while those that do not, include phage T4 primase and *E. coli* DnaG. *E. coli* DnaG has been shown to interact with DnaB helicase via this region (Tougu *et al*, 1994). When the C-terminal domain of DnaG is removed by proteolysis, the protein can still synthesise primer strands but is unable to engage in any helicase activity (Tougu *et al*, 1994) because the DnaG no longer associates with DnaB.

1.4.3.1.3. DNA polymerases

DNA polymerases are essential enzymes for DNA replication, recombination and repair. They are involved in the catalysis of the synthesis of DNA in a sequence-independent manner that results in a faithful copy of the original DNA molecule. Because DNA polymerases are involved in many separate functions, an extreme diversity of DNA polymerases exists, not just among different organisms but also within individual organisms. For example, in *Saccharomyces cerevisiae*, there are at least eight different DNA polymerases (Hubscher *et al*, 2000). These enzymes can function alone or with accessory factors. They can vary not only in their fidelity, speed and processivity with which they replicate DNA but also in the functions that they perform.

Based on their functions, DNA polymerases can be broadly classified into two groups; the first are the replicative DNA polymerases and the second, the repair polymerases. Replicative DNA polymerases are responsible for DNA duplication while repair polymerases fix damaged DNA. The discovery of *E. coli* DNA polymerase I in 1956 (Kornberg, 1960) yielded the first insight into the structure and mechanisms of this group of enzymes. Recently, with the explosion of genome sequencing projects, more DNA polymerase sequences have been discovered which allows us to use this information to further understand the workings of the enzyme.

Name	DNA polymerase family	Function
pol (alpha) α	В	Replication
pol (beta) β	X	Repair
pol (gamma) γ	A	Mitochondrial replication
		and repair
pol (delta) δ	В	Replication, repair
pol (epsilon) ε	В	Replication, repair
pol (zeta) ζ	В	Translesion synthesis
pol (eta) η	Y	Translesion synthesis
pol (theta)θ	A	Cross-link repair
pol (iota) ι	Y	Translesion synthesis
pol (kappa) к	Y	unknown
pol (mu) μ	X	Somatic hypermutation
pol (lambda) λ	X	Meiotic mutation.

Table 1.2. Classification of DNA polymerases. (Adapted from Friedberg et al, 2000).

The DNA polymerases have also been grouped into different families based on their amino acid sequence homologies to *E. coli* DNA polymerases (Ito & Braithwaite, 1991). Family A polymerases are homologous to the *pol*A gene product, *E. coli* DNA polymerase I. Other members include bacterial and bacteriophage polymerases like *Thermus aquaticus* DNA polymerase I, phage T7 DNA polymerase, eukaryotic mitochondrial DNA polymerase gamma (γ) and DNA polymerase theta (θ). DNA polymerase θ is a polymerase that was identified in *Drosophila melanogaster* as Mus308 that is shown to be involved in the repair of DNA crosslinks (Harris *et al*, 1996) and later the human homologue of Mus308 was identified too (Sharief *et al*, 1999). Like DNA polymerase θ , the family A DNA polymerases function in repair (e.g. PolI in nucleotide excision repair (Sancar, 1996) and also in replication (e.g. phage T7 DNA polymerase (Fuller & Richardson, 1985). Family A DNA polymerases contain three conserved motifs (A-C) (Sousa *et al*, 1996), where motifs A and C are found in the catalytic site of the enzyme while motif B is involved in the binding of dNTP.

Family B DNA polymerases comprise bacterial, bacteriophage and eukaryotic enzymes, which are homologous to the *E. coli* DNA polymerase B (pol II). Members include eukaryotic replication polymerases like polymerase alpha (α), polymerase delta (δ) and polymerase epsilon (ϵ) (Hubscher *et al*, 2000) and also phage T4 replicative polymerase (Karam & Konigsberg, 2000). In type B polymerases, there are six conserved motifs identified. Motifs I and II are found in the catalytic site. These two motifs are considered to be equivalent to family A polymerase motifs C and A respectively (Joyce, 1997; Steitz, 1999).

There are other polymerase families like the family C DNA polymerases and family X polymerases. Family C polymerases are homologous to *E. coli* polymerase C (pol III) (Kelman & O'Donnell, 1995) while family X members do not share significant sequence similarity with DNA polymerases but are related to terminal transferases. Members of family C comprise most of the bacterial replication polymerases while family X members includes eukaryotic DNA polymerase beta (β), lambda (λ) (Garcia-Diaz *et al*, 2000) and mu (μ) (Dominguez *et al*, 2000). With the ever-growing size and



Figure 1.9. 3D structure of *E. coli* Klenow fragment and *Thermus aquaticus* Taq polymerase. (Source:Klenow- PDB file 1KFD, Beese *et al*, 1993 and Taq –PDB file 1KTQ, Korolev *et al*, 1995). Both structures possess similar overall structure which comprise the characteristic finger, thumb and palm.

Other residues including Arg 754, Arg 682, Lys 758 and His 734 have been implicated in the interaction with dNTPs. These sets of residues are well conserved among family A DNA polymerases. They are all positively-charged residues and are thought to interact with the phosphate backbone of dNTPs (Astatke *et al*, 1995). Residues that are important and play roles in primer strand attachment are Glu 611, Thr 609, Asn 678, Asn 675, Lys 635 and Arg 631 (Basu *et al*, 1988).

The detailed structure of the *E. coli* DNA polymerase provides a platform for the functional studies of other DNA polymerases I. Still only a small area of the structure has been studied. More has to be carried out to study the interactions of the various residues with the dNTPs, DNA template, primer and other replication proteins found on the replication fork.

1.4.3.2. DNA Replication in *Plasmodium*

One reason that the replication of DNA in malaria parasites is the subject of research is because it provides a gateway for potential therapeutic applications. DNA replication occurs at several points during the life cycle of the parasite. One time point is when after parasites invade hepatocytes of the host and undergo exo-erythrocytic schizogony. DNA replication also occurs during both erythrocytic schizogony and gametogenesis, and during sporogony in the mosquito. In fact, DNA replication during gametogenesis and exflagellation appears to be one of the most rapid known examples of DNA replication. It would be of interest to understand this process in more detail.

In *P. falciparum*, two distinct DNA polymerase classes were purified (Chavalitshewinkoon *et al*, 1993), one being aphidicolin-sensitive (A^{S}) and the other aphidicolin-resistant (A^{R}). To further characterise the two polymerases, they were treated with concentrations of different compounds to obtain IC_{50} values. The A^{R} polymerase was sensitive to N-ethylmaleimide (NEM) and dideoxy-TTP (ddTTP) while the A^{S} polymerase was sensitive to 1- β -D-arabinofuranosyladenine-5'-triphosphate (araATP). The IC₅₀ values obtained for both polymerase classes suggested that the A^{R} fraction resembled the eukaryotic mitochondrial DNA polymerase γ while the A^{S} fraction possessed primase activity and is classified as DNA polymerase α . In *P. berghei*, three distinct DNA polymerase classes were purified (De Vries *et al*, 1991), two being A^{S} and one A^{R} . However, unlike in *P. falciparum*, the fraction that was shown to be processive was able to use an RNA-primed template but lacked primase activity. It was thus classified as a DNA pol α -like enzyme. Based on the inhibition profile of the second A^{S} fraction, it was classified as DNA polymerase δ . The A^{R}

fraction of *P. berghei* unlike the *P. falcipurum* fraction was resistant to both ddTTP and NEM and was classified as a DNA polymerase β .

Inhibitors of DNA polymerase have been investigated as potential therapeutic agents. For example acyclic nucleoside analogues, which are related to the S-9-(3-hydroxy-2-phosphoryl methoxy-propyl)adenine (HPMPA), were shown to strongly inhibit the growth of *P. falciparum* (De Vries *et al*, 1991). A single injection of HPMPA prevents any increase in parasitaemia in mice infected with *P. berghei* (De Vries *et al*, 1991) for several days. Repeated drug administration of HPMPA blocked parasite growth for long periods but when treatment was stopped, parasitaemia returned.

Nuclear DNA replication is much better understood than organellar DNA replication. However, it is critical to understand this latter process as enzymes of organellar DNA replication provide good targets for drugs against the parasites, especially if they are of plastid and cyanobacterial/prokaryotic origin and as such, likely to be different from the host's DNA replication enzymes. As yet, plastid DNA replication has not been well explored or understood in *Plasmodium*. Even information from the conventional plastid genomes is of little help. What is known is that the malaria plastid DNA does not carry any genes for plastid DNA replication. They are most probably carried by the nuclear genome but had not been identified prior to this thesis.

1.4.4. Plastid DNA replication

Until recently, nothing much was known about the replication of the *Plasmodium* plastid DNA. The plastid DNA, during the erythrocytic cycle, replicates early in schizogony at about the same time as the nuclear and mitochondrial DNA. The pattern of replication was presumed to be similar to higher plant chloroplast, i.e. the plastid replicates bidirectionally using the rolling circle mode of replication. Williamson *et al*, (2002) discovered that the plastid of *Plasmodium* in fact replicates by two different mechanisms, one that adopts the rolling circle strategy, initiating at an unknown location while the second mechanism, which initiates at the twin D-loops located in the large inverted repeat. This mechanism is relatively sensitive to the topoisomerase inhibitor, ciprofloxacin unlike the first mechanism which is less sensitive to it.

The enzymology of plastid replication is not well studied. There are only two replication genes that have been found that are associated with plastid DNA replication. The genes are gyrA and gyrB which encode gyrase subunits have been shown to be nuclear encoded plastid proteins (Wilson *et al*, 2002). Therefore, they are most probably involved in the plastid DNA replication. White & Kilbey (1996) proposed that the aphidicolin-insensitive DNA polymerase identified by De Vries *et al* (1991) could play a role in plastid DNA replication. As yet though, no other direct evidence has been obtained or enzymes isolated that have been implicated directly in the replication of the plastid DNA, the enzymology of which remains elusive.

1.5. Apicoplast

1.5.1. Introduction

Plastids include the photosynthetic chloroplasts of plants, the rhodoplasts of red algae and the chromoplasts of brown algae. The apicoplast also belongs to this family of organelles. The apicoplast is found in a group of parasites which belong to the phylum Apicomplexa. The members of the Apicomplexa include *Toxoplasma*, *Plasmodium* and *Theileria* among many other groups. Several of them cause severe diseases in livestock and humans such as theileriosis, toxoplasmosis and malaria.



Figure 1.10. Ultrastructure of a ring stage *Plasmodium falciparum* showing the various organelles like the plastid and mitochondrion. (Source: Bannister *et al*,2000)

When the non-photosynthetic plastid was first discovered in apicomplexan parasites, it attracted a lot of interest for its potential as a drug target. Its evolutionary history indicates that an ancestor of the apicomplexan parasite must have endocytosed a symbiotic alga. A great deal of work has been carried out to learn and understand more of this interesting organelle.

1.5.2. Origin and evolution of the apicoplast

The apicomplexan plastid is thought to be a product of a secondary endosymbiotic event where instead of engulfing a cyanobacterial prokaryote, the apicomplexan ancestor gained its plastid by taking up a eukaryote that had already acquired a primary plastid. Ultrastructural and molecular data support the notion that the apicoplast originated from such an event. Hopkins *et al* (1999) postulated that three membranes surround the *Plasmodium* plastid while others have shown that there are four membranes surrounding the plastid (Kohler *et al*, 1997). The two inner membranes correspond to the typical plastid envelopes, while the outermost membrane belongs to the host endomembrane system. Regardless of the finer details, the plastid does seem to originate from a serial endosymbiotic event (Delwiche & Palmer, 1997).

However, there seems to be confusion over the precise origin of the plastid in Apicomplexa. There are arguments over whether the plastid originated from a red alga or green alga. Barly evidence supported the green alga lineage where phylogenies constructed from ribosomal RNAs and ribosomal proteins suggested an evolutionary affiliation with the euglenoid plastids (Gardner *et al*, 1994; Howe, 1992). Sequence analysis of the *tuf*A gene, coding for protein synthesis factor, Tu, has placed the



Figure 1.11. Secondary endosymbiotic event where the apicomplexan ancestor engulfed a eukaryote that already harboured a plastid. (Source: adapted from Roos *et al*, 1999).

apicomplexan plastid with the green alga lineage. Unfortunately, these studies were flawed because of a common problem; the apicomplexan plastid genome is very A-T rich, thus skewing both the ribosomal RNA sequences and amino acid composition of proteins (Kohler *et al*, 1997; Wilson *et al*, 1996) which posed a serious problem for sequence analysis and weakened the interpretation of the data obtained.

The presence of a gene, *orf470*, on the *P. falciparum* plastid DNA was the key that led to the proposal of a red alga origin (Williamson *et al*, 1994; Wilson, 1993). The same gene is found on the plastid genome of red algae, where it is designated *ycf24* (Law *et al*, 2000; Williamson *et al*, 1994; Wilson, 1993). The support of a red algal origin of the apicoplast is stronger than that of a green algal lineage (McFadden *et al*, 1997; Zhang *et al*, 2000).

1.5.3. Protein targeting to the apicoplast

The apicoplast genome contains a total of 57 genes (Table 1.3), the majority of which are involved in gene expression. There are genes encoding three subunits of RNA polymerase, 17 ribosomal proteins, elongation factor Tu, duplicated large and small subunit rRNAs and 25 tRNAs. There are also nine open reading frames, one of which encodes a subunit similar to the Clp family of molecular chaperones, one is a conserved ORF found in red algal plastids and the rest are of unknown function. These genes are not sufficient to provide all the required elements for the maintenance and survival of the plastid. During the evolutionary development of plastids, many of the genes that encode proteins that are essential to plastid function have been transferred to the nucleus of the host. The same scenario is observed in plants where it is estimated

Class	Genes	
Ribosomal RNA	168, 238	
Transfer RNA	$ \begin{split} A^{UGC}, C \stackrel{GCA}{\to} D^{GUC}, E^{UUU}, F^{GAA}, G^{ACC}, G^{UCC}, H^{GUG}, I^{GAU}, \\ K^{UUU}, L^{UAG}, L^{UAA}, M^{CAU}, M^{CAU}, N^{GUU}, P^{UGG}, Q^{UUG}, R^{UCU}, \\ R^{ACG}, S^{GCU}, S^{UGA}, T^{UGU}, V^{UAC}, W^{CCA}, Y^{GUA} \end{split} $	
Ribosomal proteins	rps 2, 3, 4, 5, 7, 8, 11, 12, 17, 19	
	rpl 2, 4, 6, 14, 16, 23, 36	
RNA polymerase	$rpoB, C_1, C_2$	
Other proteins	clpC, tufA, ORF430	
Unassigned ORFs	51, 78,79, 91, 101, 105, 129	

Table 1.3. Gene content of the 35 kb circular DNA of P. falciparum. (Source: Wilson et al, 1996).

that between 1,000 to 5,000 chloroplast proteins are encoded by nuclear genes (Martin & Herrmann, 1998). These nuclear encoded proteins must be imported back into the

plastid. In order for import, plastid targeted proteins possess N-terminal extensions called transit peptides.

Due to the fact that the apicomplexan plastids are of secondary endosymbiotic origin, protein targeting is more complex because of the presence of multiple membrane layers for the proteins to cross. Therefore, a two-part targeting system is used for plastid entry. Instead of having just a transit peptide, these plastids possess bipartite presequences (Waller et al, 1998). These bipartite presequences of proteins that are targeted to the apicoplast have two functional domains:- a signal peptide and a transit peptide. The signal peptide consists of a short run of hydrophobic residues folowed by a cleavage motif (Nielsen et al, 1997). The transit peptide domain carries a net positive charge (Waller et al, 1998). Recently, the Plasmodium transit peptide has been defined (Foth et al. 2003) and it also contains a net positive charge due to the high number of lysine residues found within the domain. It is also depleted in acidic residues (glutamate and aspartate) among its first 20 amino acids. In addition, Plasmodium transit peptides are enriched in isoleucine while possessing low numbers of small apolar amino acids like glycine, valine, alanine and leucine. With this consensus in mind, Foth et al (2003) identified 466 putative apicoplast-targeted proteins in the P. falciparum genome.

The presence of a bipartite presequence means that the protein targeting process involves at least two steps. The apicoplast-targeted protein with the help of the signal peptide, first enters into the secretory pathway At some point in this pathway, the transit peptide mediates the transfer committing to, and allowing entry into, the apicoplast (Waller *et al*, 2000). As for the exact mechanism and proteins involved in

this transport process, they are unknown. In plant chloroplasts, protein targeting is mediated through a system called the Tic-Toc system. The translocon outer chloroplast complex (Toc) and the translocon inner chloroplast complex (Tic) are located on the outer and inner membrane of the chloroplast respectively (Soll & Tien, 1998), where they are responsible for translocating proteins across the chloroplast membrane into its interior. Instead of two membranes, the apicoplast is surrounded by at least three membranes. Therefore, the tic-toc system cannot operate precisely as in plants to ensure the translocation of apicoplast-targeted proteins. However, Tic and Toc homologues have been identified in *P. falciparum* (vanDooren *et al*, 2000) which suggests that the protein translocation system in *Plasmodium* is probably related to the tic-toc system known to operate in plant chloroplasts.



Figure 1.12. Diagrammatic representation of the chloroplast protein transport system (Tic-Toc). (Source: Fuks & Schnell *et al*, 1997)

1.5.4. Functions of the apicoplast

Various experiments have shown that the apicoplast is indispensable for the survival of the apicomplexan parasites (Fichera & Roos, 1997; McConkey *et al*, 1997; McFadden & Roos, 1999). As the apicomplexan plastid is non-photosynthetic, a question arises to its precise cellular function. Since there are estimated to be 466 putative nuclear-encoded plastid targeted proteins – the net plastid proteome could point to its function. Several laboratories have successfully demonstrated the presence of key metabolic activity taking place in the apicoplast. Key pathways include *de novo* fatty acid biosynthesis (Jelenska *et al*, 2001; Waller *et al*, 1998; 2000), isoprenoid biosynthesis (Jomaa *et al*, 1999; Wiesner *et al*, 2000) and haem synthesis (Sato & Wilson, 2002; van Dooren *et al*, 2002). The principal features of these pathways are outlined below.

1.5.4.1. Fatty acid synthesis

Fatty acid synthesis is a process where two-carbon precursors are assembled into fatty acids. Two important cofactors involved in the process are coenzyme A (CoA) and the acyl carrier protein (ACP). Even though it was previously thought that *P. falciparum* does not synthesise fatty acids *de novo* (Matesanz *et al*, 1999), several proteins that are involved in the process including ACP, β -ketoacyl-ACP synthase III (FabH) and acetyl coA carboxylase (ACC) have been shown to localise to the apicoplast (Jelenska *et al*, 2001; Waller *et al*, 1998; 2000). In addition, using radilabelled precursors, Surolia & Surolia (2001) have shown that they are incorporated into fatty acids by *P. falciparum*. The evidence therefore suggests that *P. falciparum* does possess a *de novo* fatty acid synthesis pathway. Judging by the presence of apicoplast-targeted enzymes responsible for fatty acid synthesis, the

apicoplast is likely to be site for fatty acid synthesis. These fatty acids can then be incorporated into phosphoglycerides, sphingolipids and glycosylphosphatidylinositols (GPI).

1.5.4.2. Isoprenoid synthesis

Isoprenoids are a group of natural products that include sterols, carotenoids and terpenoids. They are usually incorporated into proteins that are involved in signal transduction, protein-protein interactions and membrane-associated protein trafficking. A 1-deoxy-D-xylulose-5-phosphate (DOXP) pathway or non-mevalonate isoprenoid biosynthesis pathway in *P. falciparum* linked the apicoplast with the plastids in plants and algae which also possess the DOXP pathway. The DOXP pathway provides a pathway to isoprenoids that is distinct from the pathway in eukaryotes. In *P. falciparum*, the genes encoding DOXP synthase and DOXP reductoisomerase have been found (Jomaa *et al*, 1999) and they contain plastid-targeting domains.

1.5.4.3. Haem synthesis

Haem is the prosthetic group of oxygen-carrying electron-transferring proteins (for example cytochromes). In *P. falciparum*, protein synthesis appears to be haem-dependent (Surolia & Padmanaban, 1991; 1992). *Plasmodium* also appears to encode its own enzyme δ -aminolevulinate dehydratase (ALAD) which is involved in the condensation of two molecules of δ -aminolevulinate to give porphobilinogen (Sato *et al*, 2000; Sato & Wilson, 2002). The gene encoding ALAD possesses a putative apicoplast-targeting leader sequence (Sato & Wilson, 2002), thus it appears that it might be active within the apicoplast. However, there is no evidence for the presence of another important protein, δ -aminolevulinate synthetase (ALAS) which is responsible

for the formation of δ -aminolevulinate from other precursors. Therefore, there is still a need for further clarification to the precise role that the apicoplast plays in haem synthesis.

1.5.5. The apicoplast as a potential drug target

The discovery of the presence of a relict endosymbiont within the apicomplexan parasites has opened up doors for novel chemotherapies exploiting the apicoplast as a drug target. Even though the true function of the apicoplast has yet to be established, it has been shown to be indispensable to the parasite (Fichera & Roos, 1997). Interestingly though, when *Toxoplasma gondii* parasites lose apicoplast function, they do not perish immediately, instead they fail to invade new host cells (He *et al*, 2001). This suggests that the apicoplast may play a role in the invasion of the host cell. In addition, due its prokaryotic origin, the apicoplast contains key cellular processes such as DNA replication, transcription, translation, catabolism and metabolism differ substantially from similar processes in their host. Therefore, the apicoplast can make a good drug target. In the next section, several pathways that have already been targeted by compounds that lead to parasite death are discussed.

1.5.5.1. Fatty acid synthesis

As reviewed in Section 1.5.4.1, the presence of a fatty acid synthetic pathway in *Plasmodium* has been noted. The plasmodial pathway appears to be similar to the type II fatty acid pathway present in bacteria and chloroplasts which uses distinctive enzyme systems when compared with eukaryotic pathways. Several antibiotics target these type II fatty acid biosynthetic enzymes. For example β -ketoacyl ACP synthase (FabH) is a target for thiolactomycin (Waller *et al*, 1998) and triclosan

targets enoyl-ACP reductase (FabI) (Surolia & Surolia, 2001). Both of these antibiotics show anti-plasmodial activity.

1.5.5.2. Isoprenoid synthesis

The DOXP reductoisomerase has been identified in *P. falciparum* and has been shown to be a target of fosmidomycin (Jomaa *et al*, 1999). Furthermore, fosmidomycin can inhibit the growth of *P. falciparum* in culture (Jomaa *et al*, 1999). Many other inhibitors of the DOXP pathway are being tested against bacterial (Altincicek *et al*, 2000) and plant (Zeidler *et al*, 2000) systems which may be of use to test for antimalarial activity.

1.5.5.3. DNA replication and transcription

Since the apicoplast is indispensable for the apicomplexan parasite, inhibitors of DNA replication of its genome will affect the well-being of the parasite itself. Ciprofloxacin blocks prokaryotic DNA replication by inhibiting DNA gyrase, a prokaryotic type II topoisomerase. This antibiotic also inhibits plastid DNA replication in *P. falciparum* (Weissig *et al*, 1997). Interestingly in *T. gondii*, the antibiotic does not kill the parasite immediately but progressively reduces the copy number of the plastid genome, and it affects the ability of the parasite to invade host cells (Fichera & Roos, 1997). As yet, other than gyrA and gyrB (two components of the DNA topoisomerase system), no other plastid DNA replication related genes have been described. However, enzymes involved in plastid DNA replication and maintenance should be good targets for antimalarials.

Plastid transcription uses prokaryotic-like RNA polymerase homologues (Gray & Lang, 1998). Antibiotics that inhibit transcription can be used against these parasites. For example, rifampicin inhibits the function of prokaryotic type RNA polymerase B, is effective against apicoplast transcription (Wilson *et al*, 1996) and has activity against *Plasmodium*.

Aims of the project

The discovery of a gene that encodes a putative DNA polymerase, DNA helicase and DNA primase activity, and which possesses a putative apicoplast-targeting presequence, underlies the aims of this project. The main aim of the project was to determine and clarify the true status of this gene, *PfPREX*. My PhD project aimed to clone and express the gene and to determine whether it encodes all of the predicted functions. Moreover, the project aimed to investigate the origins of *PfPREX*, and the evolutionary route that has led to it now functioning in *P. falciparum*.

١ť.

Chapter II

÷

45

Materials and Methods

2.1. Materials

All chemicals used were reagent grade and were purchased from Sigma, unless otherwise stated. DNA restriction and modification enzymes were from Promega. Ribo- and deoxoyribonucleoside triphosphates were purchased from both Invitrogen and Sigma. Oligonucleotides were synthesised by Invitrogen and MWG Biotech. All radioactive materials were purchased from Perkin Elmer. Poly (dT) from Sigma, had an average length of 200-300 nucleotides. Bacteriophage M13mp18 singlestranded DNA was obtained from New England Biolabs. The recombinant helicase, PerA was purchased from Cambio Science. Klenow fragment, *Taq* and *Pfu* DNA polymerases were purchased from Promega. The Gateway Cloning kits and competent cells, DH5 α and BL21si were obtained from Invitrogen.

2.2. Parasite lines/clones and culture

Plasmodium falciparum clone 3D7 was cultured *in vitro* in a class I sterile hood using the method of Jensen & Trager (1978) with some modifications. The parasites were grown at 5% haematocrit in RPMI 1640 medium (Gibco BRL), supplemented with 5.94 g/l HEPES, 42 ml/l 5% (w/v) sodium bicarbonate and 10% pooled heat-inactivated human serum (group AB). Whole blood was obtained from the Blood Transfusion Service. The cultures were gassed with a mixture of 3% CO₂, 1% O₂, 96% N₂ and incubated at 37°C. The culture was kept at between 1 – 5% parasitaemia. If the parasitaemia exceeded 5%, the culture was diluted to a lower parasitaemia level in fresh supplemented RPMI 1640 medium.
2.3. Bacterial strains and cultures

Several *E. coli* strains were used for cloning, sub-cloning and expression studies. *E. coli* JM109 (Promega) and DH5 α (Invitrogen) were used for standard subcloning experiments. For the expression studies, the *E. coli* BL21si (Invitrogen) strain was used. The liquid cultures of *E. coli* were grown in Luria-Bertani (LB)(Appendix A.1) broth at 37°C with shaking while for BL21si strains were grown in LB broth with no NaCl.

Antibiotics were used at final concentrations of 30 μ g/ml for kanamycin (km) and 50 μ g/ml for ampicillin (amp) depending on particular experiments. Stock solutions of the antibiotics were made at 100 mg/ml with water for ampicillin and 30 mg/ml for kanamycin and were kept at -20°C.

2.4. Basic molecular biology techniques

2.4.1. Preparation of parasite extracts

For the preparation of extracts from the malarial parasites, 1 ml of parasite culture (5% haematocrit and 5% parasitaemia) was used. The infected red blood cells were pelleted by spinning at 800 g for 5 minutes. The pelleted red blood cells were lysed by the addition of 0.15% saponin in PBS. The suspension was incubated at room temperature for 15 minutes, followed by centrifugation at 800 g for 10 minutes. The pellet was further washed twice with 1 ml of PBS (Appendix A.2). The parasite pellets were stored at -20° C until use in Western blot experiments.

2.4.2. Preparation of genomic DNA from P. falciparum

Genomic DNA from *P. falciparum* 3D7 was isolated from 10^8 mixed asexual erythrocytic stage parasites using the Bio-Rad InstaGeneTM whole blood kit. Asynchronous cultures of 3D7 were spun down to obtain the red blood cell pellet and genomic DNA extraction was carried out according to the manufacturer's instructions.

The kit consisted of the lysis buffer and a suspension of InstaGeneTM matrix. The cell membrane was broken up by the addition of the lysis buffer and the haemoglobin was released from the red blood cells along with the nuclei from the white blood cells. The white blood cell nuclei were pelleted by centrifugation and the InstaGeneTM matrix was added. The matrix is a polyvalent cation-chelating resin which binds cations that would catalyse the degradation of DNA at high temperature. The parasite nuclei were lysed by incubation at 60°C. After centrifugation, the DNA released from the nuclei localised to the supernatant while the debris was in the matrix pellet. The DNA obtained does not need further purification and can be used straight for PCR and other applications.

2.4.3. Preparation of total RNA from *P. falciparum*

Total parasite RNA was prepared using the TRIzol[®] Reagent (Invitrogen). The TRizol[®] Reagent is a mono-phasic solution of phenol and guanidine isothiocyanate and the RNA extraction using this reagent was carried out based on the method developed by Chomczynski & Sacchi (1987). The rbc pellet was lysed by the addition of 1 ml of the reagent followed by repetitive pipetting. The rest of the extraction was carried out according to the manufacturer's instructions. The RNA obtained was dissolved in DEPC-treated water and stored at -70° C until use.

2.4.4. Preparation of plasmid DNA from E. coli

Plasmid DNA was prepared from 5 ml of overnight bacterial cultures by alkaline lysis. Overnight cultures were harvested at 3,000 g for 15 minutes. Plasmids were extracted using QIAprep spin Mini Prep kit from Qiagen according to the manufacturer's instructions. The plasmid DNA was resuspended in sterile water and stored at -20°C until use.

2.4.5. Restriction digestion of plasmid DNA

Plasmid DNA was digested with various restriction enzymes (Promega). Approximately $0.2 - 2 \mu g$ of plasmid DNA was cut using 10 units of enzyme in 1X of the corresponding reaction buffer and left at 37°C for 1 - 2 hours. The digestions were stopped by inactivating the enzyme at 68°C for 10 minutes when using heat-sensitive restriction endonucleases.

2.4.6. Agarose gel electrophoresis of DNA

The DNA samples were separated on 0.8 - 1.0 % agarose (Gibco BRL) prepared in 1X TAE buffer (Appendix A). The gel was cast on a standard gel apparatus (Horizon [@] Gibco BRL) and DNA samples containing 1X loading buffer (Appendix A.3) were loaded into the wells of the gel submerged in 1X TAE. 500 ng of 1 kb DNA ladder (Promega) was loaded alongside the samples on the same gel as a DNA size marker. Electrophoresis was carried out at 10 V/cm. The gel was then stained with EtBr (0.5 µg/ml) for at least 15 minutes and viewed under the short range UV transilluminator (Hoefer) and photographed.

2.4.7. Polymerase chain reaction (PCR)

The PCR reactions were carried out in 100 μ l volumes containing 200 μ M of each dNTP, various concentrations of MgCl₂ ranging from 1.5 mM to 3.0 mM, 10 μ l of 10X *Pfu* polymerase buffer (Promega), 0 to 10 % DMSO, 1U of *Pfu* polymerase (Promega) and 50 ng of *P. falciparum* 3D7 genomic DNA. The typical amplification profile for the PCR for 25 or 40 rounds was as follows: - denaturation step at 94°C for 30 seconds; annealing step at the appropriate annealing temperature for 30 seconds, followed by an extension step at 62°C for 2 – 10 minutes depending on the end product to be obtained and a final elongation step at 72°C for 5 minutes. The PCR amplifications were performed on the DNA engine Gradient Cycler (MJ Research). 10 μ l of the amplified products were analysed on the appropriate percentage agarose gel.

Primer Name	Gene region	Gene-specific primer sequence		
and have a state of				
GWpolF1	Polymerase	5'-ATCAAGAAATGGAAAAGAATAAC-3'		
GWpolR1	Polymerase	5'-ATCATTCCATGGGAACATTT-3'		
GWHeliSignalF1	Helicase/Primase	5'-TTCTATGATATACTACCATGCTTTTG-3'		
GWHeliHingeR1	Helicase/Primase	5'-TGTTTTGATTCATATTCATCATCAGCTA-3'		
GWHeliF4	Helicase	5'-ATGCGAATGATTGTTTGAAACAT-3'		
OWE PD4	TT 1			
GWHCIIK4	Helicase	5'-ICICIAAACIAICCACA4A-3'		
GWPriF1	Primase	5'-GTCATAGATGTGGATATAAAGGA-3'		
(1111)				
GWPriR1	Primase	5-TGGCATUAATTICTCCTTCTGT-3		
GWHingeF1	Hinge	5'-ATGAGCATGTAGTTTCAAATACT-3'		
	_			
GWHingeR2	Ilinge	5'-TACTACTGTTCTTATTCTTATT-3'		
AttB1		5'-GGGGACAAGTTTGTACAAAAAGCAGGCT-3'		
		· · · · · · · · · · · · · · · · · · ·		
AttB2		5'-GGGGACCACTTTGTACAAGAAAGCTGGGT-3'		
Add attB1 sequence to forward primers and attB2 to reverse primers				
ssurRNAF1	ssurRNA	5'-GATCCTAGCTCAGAATTAAACGCT-3'		
ssurRNAR1	ssurRNA	5'-GTAAGGTTTATCGTGTTGCATC-3'		

Table 2.1. Primers used for all PCR reactions are listed in this table.

2.4.8. DNA extraction from low-melting point agarose gels

The DNA products were analysed first on a 1% low-melting agarose gel. Gel fragments of the desired size were cut from the gel and incubated with equal volume of TE buffer (0.1M) (Appendix A.2) at 37°C for 30 minutes after disrupting the gel piece with a wide-bore needle. One unit of Agarace (Promega) was added and the suspension was incubated for a further 45 minutes at 37°C. An equal volume of phenol was added and the suspension incubated on ice for 20 minutes. This was followed by centrifugation at 13,000 x g at 4°C for 30 minutes. The upper aqueous layer was removed and the DNA was recovered by standard ethanol precipitation (Sambrook *et al*, 1989)

2.4.9. Reverse-transcriptase PCR amplification (RT-PCR)

2.4.9.1 First strand cDNA synthesis

The RNA used was treated with 2 μ l (1 U/ μ l)(Sigma) of DNase I at room temperature for 15 minutes. 2 μ l of RNase inhibitor (10 U/ μ l) (Gibco Life Technologies) and 2 μ l of 200 ng/ μ l oligo dTs were added to 20 μ l of DNase-treated RNA. The RNA was denatured by incubating the mix at 80°C for 10 minutes and then cooled rapidly on ice. After the addition of 4 μ l of 0.1 M DTT, 8 μ l 5X FSB and 2 μ l of 10 mM dNTPs, 20 μ l of the mixed solution was removed as a negative control (minus reverse transcriptase). To the rest of the mixture, 1 μ l of superscript II (Gibco) was added to make the first strand of cDNA. The primer extension was carried out at 42°C for 1 hour.

Primer	Gene	Primer sequence	Product	Tm
Name	region		size (bp)	(°C)
RTHeliF1	Primase	5'-GCTCGTACATTTGTCGTTATGC	w	59
RTHeliF2	Helicase	5'-AAGAAGCTGAAAACGCAAGG-3'	453	59
RTHeliR2		5'-TGGCATCAATTTCTCCTTCTG-3'		60
RTHingeF1	Hinge	5'-TCGGCACAAAATAACATTCC-3'	420	58
RTHingeR1		5'-TCCATCACCGATGTTAGATCC-3'		59
RTPolF1	Polyme-	5'-CAACAGGTTTAGAAGTGTTCGATG-3'	400	60
RTPolR1	rase	5'-CAACTAGAATCACGAGCAGCA-3'		59

Table 2.2. Primer sequences used for the RT-PCR reactions.

2.4.9.2. Reverse transcriptase-PCR

The cDNA was amplified using various oligonucleotides, for 40 cycles: $94^{\circ}C$ -30 seconds, $50^{\circ}C$ -30 seconds, $62^{\circ}C$ -2-5 minutes. The amplification products were analysed on an agarose gel using ethidium bromide according to standard procedures.

2.4.10. Formaldehyde gel electrophoresis of RNA

A formaldehyde gel was made by adding a solution containing 16 ml of formaldehyde, 5 ml of 20X Northern gel buffer (Appendix A.4) and 29 ml of DEPC-treated ddH₂O to a solution of 50 ml of 2% agarose melted in a microwave and allowed to set in a standard gel apparatus (Horizon [®] Gibco BRL). A 1X Northern Gel buffer was prepared by mixing 50 ml 20X Northern Gel buffer, 88 ml formaldehyde and 860 ml DEPC-treated ddH₂O.

The RNA samples were prepared in a final volume of 20 μ l by adding to the RNA sample, 1 μ l of 20X Northern gel buffer, 3.5 μ l formaldehyde, 10 μ l of formamide and DEPC-treated ddH₂O to 20 μ l. The mixture was heated at 55°C for 15 minutes, then chilled on ice and 2 μ l of RNAse-free loading buffer (Appendix A.4) was added before loading the samples on the gel. The gel was run at 60V for 4 to 5 hours with a magnetic stirrer placed inside the box once the RNA samples had entered the gel matrix. The stirrer keeps the buffer circulated.

2.4.11. Northern Hybridisation

RNA was extracted and approximately 10 μ g was electrophoresed using denaturing conditions on a formaldehyde gel. The gel was then directly blotted on to a nylon membrane (Hybond N, Amersham) according to the technique described in Sambrook *et al* (1989), using 20x SSC.

2.4.11.1. Labelling of DNA probes

DNA probes were labelled using the Prime-It random primer labelling kit (Stratagene). The probes were labelled according to the manufacturer's instructions. The procedure relies on the use of random hexanucleotides that anneal to multiple sites along the DNA template. 25 μ g of purified DNA templates were boiled for 5 minutes in the presence of the random oligonucleotides in a final volume of 34 μ l. The mix was then allowed to cool and 10 μ l of 5x primer buffer, 5 μ l of [α^{32} P] ATP (3000 Ci/mmol, 10 mCi/ml, NEN Life Science Products), and 1 μ l of Klenow fragment (5U/ml) was added. The reaction was incubated to 37°C for 10 minutes and stopped by adding 2 μ l of stop mix. The probe was then purified to separate the unincorporated dATP from the

labelled probe using a microspin column (Sigma). The labelled DNA was then denatured by boiling for 5 minutes.

2.4.12. Chemical transformation of *E. coli*

The cells used for chemical transformation were either DH5 α (Promega) or BL21si (Invitrogen) and a standard heat shock method was employed (Sambrook *et al*, 1989). 50 µl of cells were used. The cells were pre-incubated with about 1 – 50 ng of DNA to be transformed on ice for 30 minutes. This was to allow the DNA to attach itself to the cells. This was followed by heat shocking the cells at 42°C for 45 seconds. The cells were placed immediately on ice for 2 minutes for the cells to recover before the addition of 450 µl SOC medium (Appendix A.1). The cells were then allowed to multiply at 37°C for 1 hour. 100 µl of the cell suspension was plated onto LB plates containing the appropriate antibiotics and incubated at 37°C overnight.

2.5. Molecular cloning techniques for PfPREX

2.5.1. Gateway cloning technology

Gateway cloning technology (Invitrogen) is a powerful system. It provides a faster and more efficient route to traditional cloning methods. It is based on the lambda phage site-specific recombination. This system allows the transfer of DNA fragments between vectors with compatible recombination sites while maintaining orientation. Therefore, it is a versatile system which allows the simple relocation of a gene of interest into any number of different expression vectors for expression in a variety of hosts. The Gateway Cloning Strategy involves two steps, the first of which is to insert the gene of interest into an entry vector. The second step involves the shuffling of the gene fragment into a destination vector to generate an expression clone.



Figure 2.1. Schematic representation of the Gateway cloning technology. The diagram shows the two key reactions for Gateway cloning, the BP reaction and the LR reaction. (Source: Invitrogen)

With our intended genes of interest, we anticipated problems with expression due to the high A-T content of the *Plasmodium* genome. Therefore, this system was selected because it is ideal for us to attempt to clone our gene into different expression vectors with the minimum number of passages through *E. coli* and also permits one to express the recombinant proteins in various expression systems with minimal sub-cloning.

2.5.2. Generation of expression constructs

2.5.2.1. Generation of entry clones

The generation of an entry clone is achieved by designing primers that contain *att*B recombination sites. From the PCR, one obtains PCR fragments containing

the *att*B sites which are then converted into entry clones by recombining into the *att*P sites of pDONR201.

2.5.2.1.1. Preparation of *att*-PCR products

The PCR primers used for the amplification were obtained from Invitrogen. The forward and reverse primers contain additional sequences of attB recombination sites at the 5' ends which are $attB_1$ and $attB_2$ respectively. The attB sites are specific recombination sites that are normally present in *E.coli* which will recombine with attPsites found on the lambda phage for integration process to proceed. The PCR reactions were carried out as stated in Section 2.4.7.

2.5.2.1.2. 'BP' reaction for the generation of the entry clones

The BP reaction involves the generation of the entry clones, it starts with the mixing of the specially generated PCR products containing the gene of interest and the entry vector of your choice. More specifically, the BP reaction facilitates the recombination of the *att*B-containing substrate (*att*<u>B</u>-PCR product) with an *att*<u>P</u> substrate (pDONR201- Donor vector), which gives the reaction the name BP. This reaction is catalysed by the BP ClonaseTM enzymes mix.

The *attP* containing entry vector, pDONR201 (300 ng) was mixed with 2 μ l of the purified PCR product. It was added to a reaction mixture of 20 μ l, containing 4 μ l of the BP clonase (mixture containing both the lambda phage integrase and the *E. coli* protein, IHF). After incubation at 25°C for 1 hour, proteinase K (4 μ g in 2 μ l) was added to inactivate the BP clonase by incubating the mixture at 37°C for a further 20 minutes. 2 μ l aliquots were transformed into *E. coli* DH5 α (library efficient strain)

(Invitrogen) and plated onto Kanamycin (100 μ g/ml) containing LB plates. Since the Gateway reaction has a high efficiency of generation of positive clones (usually > 99%), it minimises the need for screening.

2.5.2.2. Generation of destination clones

2.5.2.2.1. 'LR' reaction for the generation of destination clones

The second step of Gateway Cloning involves the LR reaction. This reaction facilitates the recombination of the attL containing entry clone with a attR containing destination vector. This reaction is catalysed by the LR ClonaseTM enzyme mix.

For the LR reaction, the entry clones obtained from the BP reaction were used. The destination vectors used were pDEST15 (GST-Tag) and pDEST17 (His-Tag). Approximately 200 ng of entry clone was added to 300 ng of destination vector in 20 µl reactions containing 4 µl of LR ClonaseTM (which contains lambda recombination proteins Int, Xis and an *E. coli*-encoded protein IHF). The reaction mixture was incubated at 25°C for 60 minutes. Proteinase K was added and the reaction was further incubated at 37°C for 20 minutes. 2 µl aliquots of each reaction were transformed into *E. coli* (DH5 α) and plated onto LB-Amp plates (100 µg/ml). Colonies were picked and minipreps to extract the plasmids were performed to identify positive clones.

2.6. Protein analysis

2.6.1. Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE)

The purified proteins or parasite extracts were prepared by the addition of an equal volume of 2X sample buffer (Appendix A.5). The samples were boiled for five

minutes before being loaded onto a gel system (Laemmli, 1970) consisting of a 10% resolving gel and a 5% stacking gl (Appendix A.5). The gel was run at 150V on a mini-protean dual slab gel electrophoresis apparatus (Bio-Rad) in 1X running buffer (Appendix A.5).

After electrophoresis, the gels were immersed in a Coomassie Blue staining solution (Appendix A.5) for 1 hour at room temperature with shaking. Excess dye is then allowed to diffuse from the gels by immersing the gel in the destain solution (Appendix A.5) for one hour.

2.6.2. Western blotting

The proteins were size-fractionated by electrophoresis on SDS-PAGE gels and transferred to Hybond nitrocellulose membranes (ECL) using the Bio-Rad minitrans blot electrophoretic transfer cell at 100 V for 1 hour. The membrane was blocked with 5% skimmed milk containing 0.2% Tween 20 in 1X TBS (Appendix A.6). It was probed with a dilution of the primary antibody at 4°C overnight with shaking in 1% skimmed milk with 0.1% Tween 20 in TBS. The membrane was then washed 4 times with 1X TBS, 1% skimmed milk. After washing the membrane, it was probed with horseradish peroxidase-labelled secondary antibody (Diagnostics Scotland), diluted to 1:2000 in 10X TBS and 1% skimmed milk. The membrane was incubated with the secondary antibody at room temperature for 2 hours with shaking, followed by 3 washes for 30 minutes. The immunoreactive proteins were detected using the enhanced chemiluminscence protocol (Pierce). Blots were exposed to film for between 30 seconds to 15 minutes depending on the signal intensity.

2.6.3. Protein quantification

The protein concentrations were determined using the Bio-Rad protein assay which is based on the method of Bradford (Bradford, 1976). It involves the addition of an acidic dye to a protein solution where the colour change is determined by measurement at 595 nm with a spectrophotometer. Protein standards using bovine serum albumin (BSA) were prepared within the range of 0.5 to 5 μ g of proteins. The dye reagent (Bio-Rad) was diluted 1:4 with distilled water and filtered to remove particulate matter. 10 ml of each standard and sample solution were pipetted into separate microtiter plate wells. 200 μ l of the diluted dye reagent was added to each well and mixed thoroughly. The plate was then incubated at room temperature for at least 5 minutes when the absorbance was read at 595 nm using the microplate reader (Titertek Multiskan[®] MCC/340). The protein concentrations were calculated with reference to the standard curve drawn with the values obtained from the protein standards prepared.

2.7. Expression and purification of recombinant proteins

2.7.1 Expression and purification of His-tagged recombinant proteins

The polymerase domain (PfPREXpol) was first cloned into an entry vector, pDONR201 then transferred into the Gateway vector pDEST17 (Invitrogen) to give a hexahistidine amino-terminal fusion protein. This was to allow the ease of purification of the recombinant protein. The pDEST17 vector has salt-inducible T_7 promoter, therefore expression of the fusion protein was carried out in *Escherichia coli* strain BL21si (Invitrogen) which enables differential expression in response to salt.

Positive colonies were inoculated onto LB without salt with ampicillin (100 μ g/ml). From an overnight culture of 5 ml of LBamp without NaCl, 1 in 100 dilution of

the culture was transferred to fresh LBamp without NaCl, containing the respective antibiotics and incubated at 37°C until OD₅₉₀ was approximately 0.3. Expression of the recombinant protein was induced by the addition of 5 M NaCl to a final concentration of 0.3 M and incubated at 37°C for a further 4 hours. The culture was then harvested and resuspended in the lysis buffer (1 mM EDTA, 50 mM Tris-HCl, pH 8.0, 1 mM DTT). The cells were lysed by sonication (20 microns, 30s on, 30s off for 6 cycles). Lysates were cleared by centrifugation at 13,000 x g for 20 minutes at 4°C. The cell lysates were filtered using a 0.2 µm syringe filter (Sartorius). The recombinant protein was then purified using a Nickel-chelate column through the BioCad purification system, which allowed high-pressure affinity chromatography to be carried out. A wash was carried out at the beginning with 50 mM of imidazole to remove non-specific binding of E. coli proteins from the column. Elution of the His-tag proteins was performed at an imidazole gradient ranging from 50 mM to 500 mM with 1.5 ml fractions over 10 fractions. Fractions containing the eluted recombinant protein were collected and pooled together. The pooled fractions were concentrated using a stirred cell concentrator (Amicon) and a cellulose membrane of 30 kDa cut off size (Flowgen). Concentration of at least 15 fold was achieved. The purified proteins were then checked for purity by running on a protein SDS-PAGE gel.

2.7.2 Expression and purification of GST-tagged recombinant proteins

The helicase/primase domain was also cloned into pDONR201 to make the entry clones and then transferred to the pDEST15 vector to give a N-terminal GST-Tagged fusion protein. Like the pDEST17 vector, the pDEST15 vector has a salt-inducible T7 promoter for which the induction of expression of the fusion protein was carried out in the *E. coli* BL21si strain (Invitrogen).

The positive colonies were plated on LB plates without salt with ampicillin (100 μ g/ml) and the procedure for the induction of expression of the recombinant protein was essentially the same for the His-Tagged recombinant protein.

The purification of the GST-Tagged recombinant proteins was performed using the GST-purification kit from Novagen. The GST purification resin slurry (50%) was provided with the kit. A 2 ml bed volume column was made and the 20% ethanol was allowed to drain from the column. The column was allowed to equilibrate in 5 volumes of 1X GST wash buffer provided in the kit. Subsequent washing steps using 10 volumes of 1X GST wash buffer was carried out. For the elution of the GST-Tagged proteins, the 1X elution buffer containing reduced glutathione (10 mM) was used. Three eluted fractions of 2 ml volume were collected and the fractions were concentrated using a stirred cell concentrator (Amicon) and a cellulose membrane of 10 kDa cut off size (Flowgen). Concentration of at least 5 fold was achieved. The purified protein quality was checked by SDS-PAGE.

2.7.3. Production of antibodies

For the production of anti-serum against the polymerase domain, the recombinant polymerase was sent to 'Diagnostics Scotland'. The polyclonal antiserum was raised in rabbit by the inoculation of 4 aliquots of the recombinant polymerase (100 μ g per aliquot of 1 ml).

For the production of the antisera against the hinge, primase and helicase domains, new primers were designed for those domains to clone them into Gateway pDEST15 vector to generate the new recombinant proteins. The GST-Tagged fusion proteins were purified as stated in the previous section and four aliquots of proteins were sent to Diagnostics Scotland to have polyclonal antisera raised in rabbits.

All the antisera (with the exception of the GST-tagged helicase domain antisera due to time constraint) were all checked using dot blotting method against the corresponding recombinant proteins to ensure the presence of antibodies in the antisera specific for the proteins in question. In addition, negative controls were carried out to ensure that first there was no non-specific cross-reaction between the antisera produced and non-related proteins and that the pre-immune sera also did not produce any reaction to the corresponding recombinant proteins.

2.8. Functional analysis of the recombinant proteins

2.8.1. DNA primase activity assays

Reaction buffer for the DNA primase assay contained 50 mM NaCl, 2 mM MgCl₂, 200 μ M ATP, 0.2 mg/ml BSA, 50 mM Tris (pH 7.5). An appropriate amount (1 μ g) of primase was added to the reaction buffer alongside 500 ng of poly (dT) and incubated at 30°C for 15 minutes. Subsequently, buffer exchange was performed to remove the salts from the previous reaction buffer in buffer-exchange columns (Amersham). The flow-through from the buffer exchange columns was then used for the next step of the assay whereby 1 U of Klenow fragment (Promega) and 10 μ Ci of [³H]-dATP (15 Ci/mmol, Perkin Elmer), along with the buffer provided with the Klenow fragment were added. The reaction mixture was incubated at 30°C for 30 minutes before it was terminated by the addition of 10 μ l of 0.5 M EDTA. The unincorporated radionucleotides were removed using either microspin G-50 columns (Amersham Biosciences) or the Sigma Spin Post-reaction spin columns, spun at 750 x g

for 2 minutes. The flow-through was then counted for 1 minute in a beta-scintillation counter (Perkin Elmer Wallace Trilus 1450 Microbeta). Negative controls were carried out using the recombinant PfPREXpol. This was carried to ensure that the activity observed was due to DNA primase alone and not some *E. coli* contaminating protein that may be present in the purified recombinant protein.

2.8.2. DNA helicase activity assays

DNA helicase activity was characterised using the Helicase [³H] scintillation proximity enzyme assay (SPA) system (Amersham Life Science). The SPA system uses a weak beta-emitter like [³H]. On contact with the SPA beads, the deceleration of its particles will emit a light signal which can be measured. The starting material of the assay is a M13 single-stranded DNA bound to a [³H]-oligonucleotide. If a helicase is present, the oligonucleotide will be unwound from the M13 single-stranded DNA. The [³H] -oligonucleotide is then free to bind to a biotinylated complementary oligonucleotide. At this point streptavidin-SPA beads are added to the reaction buffer which will bind to the biotinylated/[³H]-oligonucleotide (Figure 2.2). This association of all the appropriate molecules will lead to a production of a signal which can be measured by a scintillation counter.

The assay is set up using 20 μ l of the assay buffer containing 30 mM Tris-HCl, pH 7.6, 5 mM ATP (or various other nucleotides), 5 mM MgCl₂, 0.075% (v/v) Triton X-100, 0.05% (w/v) sodium azide to which 10 μ l of the diluted (1:10) substrate solution which contains 10 mM Tris-HCl, pH 8.0, 25 mM NaCl, 1 mM EDTA and 0.05% (w/v) sodium azide and 0.08 μ Ci of [³H]-dNMPs was added. 10 μ l of the purified recombinant helicase (approximately 1 μ g) was added to the reaction mixture. The reaction was started by incubating the tubes at 33°C for 1 hour. The reaction was terminated by the addition of 10 μ l of the stop/capture reagent and incubated for a further 15 minutes at 33°C. To the mixtures, 200 μ l of streptavidin SPA bead suspension was added and incubated at room temperature for 15 minutes. The tubes were then counted for 1 minute in a beta-scintillation counter (Perkin Elmer Wallace Trilus 1450 Microbeta). All reactions were carried out in triplicate.



Figure 2.2. Helicase [³H] Scintillation proximity enzyme(SPA) assay (Amersham). (Adapted from Amersham Bioscience).

2.8.3. DNA polymerase activity assays

Reaction mixtures contained 20 mM Tris-HCl, pH 8.0, 1 mM DTT, 5 mM MgCl₂, 20 μ g BSA, 100 μ M each of dCTP, dTTP, dGTP and 10 μ Ci [³H]-dATP (15 Ci/mmol, Perkin Elmer) (method adapted from Huang & Levin, 2001). The template

used for the assay was M13mp18 single-stranded DNA (1.2 pmol), annealed to 1.6 pmol of M13 specific primer (5'-TTCCCAGTCACGACGTTGTAAAACGACGG-3'). A volume of 1.2 μ l of the annealed primer-template mixture was added to the reaction mixture along with 5 μ g of the recombinant enzyme. The mixture was incubated for 1 hour at 37°C. The unincorporated radionucleotides were removed using either microspin G-50 columns (Amersham Biosciences) or the Sigma Spin Post-reaction spin columns (Sigma), spun at 750 x g for 2 minutes. The radioactivity that passed through the column representing incorporated nucleotides, was then counted in a liquid scintillation counter (LKB Wallace 1219 Rack Beta). The relative activity in any given experiment is defined as a percentage relative to the highest activity recorded in a given experiment.

The DNA polymerase activity was also evaluated in the presence of various concentrations of MgCl₂, KCl and various pHs and temperature variations are recorded the relevant sections. The effect of aphidicolin was also examined by the addition of various concentrations of the inhibitor. All reactions were carried out in triplicate.

2.8.4. 3'-5' exonuclease activity assays

The substrate for the 3'-5' assay was prepared similarly to the reaction mixtures for the DNA polymerase activity assays (method adapted from Huang & Levin, 2001). Reaction mixtures contained 50 mM Tris-HCl, pH 8.0, 1 mM DTT, 5 mM MgCl₂, 20 µg BSA, 100 µM of each of dCTP, dTTP, dGTP, 50 µCi α -³²P-dATP and 5 pmol of M13-ssDNA to a final volume of 50 µl. Five units of Klenow fragment were added to catalyse the reaction and the mixture was incubated at 37°C for 1 hour. The unincorporated radionucleotides were removed using microspin G-50 columns. This labelled, double-stranded DNA substrate was then used for the exonuclease assay.

For the assay, the labelled DNA substrate was added to 50 mM Tris-HCl, pH 8.0, 1 mM DTT, 20 μ g BSA and 5 μ g of purified recombinant polymerase. The reaction mixture (25 μ l) was incubated at 37°C for 1 hour in the absence of deoxyribonucleotide triphosphates. The reaction was stopped by the addition of 5 μ l of 0.5 M EDTA, pH 8.0. DNA was precipitated by the addition of 10 μ l of 100% TCA and by centrifugation at 13,000 x g for 30 minutes. The radioactivity of the supernatant fraction that contained the released nucleotides, was determined in a liquid scintillation counter. A control reaction where the recombinant protein was not added was also carried out. All reactions were carried out in triplicate.

2.8.5. Nucleotide hydrolysis assays

DNA helicase activity requires the hydrolysis of NTP substrates to release the free phosphate to power its actions. NTP hydrolysis was measured using the method adapted from Tseng *et al* (2000). The method used is colorimetric and involves the detection of the released phosphate by the absorption of a 'molybdate blue' complex which forms upon the reduction of an ammonium molybdate-phosphate complex in strong acid solution.

The reaction was carried out in a 96-well plate in a 30 μ l assay volume which contained the assay buffer (50 mM Tris-HCl, pH 7.6, 10 mM MgCl₂, 0.2mg/ml BSA, 2.0 mM of individual NTPs/ dNTPs and M13 ssDDNA (17 μ g/ml)). The reaction was incubated at 37°C for one hour after which 125 μ l of a 1:1 mixture of 1.4% ascorbic acid in 1M H₂SO₄ and 0.36 % in 1M H₂SO₄ was added. The reaction was incubated at 37°C for 15 minutes for colour development to take place. Negative

controls were carried out where no enzyme was added to detect if there was spontancous hydrolysis of the dNTPs. The plate was then read at A_{690} using a microplate reader (Titertek).

2.9. Subcellular localisation of *PfPREX* gene products

2.9.1. Indirect immunofluorescence

For experiments with mitotracker staining, the live *Plasmodium* infected erythrocytes were incubated in the presence of 100 nM of Mitotracker[®] Red CMXRos (Molecular Probes) for 15 minutes at 37°C and washed 2 times in PBS, after which, normal blood smears with parasitised red blood cells were performed on slides that were then air-dried for 10 minutes. The slides were fixed in acetone and allowed to air dry. Later, the slides were blocked in blocking buffer for 20 minutes at room temperature. The slides were then incubated with the appropriate primary and secondary antibodies. Secondary antibodies were fluorescein isothiocyanate (FITC)-coupled antirabbit IgG (Diagnostics Scotland). Primary antibodies included those raised to the DNA polymerase domain, the DNA primase region and the DNA helicase region (peptide fused to GST – more details included in the results section).

The primary and secondary antibodies were diluted appropriately in the immunobuffer. Dilutions are detailed in the results section. In between incubation with the primary or secondary antibodies, the slides were washed twice in 1X PBS containing 0.2% Tween 20. Where indicated, slides were stained in 5 mM 4,6-diamidino-2-phenylindole (DAPI) (Sigma) for 30 minutes at 37°C. Coverslips were mounted with the anti-fade agent AF2 (Citifluor).

2.9.2. In situ hybridisation

2.9.2.1. DNA probe preparation

The ssurRNA PCR product was amplified using the ssurRNAF1 and ssurRNAR1 primers (Table 2.1). The appropriate PCR product was used as the DNA probes for the *in situ* hybridisation were denatured at 80°C for 2 minutes. The probes were then allowed to cool slowly down to 42°C. Approximately 75 ng of DNA probed was added to the same volume of 50% formamide for use in the experiment.

2.9.2.2. In situ hybridisation

Slides were prepared and fixed as stated in section 2.9.1. The slides were prchybridised in the hybridisation buffer (HB) for 2 hours at 42°C. Once the excess hybridisation buffer was removed, the DNA probe was added (6 μ l in 24 μ l HB) was added. The coverslips were mounted and the slides were incubated overnight at 42°C on a PCR machine (Hybrid).

The slides were washed 3 times in 0.1X SSC for 30 minutes at 42°C, followed by 2 washes in 1X SSC for 10 minutes at room temperature. Finally, the slides were washed in 1X PBS for 5 minutes.

2.9.2.3. Fluorescent detection of the *in situ* hybridisation DNA probe

The slides were placed in detection buffer for 5 minutes. The slides were blocked in BSA blocking buffer (2% (w/v) BSA in Detection buffer) at 37°C for 30 minutes in the dark. After the blocking step, the detection solution (10 μ g/ml of Streptavidin-Texas red in BSA blocking buffer) was added and the slides were incubated at 37°C for 1 hour. Three washes in detection buffer were done at 42°C for 20

minutes. Coverslips were mounted with the anti-fade agent for viewing under the microscope.

2.9.3. Fluorescent Microscopy

Slides prepared were viewed using a Zeiss Axiovert 35 inverted microscope equipped with a 100W Mercury-vapour lamp. The microscope was equipped with a Axiovision camera Axiovision Software (Zeiss). The DAPI filter had an excitation and emission peaks of 360 nm and 460 nm, the Texas red filter- 560 nm and 630 nm and the FITC filter- 484 nm and 510 nm respectively. Image analysis was performed using Axiovert software (Axiovert). The images were merged using the Adobe Photoshop 6.0 software.

2.10. DNA sequencing

Plasmids that contained the cloned PCR fragments were sent to MWG-BIOTECII (Milton Keynes, UK) for sequencing. The plasmids were sequenced in both directions to confirm their sequences. The data obtained was then analysed by comparing chromatograms using Vector NTI version 6 (Informax, Inc.).

2.11. Software

Prism version 3.02 (Graphpad software) (<u>http://www.graphpad.com/prism/Prism.htm</u>) was used to handle all the biochemical graph-fitted data.

The Vector NTI suite 6 (<u>http://www.informaxinc.com</u>) package was used to handle DNA and protein sequences. The suite contains different functional modules like $AlignX^{\textcircled{0}}$ for carrying out alignments and ContigExpress⁰ for the construction of contigs from shorter sequences.

ClustalX version 1.81 (Jeanmougin *et al*, 1998) was used to do the protein multiple alignments where stated. Specific parameters for the protein alignments are described the relevant results sections.

Mega (Molecular evolution and genetic <u>analysis</u>) version 2.1 (<u>http://www.megasoftwarc.net</u>) was used to build all phylogenetic trees. There are four different tree-building methods available with the software – <u>Neighbour Joining</u> (NJ), <u>Unweighted Paired Group Method with Arithmetric Mean (UPGMA), Minimum Evolution (ME) and Maximum Parsimony (MP). The default settings used for tree generation are discussed in the relevant result sections.</u>

VMD (Visual Molecular Dynamics) version 1.8.1

(http://www.ks.uiuc.edu/Research/vmd/) was used to visualise 3D structures of proteins.

A number of different databases were used to clarify *PREX* homologues, predominantly using BLAST servers associated with the sites listed below: -

Databases

- TIGR databases: <u>http://www.tigr.org/tdb/parasites</u>
- Parasite-genome (European Bioinformatic Institute): http://www.ebi.ac.uk/blast2/parasites.html
- The Sanger Center: <u>http://www.sanger.ac.uk/Projects/</u>
- NCBI: <u>http://www.ncbi.nlm.nih.gov/BLAST</u>
- Plasmodb: <u>http://www.plasmodb.org</u>

Alignment programme

DIALIGN: <u>http://www.genomatix.de/cgi-bin/dialign/dialign.pl</u> (Morgenstern *et al*, 1996). The parameters used for the protein alignments:- Threshold T =0 (DIALIGN uses diagonals to construct an alignment. Therefore, the threshold T influences the set of used diagonals: with T > 0, a diagonal is considered for alignment only if its *weight* exceeds this threshold. Regions of lower similarity are not aligned.)

Prediction of the presence of signal peptide

SignalP: <u>http://www.cbs.dtu.dk/services/SignalP/#submission</u> (Nielsen *et al*, 1997).

Prediction of apicoplast-targeted Plasmodium proteins

- PlasmoAP: <u>http://www.plasmodb.org/restricted/PlasmoAPcgi.shtml</u> (Foth *et al*, 2003).
- PATS (Prediction of <u>apicoplast targeted sequences</u>): <u>http://gecco.org.chemie.uni-frankfurt.de/pats/pats-index.php</u> (Zuegge *et al*, 2001).

ORF prediction software

• ORF finder : http://www.ncbi.nlm.nih.gov/gorf/gorf.html

Chapter III

Identification of a novel gene, *PfPREX*, from *Plasmodium falciparum*

j

3.1. Introduction - background to the discovery of the gene, *PfPREX*.

During a study carried out by Dr P.Neuville with Professor RS Phillips, (University of Glasgow) analysing antigenic variation in P. chabaudi AS strain, a gene with a putative role in this process was identified and named PomI (Genbank accession No: AAA84746.1). The gene was identified using hyperimmune scrum raised in mice against antigenically distinct variants of P. chabaudi, one from the parent AS clone and a second raised from a recrudescent variant clone. The serum was used to screen phage λ gt 11 expression libraries of cDNA from the parasite and several clones were recognised by the sera. The clones were all sequenced and one particular clone of 597 bp translated to give a putative protein containing an ATP-binding site (PomI). The PomI sequence was used in BLAST searches against the P. falciparum genome database (Sanger Center, TIGR and Stanford University). The search gave a longer sequence (an apparently full-length open reading frame, ORF of 6057 bp) in P. falciparum. On further investigation, it was revealed that it contained a single open reading frame encoding 2016 amino acids (Genbank accession No: AAN36724.1) with a predicted molecular mass of 235.8 kDa. Since work reported in this thesis indicates that the gene encodes proteins involved in plastid DNA replication, it was named PfPREX (P. falciparum Plastid Replication/Repair Enzyme Complex)

This chapter describes the work involved in elucidating the properties and arrangement of the *PfPREX* gene. The purpose of this chapter is to provide insight into the nature of the gene and also the background to the predicted functions based on sequence information.

3.2. Identification of the gene and sequence analysis

The nucleotide sequence of the gene studied in this thesis has been termed *PfPREX*. The predicted amino acid sequence, PfPREX was used in BLASTP searches in the NCBI website (<u>http://www.ncbi.nlm.gov/blast</u>) against non-redundant protein databases. Initially, the C-terminal domain of PfPREX produced a match to a thermostable DNA polymerase from *Aquifex aeolicus* (Genbank accession Number: AAC07735.1) with an E value of 1×10^{-53} (Figure 3.1a). The N-terminus of PfPREX showed homology to DNA helicases and primases from bacteriophages (E values of $7 \times 10^{-7} - 4 \times 10^{-5}$) and bacteria (Figure 3.1b). It seems that the *PfPREX* sequence encodes for a polypeptide with more than one functional domain (Figure 3.2). Therefore, it was decided to split the *PfPREX* sequence into separate regions to do more detailed homology modelling to provide a clearer indication of the roles of different domains of PfPREX.

3.2.1. Identification of a DNA polymerase-like domain in the *PfPREX* sequence

The C-terminal end of the sequence, which has been designated PfPREXpol, was used in a BLASTP search against the non-redundant protein database (Figure 3.1a) The predicted sequence shares 25-30% identity and 40-46% similarity with prokaryotic DNA polymerases from the family A group (Section 1.4.3.1.3). The results gave E values of 1×10^{-53} to 6×10^{-7} with the highest significant homology to the DNA polymerase of *Aquifex aeolicus* (Genbank accession Number: AAC07735.1). The other matches all belong to the members of the family A group such as *E. coli* DNA polymerase I, *Thermus aquaticus* DNA polymerase I and T7 DNA polymerase I.

<pre>(a) sp[067779 DP01_AQUAE_DNA_POLYMERASE I (POL I) >gi_7514773 p sp[051498 DP01_BORBU_DNA_POLYMERASE I (POL I) >gi[7434318 p pir][E72232_DNA-directed_DNA_polymerase I - Thermotoga_mari sp[P00582 DP01_ECOLI_DNA_POLYMERASE I (POL I) >gi[67054 pir pdb]1KFD Dna_Polymerase I (Klenow Fragment) (E.C.2.7.7 pdb]2KFN A_Chain A, Klenow Fragment With Bridging-Sulfur Su sp[Q59156 DP01_ANATH_DNA_POLYMERASE I (POL I) >gi[1405438]e pdb]1KLN A_Chain A, Dna_POLYMERASE I (POL I) >gi[1405438]e sp[Q59156 DP01_ANATH_DNA_POLYMERASE I (POL I) >gi[1074025]p pdb]1KLN A_Chain A, Crystal_Structure Of The Complex Of Dna sp[05949]DF01_RICPR_DNA_POLYMERASE I (POL I) >gi[7434819]p emb]CAB72805.1] (AL139074)_DNA_polymerase I [Campylobacter</pre>	$\begin{array}{c} \underline{213} & 1e-33 \\ \underline{189} & 2e-46 \\ \underline{175} & 4e-42 \\ \underline{159} & 4e-37 \\ \underline{159} & 4e-37 \\ \underline{159} & 4e-37 \\ \underline{159} & 4e-37 \\ \underline{157} & 2e-36 \\ \underline{155} & 4e-36 \\ \underline{155} & 6e-36 \\ \underline{154} & 8e-36 \\ \underline{154} & 8e-36 \\ \end{array}$	
<pre>(b) <u>gi19627449 ref NP_041977.1 </u> <u>gi130387470)ref NP_041977.1 </u> <u>gi19627447 ref NP_041975.1 </u> <u>gi19627447 ref NP_041975.1 </u> <u>gi17570806 ref NP_523316.1 </u> <u>gi17570806 ref NP_523315.1 </u> <u>gi19634014 ref NP_052088.1 </u> <u>gi19634013 ref NP_052087.1 </u> <u>gi19634013 ref NP_064752.1 </u> <u>gi19964622 ref NP_064752.1 </u> <u>gi129366716 ref NP_813761.1 </u> gi129366716 ref NP_813761.1 </pre> gene 4B/helicase [14,15] [Enter primase/helicase protein [Yers gene 4B/helicase protein [Yers gene 4B/helicase [I4,15] [Enter primase/helicase [14,1 helicase [Bacteriophage T3] >g DNA primase/helicase [Bacteriop RFprimase/helicase [Roseophage putative primase/helicase [Pse	59 58 55 55 55 55 55 55 55 55 55 55 55 55	7e-07 1e-06 1e-06 7e-06 7c-06 1e-05 1c-05 3e-05 4e-05
(c) gi[2350933]ref[NP 702000.1] POM1, putative [Plasmodium fal gi[114260]gb[AAA84746.1] POM1 [Plasmodium chabaudi chabaudi] gi[23487175]gb[EAA20985.1] POM1 [Plasmodium yoeldi yoeldi] gi[27684485]ref[XP 219939.1] similar to twinkle [Mus muscul gi[11141909]ref[NP 068602.1] twinkle; likely ortholog of mo gi[14582616]gb[AAK69558.1]AF292004] putative T/-like mitoc gi[24962647]ref[NP 722491.2] twinkle [Mus musculus] >gi[244 gi[21294867]gb[EAA07012.1] ENSANGP00000000580 [Anopheles ga gi[24583154]ref[NP 609318.1] CG5924-PA [Drosophila melanoga gi[20976892]gb[AAK69559.1]AF292005] truncated putative T7 gi[22028196]gb[AAK69559.1]AF292005] truncated putative T7	$ \begin{array}{r} 1779 \\ \overline{495} \\ \overline{493} \\ \overline{245} \\ \overline{244} \\ \overline{244} \\ \overline{239} \\ \overline{231} \\ \overline{228} \\ \overline{2066} \\ \overline{198} \\ \end{array} $	$\begin{array}{c} 0.0\\ e-138\\ e-137\\ 3e-63\\ 8e-63\\ 8e-63\\ 1e-62\\ 3e-61\\ 7e-39\\ 5e-58\\ 3e-51\\ 6e-49 \end{array}$

Figure 3.1. BLAST results with the description consisting of four columns (from the left): (1) identifier for the database sequence (Genbank Entry); (2) brief description of the sequence; (3) the (bit) score of the highest-scoring HSP (High Scoring Pairs) found for each database sequence; (4) the E value. (a) BLASTP results using PfPREX against non-redundant protein database (<u>http://www.ncbi.nlm.nih.gov/BLAST</u>) showing matches to prokaryotic DNA polymerases. (b) BLASTP results using PfPREX against nonredundant protein database (pre-2001) showing matches to DNA helicases and primases. (c) More recent BLASTP (post-2001) results showing matches to 'Twinkle' and its homologues. The BLAST results obtained has changed since the beginning of the project with newer entries into the database in which our sequence is homologous to, as shown in part b and c in this figure.



Figure 3.2. Diagrammatic representation of PfPREX protein with the different functional domains it encodes for. A signal peptide at the extreme N-terminus end followed by a DNA primase and DNA helicase domain, separated by a hinge region, and a DNA polymerase domain.

3.2.2. The family A group of DNA polymerases

The DNA polymerases from family A group have diverse roles such as gap filling, DNA repair and DNA replication (Albà, 2001; Burgers, 1998; Hubscher *et al.*, 2000) (Figure 3.3). Most DNA polymerases from this family, exemplified by *E. coli* DNA polymerase I, possess three domains which include a 5'-3' exonuclease domain, a central 3'-5' exonuclease domain and a C-terminal polymerase domain. Not all members of the family have all three domains (Figure 3.3). Sequence alignments of PfPREXpol with other family A DNA polymerases (Figure 3.4) indicated the absence of a 5'-3' exonuclease domain while the 3'-5' exonuclease and polymerase domains were observed. This is similar to the phage T5 DNA polymerase I and *Aquifex aeolicus* DNA polymerase I, neither of which have the 5'-3' exonuclease domain. The Klenow fragment of *E.* coli Pol I is obtained by protease cleavage which remove the 5'-3' exonuclease domain (Klenow & Overgaard-Hansen, 1970).

3.2.3 3'-5' exonuclease domain

The 3'-5' exonuclease activity usually provides proofreading for the DNA polymerase and this helps to enhance the accuracy of the DNA synthesis activity of the DNA polymerase which helps to enhance genome stability. The exonuclease functions by catalysing the excision of the nucleoside monophosphate (dNMP) from the 3' end of the DNA. Within the PfPREXpol domain, we have located all three conserved motifs (Exo I-III) of the 3'-5' exonuclease activity (Figure 3.5). The conserved motifs are DxET in the Exo I region, NX₂YD in the Exo II region and YX₃D in the Exo III region. The aspartate residues found in all three Exo regions and the glutamate in Exo I are thought to be essential for divalent metal ion binding which participate in the catalysis of phosphoryl transfer (Beese & Steitz, 1991; Bernad *et al.*, 1989; Derbyshire *et al.*,



Figure 3.3. Diagrammatic representations of the roles of different members of the family A group of DNA polymerases. 5'Exo box represents the 5'3' exonuclease domain, the 3'Exo box represents the 3'5' exonuclease domain and the Pol A box represents the polymerase domain.



Figure 3.4. Alignments of deduced amino acid sequence of the polA domain of DNA polymerases from family A. Sequences used are as follows: *Thermus thermophilus* (Genbank accession no:- CAA46900.1), *Thermus aquaticus* (Genbank accession no:- AAA27505.1), *Thermogata maritime* (Genbank accession no:- AAD36686.1), *Borrelia burgdorfei* (Genbank accession no:- AAC66909.1), *Escherichia coli* (Genbank accession no:- CAA23607.1), *Salmonella typhimurium* (Genbank accession no:- AAG43170.1), *Aquifex aeolicus* (Genbank accession no:- AAC07735.1) and the PfPREXpol sequence. These sequences were aligned using ClustalX 1.81. Black shading indicates identical amino acid, grey shading indicates conserved amino acid substitutions. The numberings on the left hand side shows the relative position of the amino acids in the respective sequences. The alignment has been culled to show the regions where the three conserved boxes are found, they are shown as motifs A, B and C above the sequence. Key residues involved in enzymatic activity are denoted below with an asterisk.



Figure 3.5. Pairwise alignment of the amino acid sequences of the Klenow fragment of *E. coli* and PfPREXpol using ClustalX 1.81 with default settings. The numbers on the left refer to the amino acid number of the individual sequence. Black shading indicates identical amino acids; grey shading indicates conserved amino acid substitutions. The various motifs are indicated with underlined text – Exo I-III (3'- 5' exonuclease activity) and Motif A-C (DNA Polymerase activity).

1995; Morrison *et al.*, 1991). The conservation of these essential residues in the PIPREX sequence suggests that it contains a functional 3'-5' exonuclease domain.

3.2.4 The DNA polymerase domain

Several amino acid motifs are conserved among all family A DNA polymerases. These conserved motifs have been defined in different DNA polymerase families based on sequence comparisons and mutation analyses (Joyce & Steitz, 1994; 1995). The PfPREXpol domain contains the A (DFKQIELK), B (AINFGLIYGM) and C (LCVHDE) motifs (Delaruc *et al*, 1990) (Figure 3.4), which are found in the polymerase domain. Motif A and C are part of the catalytic site of the polymerase. Motif A also forms a pocket for the incoming dNTP. The two aspartate residues present within them help to coordinate the two Mg²⁺ ions that stabilise the transition state of the enzyme and facilitate phosphoryl transfer (Steitz, 1998). The presence of all these essential residues in the PfPREXpol sequence suggests that it does encode a functional polymerase.

3.3. Identification of DNA helicase and DNA primase domains

Database searches using the N-terminal half of the PfPREX sequence revealed it to share homology to both a DNA helicase and a DNA primase. BLAST searches performed in 2001 or later revealed a new sequence added to the database that showed substantial similarity with the N-terminal half of PfPREX. That protein is called 'Twinkle' from rat (Genbank accession no: - XP_219939.1 Swissprot) (E value - $3x10^{-63}$). The same domain also showed similarity to DNA helicases and primases from bacteriophages like T7 and cyanophages at E values of $7x10^{-7}$ to $4x10^{-5}$. In these bacteriophages, the DNA G-like primases are fused to DNA B-like helicases. This is the same scenario observed with the T-odd bacteriophages.

3.3.1. The DNA helicase domain of PfPREX

The sequence alignment of the DNA helicase-like domain (termed PfPREXheli), revealed motifs conserved among the DNA helicases. DNA helicases are responsible for the unwinding of double-stranded DNA using energy from NTP hydrolysis. A key characteristic of these enzymes is the presence of conserved domains responsible for helicase function (Ilyina *et al*, 1992). There are several of these motifs which are essential for several biochemical features including nucleic acid binding, NTP binding and hydrolysis. The 'Walker motif', which encompasses the motifs A (H1) and B (H2), contains a typical purine NTP-binding pattern. This 'Walker motif' can be found in all helicases and also *p+16Xother NTP-utilizing enzymes. The 'Walker motif' can also be located in our PfPREXheli sequence (Figure 3.6). The Walker 'A' motif is the phosphate-binding loop ('P' loop). The Walker 'B' motif is involved in the dNTP binding. The residues responsible for this interaction are a conserved arginine and lysine. An aspartate found in the Walker 'B' motif is important as its carboxyl group is responsible for coordinating the Mg²⁺ ion, which forms a complex with the ATP/ADP through the outer sphere interactions.

In addition to the 'Walker motif', there are three other conserved motifs (H1a, H3 and H4) that can be located in DNA helicases, also present in the PfPREXheli sequence. Motif H1a contains a glutamic acid that has been suggested to play a role in catalysis (Sawaya *et al*, 1999). The role of motif H3 is unclear. It is hypothesised, based on studies of the T7 gp4 DNA helicase that it is involved in energy transduction (Sawaya *et al*, 1999). It has also been suggested the motif H3 participates in the interaction of Mg²⁺ with ATP/ADP and contributes to the binding of oligonucleotides (Caruthers & McKay, 2002).
T3 gp4 phiYeO3-12 T7 GP4 RP PfPREXheli twinkle DNABECOLI Bacillus T4 Aquifex	79 79 80 53 89 62 59 60 181	ALSIKORVK AMTSEDAVED ALSIKORVK AMTSEDAVED ALSIKORVK AMTSEDAVED SRTLFDVVTTPLHGRI AEYPI NDLRQRILESIKYPTRINSV RQLREEVIGELSNVEQAA DATVARIEQLFQQPHIGVTG VQTYDNIEQLYNRKG - TG DWDDYEARW SYMKKARK	FDECQGINDRIL AKG FDECQGINDRIL AKG FSECTGINDKIL AG FAAUSTIGILL KSKTIPSINKYLY LAM WSRFPDINRILK HK WSRFPDINRILK HK NT YDDINKKIA LQPS PT FTEIDRMIA FQRN PFKLRINKIKCAETG VIKLQEKKKELYEK	VVMVTSGSCMGRSTFVRCQAL VIMVTSGSCMGRSTFVRCQAL VIMVTSGSCMGRSTFVRCQAL SIWTGSTVVCFTTLSQLSL TVETGPTSSFTTFISEYAL DIIIVAARPSWGRTAFALNINQ TINVLMGVNVGRSLGLCSLA- EKAVNKILSRAQVVCTTNSTA
				Walker A motif (H1)
T3 gp4 phiYeO3-12 T7 GP4 RP PfPREXheli twinkle DNABECOLI Bacillus T4 Aquifex	139 139 136 113 149 122 118 120 241	AWCKRMGARVGIAMLEFSVI AWCKRMGARVGIAMLEFSVI QWCTAMGAKVGIAMLEFSVI SLINQD-QRVGYIALEFSVI DYCIQGVSTWGSFEINNKK DLCSQGVNTWGSFEISNAR NAAMLQDFPLFFSHMGA ADYLQLGHNLYISHMAA GSEVLQNLNFDWVIIDFAT	DIQDMMGLNNKVRIROS DIQDMMGLNNKVRIROS AEDLIGLHNKVRIROS SIGURLMTVAANKPUHLN LGKMMINQFCGNNLEKNI LARMMITQFAEGRLEDQL 21MMRSLASLSRVDQTKI 21VMRICAEGNINAQNI 7CAKRIDANMLDVSIDDI AEPSCHIPLINGKKLIM	EVKKAIAEDGRFDEWYDELFG EVKKAIAEDGRFDEWYDELFG SLKREIIENGKFDQWFDELFG NELPTDELRTAFDSTLG LFDIYADKFE KYDHWALRFE RTGQLDDELWARISGTMGILE RTGNLTEEFWGKLTMAMGSISN DGHISYABYKGKMEKARKST AGDHKQLPPTVLSQEAQEAISY
T3 gp4 phiYeO3-12 T7 GP4 RP PfPREXheli twinkle DNABECOLI Bacillus T4 Aquifex	199 199 190 162 198 182 178 180 301	Hla DDTFHIYDSFAEAEARLAI DDTFHIYDSFAEAEARLAI NDTFHIYDSFAEAETRIAI TGRVFIRIGESVDPSJUN LLPLKFLKFH STNIOVID DLPIYFMTFH QQSIRTVID KRNIYIDDSSLTPTSVRSR S-GIYIDDTPIRVSEIRAK LGRIIYKQYPTGGADANTFR TLFERILLYSEEIYJIIRIQ	KLAMMRTCLGCDVIVLUH KLAMMRTSLGCDVIVLUH KLAMMRSSLGCDVIILUH DIRELTKAHEVQWIVLUH AMDMAVYBYDVKHIIIIN MQHAVYVYDICHVIIN ARRIAREHGGIGLIMINY CRRLKQESG-LGMILIY SLINELKLKKNFVPTIII YRMNKKIMEFSNKVFYEG	ISIVVSASEES-DERKMIDRLM ISIVVSASEES-DERKMIDRLM ISIVVSASGES-DERKMIDNLM ISILLSGNESI-DERKMIDIVM QFMININKFS-DIYELONIAI QFMMGHEQLSTORIAAOIYII QLMRVPALS-DNRTLEIAEIS QQLIQGSGRSKDNRQQEVSEIS VDYLGICKSCRIRVYSENSYTT KLIADKSVENHTIKDLINPEK
T3 gp4 phiYeO3-12 T7 GP4 RP PfPREXheli twinkle DNABECOLI Bacillus T4 Aquifex	258 258 249 221 258 241 236 240 361	TKLKGPAKSTGVVVVICHL TKLKGPAKSTGVVVVICHL TKLRSVEETGGVVVVICHL TKLRSVEETGGVVVVICHL DFFSSSTNKNHTTVVP GVFRKFATDNNCHVTIVIP RSLKALAKELNVPVALSOL RELFSIARELOVPVIALSOL VAI BELRALAVETETVLW LHEIPEPFKSVEEPEKVVF	Wal	ker B motif (H2) DLRGSCALROLSITITALERNQ DLRGSCALROLSITITALERNQ DLRGSCALROLSITITALERNQ OLRGSCALROLSITITALERNQ OLRGSCALROLSITITALERNQ SVF SVKST.EAUNVFILQHV SIFGSAKASCEAUNVFILQHV SIFGSAKASCEAUNVFILQHV DLRESCSTECDALINFTY DE DRESCSTECDALINFTY DE DRESCSTECDALINFTY DE DRESCSTECDALINFTY DE DRESCSTECDALINFTY DE DRESCSTECDALINFTY DE DRESCSTECDALINFTY DE
		H3	14 Sec. 103	H4

Figure 3.6. Alignments of deduced amino acid sequences of DNA helicases. Sequences are as follows: These full-length sequences were aligned using ClustalX 1.81. Sequences used are as follows: T3 gp4 (Genbank accession no: - CAA35135.1), phiYe03-12 (Genbank accession no: - CAB63608.1), RP (Roseophage, Genbank accession no: - AAL73266.1), twinkle (*Mus musculus*, Genbank accession no: - AAL27647.1), DNABECOLI (Genbank accession no: - AAA23689.1), Bacillus (*B. subtilis*, Genbank accession no: - BAA05176.1), T4 (Genbank accession no: - AAD42466.1), Aquifex (*A. aeolicus*, Genbank accession no: - AAC07803.1). Black shading indicates identical amino acids; grey shading indicates conserved amino acid substitutions. The numberings on the left hand side is the relative position of the amino acid in the respective sequences. Alignment has been culled to show the region where the five conserved boxes are found; they are shown as motif H1-H4.

Finally, motif H4 has a role that may be linked with DNA binding (Gorbalenya & Konin, 1993; Washington *et al*, 1996). The presence of these conserved motifs suggests that PfPREX encodes for a functional DNA helicase.

3.3.2. The DNA primase domain of PfPREX

The DNA primase domain of PfPREX revealed less sequence homology to the T7 phage gene4 like protein products than the DNA helicase domain. As automated alignment failed to identify all motifs, potentially significant sequence similarity was ehecked by eye. Key motifs, thought to be important for DNA primase activity were found in the PfPREX sequence (Ilyina *et al.*, 1992). The best probability match obtained during the BLASTP search was to the T7 gene 4 protein (Genbank accession no: CAA24405.1) at $7x10^{-7}$, which itself encodes for a combined DNA helicase /DNA primase.

3.3.2.1. Conserved features of DNA primases

The multiple alignment of the DNA primase domain of PfPREX (termed PfPREXpri), along with various primases of prokaryotic origin consisting of both bacterial and phage primases, revealed the presence of several motifs conserved among the group (Figure 3.7). PfPREXpri also contains these motifs, DNA primase is actually a specialised RNA polymerase, which synthesises a short RNA strand which DNA polymerase then uses for strand elongation. Among DNA primases, there are key motifs responsible for function. Prokaryotic DNA primases consist of three key domains (Figure 3.8): - (1) a zinc binding domain (Kusakabe & Richardson, 1996), (2) a catalytic domain (Aravind *et al.*, 1998) and (3) either a helicase domain or a region for the interaction with a helicase (Tougu *et al.*, 1994). PfPREXpri contains all three domains; the helicase domain (PfPREXheli) will be discussed in the next section.

88

٠z



Figure 3.7. Alignments of deduced amino acid sequences of DNA primases. Sequences are as follows: phiYe03-12 (Genbank accession no: - CAB63608.1), T3 gp4 (Genbank accession no: - CAA35135.1), T7 gp4 (Genbank accession no: - CAA24405.1), RP (Roseophage, Genbank accession no: - AAL73266.1), Salmonella (S. typhimurium LT2, Genbank accession no: - AAL22084.1), E.coli (E. coli DnaG, Genbank accession no: - CAA23531.1), Aquifex (A. aeolicus DnaG, Genbank accession no: - AAC07430.1), T4 (Genbank accession no: - AAA32554.1). Sequences were aligned using ClustalX 1.81. Black shading indicates identical amino acids; grey shading indicates conserved amino acid substitutions. The alignment has been culled to show the region where the six conserved boxes are found; they are shown as motif 1-6.

From the alignment, it is evident that two pairs of cysteine residues, a key feature of zinc finger/zinc-binding motifs, are found within the first domain. This zinc motif is present in other primases from both prokaryotic and eukaryotic sources. Among prokaryotes, the zinc-binding motifs are found near the N-terminus of the protein. It was shown in T7 gp4 protein, when the zinc motif was disrupted (Mendelman *et al.*, 1994) or in the P4 primase (Ziegelin *et al.*, 1995) where the two cysteine residues were replaced by glycine residues, priming activity was lost. The zinc motif is therefore

postulated to be essential for the primase activity. Work with T7 gp4 suggested that the zinc motif might be involved in priming site recognition on the template DNA (Berstein & Richardson, 1988; Mendelman *et al.*, 1994). In the case of PREXpri, the two pairs of cysteine residues, which make up the zinc-binding motif, are found (Figure 3.7).

In addition to the zinc-binding motif, five other motifs conserved among PfPREXpri and other prokaryotic primases (Ilyina *et al.*, 1992) are observed in the alignment. These five motifs are found within the catalytic domain (also known as the RNA polymerase domain). Motif 2 is the only motif for which no functional or structural role has been yet assigned. Motifs 3 to 6 have been implicated to play roles in primer synthesis. A glutamate residue in motif 4 is predicted to be essential for primer synthesis (Strack *et al.*, 1992). More specifically, this glutamate residue provides a base to facilitate nucleotide polymerisation (Aravind *et al.*, 1998).

The key characteristic of motif 5 is an aspartate dyad (DxD) that it is believed is involved to be in the coordination of the Mg^{2+} ion (Aravind *et al.*, 1998). An additional aspartate dyad (DxD) is found in motif 6 among all bacterial DNA G-like sequences while among bacteriophage sequences, only the first aspartate is present. There is only one aspartate residue in PfPREXpri, similar to the bacteriophage primases. Motifs 4 and 5 are especially important because they contain these conserved acidic residues which are glutamate and aspartate respectively. They are preceded by runs of hydrophobic regions predicted to form beta strands. Motifs 4 and 5 comprise the TOPRIM domain which is so named because the two motifs are found conserved not only in DNA primases (DNA G-like) but also among DNA topoisomerases. The DxD motif is central to the TOPRIM domain, as it is required for the activity of all TOPRIM- on PfPREX sequence, it is likely that it may encode for a functional primase.



Figure 3.8. Diagrammatic representation of the 3 key domains in prokaryotic DNA primase consisting of a zinc-binding domain, a catalytic domain and a helicase or helicase-interaction region. In addition, the conserved motifs of PfPREXpri in those regions are shown based on the conserved motifs discovered by Ilyina *et al* (1992).

3.4. The hinge domain of PfPREX

There is a region of sequence between the helicase and the polymerase (994 a.a to 1392 a.a) that does not have any marked similarity to any protein sequence within the database as judged by BLASTP searching. This region that separates the DNA primase/helicase and DNA polymerase domains has been termed the hinge region. Other multifunctional proteins (at least five have been characterised) from *P. falciparum* have hinge regions that appear to be involved in the proper folding of the protein. One example is the bifunctional ornithine decarboxylase S-adenosylmethionine decarboxylase (Krause *et al.*, 2000; Müller *et al.*, 2000) (ODC/AdoMetDC). It is not clear whether a similar role is played by this sequence in the PREX protein – indeed evidence presented later suggests that it may play a role in the processing of the protein rather than merely a hinge separating domain.

3.5. Identification of a transit peptide at the extreme N-terminal domain

The first 50 amino acid sequence from the N-terminal domain of PfPREX sequence was entered into the Signal P online server at http://www.cbs.dtu.dk/scrvices/SignalP that uses the von Heijne algorithm (Nielsen et al, 1997) to predict the presence and location of signal peptides. A signal sequence and a cleavage site between amino acid position 19 (Cysteine) and 20 (Isoleucine) in PIPREX were detected. In addition, the N-terminal amino acid sequence was seen to possess the characteristics of a bipartite transit sequence, which consists of 2 functional domains, a signal sequence followed by a targeting sequence, where the signal sequence is responsible for channelling the protein into the secretory pathway while the transit sequence is responsible for targeting the protein into the correct compartment. On the other hand, chloroplast-targeted proteins (Bruce, 2000) only possess transit sequences. Typically in an apicoplast bipartite transit sequences, the signal sequence part contains a

short run of hydrophobic residues, followed by the von Heijne cleavage site. The apicoplast targeting sequence carries an overall positive charge, very much like in the plant plastid targeting sequences (Waller *et al*, 1998). It is also known that *Plasmodium* apicoplast targeting sequences have a net positive charge due to the large number of lysine residues. They also possess a large number of asparagine residues (Foth *et al*, 2003). These features are evident in PfPREX putative apicoplast targeting sequence. Therefore, PfPREX has the key characteristics consistent with it being an apicoplast targeted protein.

To confirm its status, we submitted the first 50 amino acids to PATS (Prediction of Apicoplast-Targeted Sequences) (Zuegge *et al*, 2001) and PlasmoAP (Foth *et al*, 2003), two programmes designed to specifically determine potential *Plasmodium* apicoplast targeting sequences. Both programmes returned results that indicated the presence of an apicoplast targeting sequence in PfPREX (Figure 3.9).

In addition, we wanted to test whether the programmes are sufficiently robust to detect apicoplast targeting sequences and differentiate them from other targeting sequences such as mitochondrion-targeting sequences. Both programmes were therefore tested with other known targeting sequences including the leader sequence of HSP60 from *Plasmodium*, which has been shown to target to the mitochondrion (Sato *et al*, 2003). Both PATS and PlasmoAP predicted this sequence to lack apicoplast targeting potential. The leader sequence from acyl carrier protein (ACP) from *Plasmodium*, which has been shown experimentally (Waller *et al*, 2000) to be targeted to the apicoplast, was also analysed by the two programmes, both predicted the presence of an apicoplast targeting sequence. Therefore, both programmes seem to be

Apicoplast targeting peptide prediction

Query sequence: (Please note, that only the first 150 AA have been taken into account for the analysis.)

>query

MLLYKFYFLYFLLVHLSLCIRYRNQNKTDSYLKTNYKLLK KRKKYENRRY

Result:

The protein has a good signal sequence.

The submitted peptide sequence seems to have an apicoplast-targeting sequence.

Complete PlasmoAP output for query :

Criterion	Value	Decision
apicoplast-targeting	5 of 5 tests positive	++
Ruleset 1		
Ratio acidic/basic residues in first 22 amino acids <=0.7	0.143	yes
Does a KN-enriched region exist (40 AA with min. 9 K or N) with a ratio acidic/basic <=0.9	0.167	yes
Ruleset 2		
number of acidic residues in first 15 amino acids (<=2)	1	yes
Does a KN-enriched region exist (40 AA with min. 9 K or N) ? Ratio acidic/basic residues in this region <0.6	0.167	yes
Is the first charged amino acid basic ?		yes

Figure 3.9. Result of the PlasmoAP (Foth *et al*, 2003) apicoplast targeting peptide prediction where the first 50 amino acids from PfPREX were entered.

sufficiently robust to differentiate between selected proteins that target to the apicoplast and those targeted to the mitochondrion, PfPREX therefore does appear to possess a putative apicoplast targeting sequence at its extreme N-terminus.

3.6. PREX homologues in other Apicomplexans

3.6.1. Homologues in *Plasmodium*

Since the partial *PREX* gene (formerly known as *POM*I) was discovered in *P. chabaudi*, indicating that the gene was conserved in the genus *Plasmodium*, it was decided to explore other apicomplexan genome databases to search for the presence of a full-length sequences corresponding to the *PREX* gene. The predicted amino acid sequence of PfPREX was used in a TBLASTN search using the Plasmodb database (http://www.plasmodb.org) to search for other *Plasmodium* homologues. Full length PREX homologues were found in *P. vivax*, *P. yoelii*, *P. knowlesi* and *P. berghei* while partial PREX sequences were found in *P. chabaudi* and *P. reichenowi*.

Full-length sequences of the homologues from other Plasmodium species from the genome databases were acquired (Appendix C). The full-length sequence of PyPREX from P. yoelii was retrieved from the NCBI protein database with the Genbank accession number of EAA20985 (Carlton et al, 2002). The PyPREX is 1813 amino acids in length and has a sequence identity to PfPREX of 65%. The homologues from P. knowlesi and P. berghei were assembled from contigs obtained from the Plasmodium genome database from the Sanger website centre (http://www.sanger.ac.uk). The PkPREX (P. knowlesi) is 1861 amino acids in length and has a sequence identity to PIPREX of 65%. Like PfPREX, it does not appear to contain any introns. PbPREX from P. berghei is 1770 amino acids in length and is 63% similar to PfPREX. PbPREX sequence is interrupted by at least 3 introns (based on the ORF finder http://www.ncbi.nlm.nih.gov/gorf.html). For *P. vivax* homologues, a TBLASTN search was carried out against the 5X contig database on the TIGR website (<u>http://www.tigr.org</u>). A full length PvPREX was assembled and is 1860 amino acids in length with no intron and has 64% similarity to PfPREX (Table 3.1). For *P. chabaudi*, it appears to possess a PREX homologue but only a partial sequence (1345 amino acids) was obtained. As more genome sequence is entered into the database, a full-length sequence of PcPREX will be assembled when more sequence data is available.

Since PIPREX contains an N-terminal bipartite apicoplast targeting sequence, it was also of interest to check whether the homologues from other *Plasmodium* species also possess a similar bipartite plastid targeting sequence. We submitted the first 50 amino acids of all the six *Plasmodium* PREX homologues into PlasmoAP (Foth *et al*, 2003) to search for the presence of such sequences. PlasmoAP first determines the number of acidic amino acids in the first 15 amino acids. The minimum number of acidic amino acids considered for inclusion as a potential targeting sequence is two. The ratio of basic to acidic amino acids among the first 22 residues must be greater than seven or equal to 10. The presence of at least nine lysine or asparagine residues and a basic to acidic residues ratio in this stretch must be either greater than or equal to 10/9 (ruleset 1) or 5/3 (ruleset 2) respectively (Figure 3.9). Finally, PlasmoAP checks for the charge of the first charged amino acid which should be basic. Table 3.2 shows the summary of the results obtained from PlasmoAP of the six *Plasmodium* PREX homologues. All of the *Plasmodium* homologues possess an N-terminal bipartite apicoplast targeting sequence.

All of the homologues (except for *P. berghei* and *P. chabaudi* – because we do not have the full *PcPREX* sequence) also have the same cleavage point, between position 19 and 20 in spite of divergent sequences. There seems to be a consensus sequence at the cleavage point among the homologues (Table 3.2) xLx-I/VR.

3.6.2. PREX homologues in other Apicomplexans

In the previous section, it has been shown that different *Plasmodium* species possess PREX, all of which contain an N-terminal bipartite plastid targeting sequence. Whether other apicomplexans also possess PREX homologues has also been assessed.

TBLASTN searches using PfPREX amino acid sequence against the various apicomplexan genome databases. The *Theileria* genome database was sed and a PREX homologue was found. Dr Shigeharu Sato (NIMR, Mill Hill) found PREX homologues in both the *T. annulata* (Ta) and *T. parva* (Tp) genome databases and assembled the full-length sequence from the contigs. Ta and Tp homologues (TaPREX and TpPREX) have 33% and 34% sequence identity respectively to the PfPREX peptide sequence (Table 3.1). The TaPREX is 1786 amino acids in length while the TpPREX is 2060 amino acids in length.

Attempts were also made to find PREX homologues in *Eimeria tenella*, *Babesia bovis*, *Cryptosporidium parvum* and *Toxoplasma gondii*. For *Eimeria tenella*, insufficient genome information prevented a robust search for a PREX homologue. For *B. bovis* and *C. parvum*, short fragments with homology to PfPREX were isolated but there is also insufficient sequence information to enable derivation of full-length sequences for either. For *T. gondii*, we found significant matches to data available in the

T. gondii genome database but we were not able to obtain the full length PREX homologue from *T. gondii* which appears to be highly interrupted with introns.

	Peprex	PkPREX	TaPREX	TpPREX	PyPREX	PfPREX	PvPREX	PbPREX
PcPREX	100	51	32	33	71	46	51	72
PkPREX		100	38	38	69	65	87	68
TaPREX			100	90	41	33	37	41
TpPREX				100	42	34	38	42
PyPREX					100	64	69	94
PfPREX						100	64	63
PvPREX							100	68
PbPREX							<u>1</u> .	100
A State of the Sta	1.1.1.1.5		1.	b	1	Same and	and the second	Sec. 1.

Table 3.1. Similarity table of PREX homologues from *Plasmodium* and other apicomplexans. Abbreviations: *P. falciparum* (Pf), *P. berghei* (Pb), *P. chabaudi* (Pc), *P. knowlesi* (Pk), *Theileria annulata* (Ta), *T. parvum* (Tp), *P. vivax* (Pv) and *P. yoelii* (Py). The numbers represent the percentage similarity between each pair of proteins.

Signal Sequence	Apicoplast	Cleavage site	Cleavage sequence
+	+	19/20	ALG-IR
-	+	-	-
+	+	19/20	VLA-VR
+	+	19/20	ALA-VR
NA	NA	NA	NA
+	+	19/20	SLC-IR
	Signal Sequence + - + + NA +	Signal SequenceApicoplast++-+++++++NANA++	Signal Sequence Apicoplast Cleavage site + + 19/20 - + - + + 19/20 + + 19/20 + + 19/20 NA NA NA + + 19/20

Table 3.2. Table showing the *Plasmodium* PREX homologues and the presence of a signal sequence and apicoplast targeting sequence. Included also is the cleavage site of the signal sequence and the cleavage sequence for all the homologues. Abbreviations: *P. falciparum* (Pf), *P. berghei* (Pb), *P. chabaudi* (Pc) *P. knowlesi* (Pk), *P. vivax* (Pv) and *P. yoelii* (Py). NA- sequence is not available.

3.7. Determination of the arrangement of the different domains within PfPREX

Sequence similarity predictions have indicated that PfPREX encodes for several different activities. Alignment with the other PREX homologues showed variation in the degree of similarity across the protein which might indicate the structural arrangement of the functional domains within the sequence.

A multiple alignment of all of the PREX homologues allowed for the determination of the approximate limits of each individual domain based on the regions of best conservation (Figure 3.10). From the alignment, it is predicted that the DNA primase domain of PfPREX spans from position 114 to 470. Within this region, a high degree of similarity among the different PREX homologues and the conservation of the key primase conserved residues are observed. After this region, there is an area where there is less conservation before the helicase domain starts. The helicase domain is predicted to start at position 511 and runs to position 684 in PfPREX. This domain again is very well conserved among the PREX homologues.

The hinge region of PREX, shows less conservation overall. However in more or less the middle of the hinge, there is an area that is relatively well conserved. The hinge region did not show homology to any known proteins in the BLAST searches. The hinge region in PfPREX spans from position 1000 to 1360. The conserved domain within the hinge region, which spans from 1228 amino acids to 1264 in PfPREX, might be functionally relevant. However in the absence of homology to other proteins, that function remains speculative. As demonstrated later, the DNA primase/helicase and DNA polymerase domains appear to be separated by cleavage

Pyprex	1	PWYL FF
PbPREX	1	P
PCPREX	1	G
PKPREX	1	VRCOL
PVPREX	1	WVRCR LF
PFPREX	1	
TAPREX	1	
TOPREX	1	EPTY TNO
iprimi	-	
		Apicoplast targeting sequence
Pyprex	10	FFINALCHRRKELINE FILTER
PBPREX	3	FTALLOKKNELKEY IN A
PCPREX	3	LIK G OR KNEL KNY INTO A COMPANY AND A
PKPREX	10	ILIHYV AV NK-SPRADE LTTCE VK-RR (GS AY-KRKV
PVPREX	10	ILITYA AVEN
PFPREX	10	FLLVH S CHYR-CONTDS LLTR
TAPREX	10	N NETT REPKORLIG STV PLFF TRLILISSLLOFSE-SV T SOGS S SKRGRK
TPPREX	10	ANNEPT K SK RYY VALLEFT TRLILISCL PESE-CVAT SOGS S SKRGRO
- P	- 75	
PyPREX	60	KINGKO AVYKONDIE
PEPREX	48	SENEKI KOGRAWYKENNDIE
PCPREX	47	KENSE KORSAWYKSNNNIE
PKPREX	55	EDERLG RSF D CLEGGRSN
PVPREX	55	RGEDLGS ARMDDS LWGRR
PFPREX	57	GUNNN YDAINNNTNN I YNNDTYKNRIFFRNEKNDEN
TAPREX	69	TERPE PONTEKLATTHKLNAVDKESMISMKSLKOLSNLRNLIREKYNKGS
TOPREX	66	FEPPET CNTERTINGESPUDKLNMTSVKSLKDTSNLRNLTEFRYNKED
. Pr. min	00	
PyPREX	82	GSNEK COMNENE I
PbPREX	70	GSNE SCONE IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
PCPREX	69	SKNEN SKIN NYTKNKIN MSHOT
PKPREX	75	CSSAGNHDNHVDRUNSGARRENSIG
PVPREX	75	NRCSGGRVCRSTRE
PFPREX	96	IKTNEYSEYEKRNKIKLNSE
TaPREX	121	RNYKN FYKC FYKNLNEPOIDGGNIIYGSOFYAYSPSSYENVPSYNGSY DE NUS
TOPREX	118	RNYRN SFYKC FVSNLNKPOMESGKIIDGGSFYSYPPTSYPNMPSYNGSYC DE NI DU
-1		
PyPREX	108	SKY WOIN IN DVY NO DIRKKY BY DE ROVING BY OPEOPERSY AND WORLD FROM WAS
PbPREX	96	Reache and a should be written and a state of a state of a state of the state of th
PCPREX	95	SKINDED DE ANTRE SKIEGE DE ANTRE DE LE
PKPREX	101	SKEEPENERS STORES STORES STORES IN THE CONTRACT STORES IN A STORES AND A STORES AND A STORES AND A STORES AND A
PVPREX	101	Skreettittiin s <mark>s</mark> toppusta s e toretypeet neeroos an erevolu hydenin aksys
PFPREX	122	SKYYKIN SCHOOL HELEY STREET TO THE COUNTY AND
TAPREX	181	NH RIMGE IVE TO NEE VOE FRESIE
TOPREX	178	SNHERNMGE TVE AND SHE IN SPIELY DESERVED FOR THE SHULFS HERE SHE
		Zinc-binding domain –H1
Pyprex	168	CHROCK COSE A DOI STORE AND COSE AND COSE PERMID DATE KASE CHORACE AND A COSE
Pbprex	156	
PCPREX	155	
PkPREX	161	CHERCENE OF A DECEMBER OF A
PVPREX	161	CHENE MILLER AND
PFPREX	182	SHROOM SHALL HERE I SHIESTY HNNNFY EEEEKINLHPMONTNEH DA
TaPREX	241	PHENERICAL PRAFTING DE PRAFTING DE TENERE
TOPREX	238	ENROPS AL FORMATING DE PGAFDIASDPITSPFGAKKVOKDPDVE IE EVNKEEDE YN

PyPREX PbPREX PcPREX PkPREX PvPREX PfPREX TaPREX TpPREX	221 209 208 215 214 239 301 298	SEARCH THE LESS STATES
		HZ
PyPREX PbPREX PcPREX PkPREX PvPREX PfPREX TaPREX TpPREX	281 269 268 275 274 299 359 356	E ENN CKS YKGIGKIEDGG I E ENN CKS YKGIGNNDGG I SIGLGNRNNNNGG T SIGLGNRNNNN
DUDDDV	205	HJ
Pyprex Pbprex PcPrex Pkprex Pvprex Pfprex Taprex	325 313 312 313 317 359 407	SOLUTION SOLUTIA
TPPREX	403	GGAWGOFGGHLIQEAQ-EKESINI SIDE HADSI HADSI HADSI KUTABUKU NGULAAN
		H4
PyPREX PbPREX PcPREX PkPREX PvPREX PfPREX TaPREX TpPREX	385 373 372 373 377 419 466 462	N T Q D V ER KNMLTKG D Q L ER KNMLTKG D Q L ER KNMLTKG D Q L ER KNLLTKG PDV R OKSRLTKK NE Y AP Q ISH A L Q VK VHPL TPTTTAT T GTT
		115
PyPREX PbPREX PcPREX PkPREX PvPREX PfPREX TaPREX TpPREX	438 426 433 437 479 517 515	GLLLPLMIGDNAIGVAEQKQDAIKENTQSGE
PyPREX PbPREX PcPREX PkPREX PvPREX	438 426 425 464 468	ERKKNKIINLNPNSSSLPNEDENVIVSTDS ERKKNKIINLNPNSSSLPNEDENVIVSTDS ERKKNKIINLNPNSSSSEENALASSNS KNGNTECIAGGNEPNTAVDPISGTKKDQITTKKSSGDGKN ENGHSENTLDGSSPNTAADPTSGTKKDKTAKEGDRRK
PfPREX	539	FNGQKISSNNINVNILKNKKNETDNITNENKSDNEKEIQNEIVIEDNNNK
TAPREX	517	
TOPPEX	515	

Primase domain



PyPREX PbPREX PcPREX PkPREX PvPREX PfPREX TaPREX TpPREX	818 779 752 850 851 943 810 808	IGSNNSINGLNING YEVLOUSS SINGAC AND
Pyprex Pbprex Pcprex Pkprex Pvprex Pfprex Taprex Tpprex	878 839 812 905 906 1000 866 863	IGKCIADSIGLILISSIR NIGKCIADSIGLILISSIG TGKCIADSIGLILISSIG GADGRISGURA TNNPNASS GADGKANGUGGTN PNASS SKLDPNLRINNID IGTEI DNSDMNNKNVTLYVDSLENIKTISTDDKTNDDKRDVNEI DQTESPTIPSGFPINGGSE JQPDTPT PSGFPINGPE
Pyprex Pbprex Pcprex Pkprex Pvprex Pfprex Taprex Tpprex	899 860 833 926 927 1060 887 884	JETERGDIFSEIKEYEKNTE EDIEKEKEKYSKISEIJKKSDE (ITNTKE DNIEA GTERGNIFNENKEYGKNTE EDIEKDIRYSKIIE KQNNE (ITNTKE DSIDV GTERGDIFSEIKENGKNAE EDIEKEKKKCIKISE KQNDE (A TKE NSVIA VHIAPNECRNDS ATDSLR ONN NSVNQVGDAE DPTSNN VVN DK EE LKCSQN- NSTDG LVDSSR ONR N PVNQVGG E DPT SNIVG A GRGEK KSIKNNERKNILKID NKSLGST N NY N NK NNNNNNNKSNQE E NNID KSIG VKSVS SNFKINLPQVNST LST PMN PI TSTMITE YQS AVPAS TNYRER TKP VS SNFKLNLPQVNGS VSN PIG PVGT TS DLQSRPTPACTNYKDK
Pyprex Pbprex Pcprex Pkprex Pvprex Pfprex Taprex Tpprex	957 918 891 981 981 1120 944 939	CS SSNET SKEKTTEKKS TE VGF N D ILLSS KN D DADKMDDKN ITRKF CS SSNET SKEKTTEKKS TE VG HN DS F SSKN D DVDKMDNTN ITGNF SIGSLE SKEKTTEKKS TE VG N DS LLGSKN E NTDTMNNKN T GKY GR KK MLKGD QGRTTTP GKTTKSDKPSGDII VO SPNTKRE O AAPD P QNV GG KN-LLGT GDTAPP GGTPTSDKPSGGSS V TPAKPTGE OLLTAD S QSVP NKNINGNSKGNN INNNNKNK SSSSSSSS SN YNNNEGIKM LNTSAQNNIPFDTIW ERNNLGITGIDLLIL PV QYV KILOYY TSNGN ESG SST OF STESSGVDLTK ERNNVGL GIDLLIL PI QQI KILOYY SSNSNSDSS SSG P GEDTSGDDNTN
PyPREX PbPREX PcPREX PkPREX PvPREX PfPREX TaPREX TpPREX	998 959 932 1017 1016 1151 1009 1003	DIDADKMDDI DIDADKMDDI DIDVDKMDNTI EINTDTMNNN CUNSPNTKRE EAAPD
PyPREX PbPREX PcPREX PkPREX PvPREX PfPREX TaPREX	1017 978 951 1041 1040 1180 1004 999	YLLSNESTEN KEUR DEKG FORVVISIS KEVINNSPULPEN VI TLSNESTEN KUDEKG FORVVISIS KEVINN PULPEN VI TLSNEET KUEDDOG FOVISKINN PULPEV RISSITIET NKR MLERITIN KOTTTTTTT TRSSITIET NKG TL DRITIN SKOTTEN TTTSTIET ENKG TL DRITIN STATE NTGTGDNSNSSGKG KKSSKSDPVE EYLSTGLK SKSTEM DV KUGLGELV NVGTNDNSNSSGKGKLKTKSELVE EYLSTGLK TKN STEM DV KUGLGELV
PyPREX PbPREX PcPREX PkPREX PvPREX PfPREX TaPREX TpPREX	1075 1036 1009 1099 1098 1238 1064 1059	SIN KURSTING SITE DREKYENKNGGDTSGNNDIIKSTOTETIE ITE GREKNGGNTSDINDIIKATOTEIFE SIN KURST STEG

PyPREX PbPREX PcPREX PkPREX PvPREX PfPREX TaPREX TpPREX	1135 1092 1057 1159 1158 1298 1123 1119	SQR.GN N. NYKGDEYN VRN VDE GDNN, IKNNMNN SQR.GN N. UNNKGDEYN VRN VDE GGENNAIKINT N SQR.GN S. VKNTGGEYSNMGS DE GDNNGIK DMNN AENHNGVI GGNPSGAL PVGGPTKNN FNIVS VKESSASGHNIGHSLRGYVSHPGKEDAL AENHNSVNGGGSPSRA PIGGPTKNI LNIVT ARESPVNSHNLGHISSGHGSAPLTGDGQ VIS IGDGOTSSSCMNIKIYSSENNIY NN NN MNKEPQTLLPNDRNDSNSHSYNN FVSTSKSLLKLERIN VPNND DE FKLKY
PyPREX PbPREX PcPREX PkPREX PvPREX PfPREX TaPREX TpPREX	1173 1130 1095 1219 1218 1358 1155 1151	
PyPREX PbPREX PcPREX PkPREX PvPREX PfPREX TaPREX TpPREX	1214 1171 1136 1263 1262 1418 1192 1188	KDKUKI EN Y EEN NIN NITTEI KUKIYI NGI Y KKKIEN Y EEN NIN NITTEI KUKIYI NGI Y KKNTEI EN Y EK G NKMSNK- EVYKOTO SILI NIKIN NITTEI G DKMGSK- EVYKOTO SILI KIN NIN NITT G DKMGSK- EVYKOTO SILI KIN NITT ONN-EITICAFE SKKOTO SILI KIN NILI DI Y Y PLLIDGIIYU SIKKESLKP
PyPREX PbPREX PcPREX PkPREX PvPREX PfPREX TaPREX TpPREX	1274 1231 1172 1322 1321 1477 1232 1228	Exo I G S N N ND G S N N ND G S S N S S S S S S S S S S S S S S S S
PyPREX PbPREX PcPREX PkPREX PvPREX PfPREX TaPREX	1320 1277 1218 1368 1367 1523 1292 1288	T Y N T T Y N T A MH CVD A MH K E A A H K VFH IN RVYG EFE GP VASK LVASRYISC TH S F V T VFH IN RVYG EFF GP VASK LVASRYISC TH S F V T
		Exo II
PyPREX PbPREX PcPREX PkPREX PvPREX TaPREX TpPREX	1380 1337 1278 1428 1427 1583 1352 1348	SIQ TA Q SIQ TA Q SIQ A Q SIQ VA Q SIQ VA Q KSE CK ET KSE CR GT KSE CR GT KSE CR GT KSE K LH SIQ FEELL SIL FV P VI EHLL IN LADIASVILL ATS SIQ VA Q KSE CK ET KSE CR GT KSE CR GT SIQ VA Q SIQ VA SIQ VA Q SIQ VA SIQ VA Q SIQ VA Q SIQ VA SIQ VA SIQ VA SIQ V

Exo III

Polymerase Domain





Figure 3.10. Multiple alignment of the PREX homologues using ClustalX 1.81 with default settings. The numbers in the brackets refer to the amino acid number of the individual sequence. Black shading indicates identical amino acids, grey shading indicates conserved amino acid substitutions. The various motifs are indicated with underlined text. Blue – primase domain , pink – helicase domain; black - hinge domain and green – polymerase domain.

during maturation. The conserved domain of the hinge region might play a role in this cleavage.

The DNA polymerase start site is predicted to begin at amino acid 1389 and stretches to the stop codon at position 2016. Yet again, the region shows a high degree of conservation among all of the homologues. All PREX homologues contain the key residues known to be essential for the function of the DNA polymerase. Figure 3.11 illustrates the PfPREX with its different domains colour-coded.



Figure 3.11. Putative amino acid of PfPREX which is colour coded for the different functional domains. Colour coding :- Red font- signal sequence (cleavage site between aa 19 and 20). Red-Hinge with box of conserved residues between *Theilera* and *Plasmodium*, Yellow background – primase domain, Green background – helicase domain, Pink-Polymerase domain, black words within polymerase domain $-3^{\circ}-5^{\circ}$ exonuclease domain, blue words within polymerase domain- pol A domain, Boxes show conserved signature motifs found within primase, helicase and polymerase domains.

3.8. Phylogenetic relationship between PREX homologues in apicomplexans

From the multiple alignments of the apicomplexan PREX homologues, the phylogenetic relationship between them could be determined. The alignment was used to obtain a phylogenetic tree by various methods. In addition to that, the genetic distances between the PREX amino acid sequences were obtained. The distances and trees were calculated and generated by the MEGA (Molecular Evolutionary Genetic Analysis) software (Kumar et al, 2001). The genetic distances were calculated with Poisson correction for amino acids (Table 3.3). Genetic distance is a measure of the difference between two sequences. The genetic distances obtained ranges from 0.0486 (between P. berghei and P. yoelii) and 0.3817 (between P. vivax and P. falciparum). This value is relatively high if you compare it with a glycolytic protein e.g. aldolase (Cloonan et al, 2001) where the genetic distance between P. vivax and P. falciparum is 0.140. However, its evolutionary rate is comparable to other proteins including adenylsuccinate lyase (ASL) (Kedzierski et al, 2002) or cysteine proteinase (Rosenthal, 1996) where the genetic distances between P. vivax and P. falciparum are 0.4577 and 0.425 respectively. Therefore, based on this result, PfPREX seems to have a rate of evolution similar to ASL.

Phylogenetic trees were constructed using the MEGA software using all 4 available methods <u>Unweighted Paired Group Method with Arithmetric Mean</u> (UPGMA), <u>Neighbour-Joining</u> (NJ), <u>Minimum Evolution</u> (ME) and <u>Maximum</u> Parsimony (MP). All 4 methods gave a similar topology (Figure 3.12). The *Plasmodium* species PREX homologues are grouped into 2 distinct clades. One branch contains the rodent malaria species, *P. berghei*, *P. yoelii* and *P. chabaudi*. It was also



Table 3.3. Genetic distances between the PREX homologues of six *Plasmodium* species. The distances were calculated with Poisson correction for amino acid residues. Abbreviations: *P. berghet* (Pb), *P. yoelii* (Py), *P. chabaudi* (Pc), *P. knowlesi* (Pk), *P. vivax* (Pv) and *P.falciparum* (Pf).

observed that *P. berghei* and *P. yoelii* are closely related with a bootstrap support of 100 (depending on the method used), which is in agreement with previous findings (Killick-Kendrick, 1978). The other branch groups the primate malaria species *P. knowlesi* and the human parasite, *P. vivax* together. This is fully supported by a bootstrap of 100. This is also observed by Kedzierski *et al*'s (2002) study with the ASL sequences. The bootstrap value for the node where *P. falciparum* joins the primate clade in the UPGMA tree is weaker (52-74) as opposed to the 100 bootstrap values for the other nodes but the bootstrap values indicate the relationships are robust.



Figure 3.12. Phylogenetic Trees for PREX homologues. The trees were generated using MEGA 2.1 for Unweighted Pair Group Method with Arithmetic Mean (UPGMA), Neighbour-joining (NJ), Minimum Evolution (ME) and Maximum parsimony (MP) methods. Bootstrap values correspond to 100 random runs. Abbreviations: *P.falciparum* (Pf), *P. berghei* (Pb), *P. chabaudi* (Pc), *P. knowlesi* (Pk), *Theileria annulata* (Ta), *T. parvum* (Tp), *P. vivax* (Pv) and *P. yoelii* (Py).

3.9. Discussion

The sequence analysis carried out so far seems to suggest that the PfPREX is composed of three putative domains, a DNA primase, DNA helicase and DNA polymerase. The C-terminal domain resembles a DNA polymerase domain. The homology search carried out identified the three key motifs characteristic of family A DNA polymerases. Therefore, to prove that indeed the C-terminal domain does encode an active polymerase, functional studies have to be carried out. The N-terminal domain resembles DNA primase and DNA helicase domains, very similar to the DNA primase/helicase complex of the T7 bacteriophage group. Homology searches carried out also identified the key motifs characteristic of both the DNA helicase and DNA primase. Since PfPREX possesses all of these key motifs, it is likely that it might code for functional DNA primase and helicase.

An interesting point is the presence of an apicoplast-targeting sequence located at the extreme N-terminus. Prediction programmes that are specific for locating such sequences in *P. falciparum* have returned positive results indicating that the sequence is involved in apicoplast-targeting. Therefore, PfPREX is most probably targeted to the apicoplast where it is postulated that it may have a role to play in plastid replication.

Homologues of PfPREX can be found in other *Plasmodium* species with very high similarity between the sequences. Furthermore, PfPREX homologues are also found in other apicomplexans ranging from *Theileria*, where the full PREX sequences have been assembled, to *Toxoplasma gondii* (partial sequences only). This seems to suggest that PREX was already present in the common ancestor of apicomplexans.

Based on amino acid homologies among the PREX homologues, we were able to use that information to determine the putative boundaries of the various domains on PfPREX. This information is useful for the design of primers for the cloning of the individual domains. In addition to this, the PREX homologues were used to make genetic distance calculations and phylogenetic analysis. The results obtained were as expected and compared to a study carried out on adenylsuccinate lyase, PREX seems to have a rate of evolution that is similar to ASL, information which can be used to contribute to studies to determine evolutionary history of *Plasmodium* species.

The information obtained from the bioinformatic studies has given insight into the function of the PIPREX and also allows the opportunity to further explore PIPREX. The main aim of this project has been to demonstrate whether PfPREX does encode a functional plastid replicase. The remaining chapters of this thesis describe the functional characterisation of this protein. Chapter IV

Functional analysis and subcellular localisation of theDNA polymerase (PfPREXpol) domain encoded by the *PfPREX* gene

4.1. Introduction

Propagation of a species depends upon the faithful replication of its genome. In order for cells to achieve this, they require an efficient machinery to maintain genome integrity during cell division, DNA repair, DNA recombination and also when confronted with DNA damage. A key enzyme involved in all of these processes is DNA polymerase. Multiple DNA polymerases have been identified in mammals, yeast and prokaryotes. A particular DNA polymerase may have multiple roles to play in a cell and a particular part of the DNA metabolic process may need more than one polymerase. So far 12 classes of DNA polymerase have been classified, α , β , γ , δ , ϵ , ζ , η , θ , ι , κ , λ , μ , (Dominguez et al, 2000, Friedberg et al, 2000, Gibbs et al, 1998, Harris et al, 1996, Hubscher et al, 2000, Masutani et al, 1999, Nagasawa et al, 2000, Nelson et al, 1996, Ohashi et al. 2000, Tissier et al. 2000) (Table 1.2) where each has a different role to play in the DNA metabolic processes. Based on sequence relationships, these enzymes are classified into five families designated A, B, C, X and Y (Burgers et al, 2001). The family A group of polymerases are typified by the *E. coli* PolI; for family B members include the mammalian pol α , δ , ε and *E. coli* pol IV. *E. coli* pol III exemplifies family C and mammalian pol β is the best known member of family X. Finally E. coli pol V represents family Y.

In *Plasmodium berghei*, seven DNA polymerases have been identified using classical biochemical characterisation (De Vries *et al*, 1991). Five of them are aphidicolin-sensitive while two are resistant. Aphidicolin is a natural product derived from the fungus, *Nigrospora oryzae*, and it is widely used to distinguish between prokaryotic and eukaryotic DNA polymerase activities (Spadari *et al*, 1984). The two DNA polymerases, which are aphidicolin-sensitive, were hypothesised to be DNA

polymerase alpha (α) and DNA polymerase gamma (γ). Among those which are aphidicolin-insensitive, two of which are believed to be either a DNA polymerase beta (β) or delta (δ). For *P. falciparum*, at least two distinct DNA polymerase activities were detected (Chavalitshewinkoon *et al*, 1993). The aphidicolin-sensitive fraction was postulated to contain DNA polymerase α activity while the aphidicolin-insensitive fraction might contain a DNA polymerase γ . To date, only the subunits of DNA polymerase alpha and delta have been cloned and characterised from *P. falciparum* (Ridley *et al*, 1991).

The *P. falciparum* genome has been fully sequenced (Gardner *et al*, 2002). This provides the opportunity to explore the database to identify DNA polymerase gene sequences within the *P. falciparum* genome. BLAST searches using the Plasmodb database (Version 4.1) and also a query search of 'polymerase' of the annotated gene database were performed. To be more specific, the query was carried out at the Plasmodb website (www.plasmodb.org), using a GO-function assignment search for function of DNA-directed DNA polymerase (GO_ID 3887). GO stands for gene ontology where it organises different proteins based on 3 principles:- (1) molecular function, (2) biological process and (3) cellular component (Ashburner *et al*, 2000). Table 4.1 shows the results from the query search through the annotated gene database (Plasmodb Version 4.1). We have found at least five different DNA polymerases:- a DNA polymerase delta (δ) catalytic subunit, small subunit to a DNA polymerase alpha (α), a DNA polymerase epsilon (ε), a DNA polymerase zeta (ζ) catalytic subunit and a DNA polymerase I. These DNA polymerases from *P. falciparum* do not possess any apicoplast-targeting sequence.

	gene identifier	location	description	references
1	PFC0340w	pfal_chr3: 350224- 351720	DNA polymerase delta small subunit, putative	Horrocks et al, 1996; Ridley et al, 1991
2	PFD0590c	pfal_chr4: 535089- 541382	DNA polymerase alpha	White et al, 1993
3	MAL6P1.175	chr6: 1071327- 1075661	DNA polymerase 1, putative	•
4	PF10_0165	chr10: 685577- 688861	DNA polymerase delta catalytic subunit	Horrocks et al, 1996; Ridley et al, 1991
5	PF10_0362	chr10: 1464928- 1472132	DNA polymerase zeta catalytic subunit, putative	• •
6	PFL1285c	chr12: 1077858- 1078766	proliferating cell nuclear antigen, putative	Horrocks et al, 1996; Kilbey et al, 1993
7	PF14_0112	chr14: 461400- 467450	POM1, putative	-
8	PF14_0234	chr14: 989361- 993071	DNA-directed DNA polymerase, putative	
9	PF14_0602	chr14: 2571936- 2573555	DNA polymerase alpha subunit, putative	White <i>et al</i> , 1993
10	MAL6P1.125	chr6: 1304902- 1314100	DNA polymerase epsilon, catalytic subunit a, putative	

Table 4.1. GO function search using Plasmodb for DNA-directed DNA polymerases in the *Plasmodium falciparum* database. The table consists of four columns, (1) Gene identifier, (2) Gene location which shows which chromosome the gene is located on, (3) Description which shows the class of DNA polymerase and (4) References from which work on the gene had been reported. Number 7 entry is *PfPREX*.

We also performed BLASTP and TBLASTN searches using the sequence of DNA polymerase gamma (γ) from *Drosophila melanogaster* which encodes a polymerase that replicates mitochondrion DNA (Lewis *et al*, 1996) in many eukaryotes. However the *P. falciparum* genome database did not appear to possess genes with homology to this sequence. Therefore, it seems that *P. falciparum* either does not possess a DNA polymerase gamma (γ) or that the gene is highly disrupted by introns and therefore not identified by classical BLAST analysis.

In addition, no DNA replication protein, i.e. the DNA polymerase involved in the replication of the *Plasmodium* plastid, has been previously isolated. The *PfPREX* sequence that is found on chromosome 14 (Genbank accession number: AAN36724.1), has homology to a prokaryotic DNA polymerase-I like sequences. The same ORF also encodes a typical plastid-targeting domain (Section 3.5). Therefore, we hypothesised that *PfPREX* may be responsible for plastid DNA replication/repair. Since the total *PfPREX* gene is over 6 kb in length and, like other *Plasmodium* genes, it is highly A-T rich, cloning and expression of the entire gene was deemed too difficult to warrant effort. Therefore, an approach was undertaken to clone and express the DNA polymerase domain in isolation from the DNA helicase and primase domains. This chapter is dedicated to reporting the cloning, expression and characterisation of the DNA polymerase domain of the *PfPREX* gene.

4.2. Results

4.2.1. Cloning of the PfPREXpol domain of the PfPREX gene

The DNA polymerase domain of the *PfPREX* gene spans a region of approximately 1,878 bp at the C-terminal domain of the gene. It was amplified using gene-specific primers, designed specifically for the use in the Gateway Cloning system (Invitrogen). The primers used were GWPolF1 as the forward primer and GWPolR1 as the reverse primer (Table 2.1). The forward primer was designed to overlap partially with the hinge domain of the gene. Figure 4.1 shows the relative positions of the primers in accordance to the PfPREXpol domain. The expected size of the PCR product was 2.9 kb (Figure 4.2). The primers used contained the *att* sites to permit of recombination with the Gateway Cloning system that was used.

1.00









The 2.9 kb PCR product was first cloned into the pDONR201 donor vector before it was transferred into the destination vector (or expression vector), pDEST17 that adds an N-terminal His-tag to the protein to facilitate purification of the expressed protein (Figure 4.3).



Figure 4.3. The pDEST17GWpol expression construct. The Gateway PCR product GWpol DNA polymerase domain (GWPolF1 and GWPolR1) (Table 2.1) was cloned into the specific recombination sites of pDEST17 (His-Tag) (Invitrogen) expression vector. The pDEST17 vector contains an ampicillinresistance gene (Amp) and a T7 promoter. Selected restriction enzyme sites are shown, as are the positions of the T7 promoter (T7), the His-tag (His), the ampicillin resistance gene (Amp) and the plasmid origin (ori).

4.2.2 Expression and purification of PfPREXpol

The *E. coli* BL21si strain was used to host the expression of PfPREXpol. Induction of expression of the recombinant protein was performed at 37° C (Section 2.7.1). The calculated mass of the PfPREXpol domain is 107 kDa (Figure 4.4). Since the expressed recombinant protein contains an N-terminal His-tag, it was purified using a nickel-affinity column on a BIO-CAD 700E workstation (PE Biosystems). Cultures were lysed and prepared for the purification of the His-tagged recombinant protein. The purification was carried out as detailed in section 2.7.1. Approximately 10 mg of soluble recombinant protein was obtained from one litre of *E. coli* culture. The purified protein was analysed by SDS-PAGE. The apparent size of the protein on the gel was 90 kDa, lower than its predicted molecular weight.



Figure 4.4. Coomassie blue-stained SDS-polyacrylamide gel of expressed and purified PfPREXpol. (A) Coomassie blue-stained SDS-polyacrylamide gel of uninduced (Lane 1) and NaCl induced cells carrying the cloned *PfPREXpol* gene fragment (lanes 2 and 3). (B) Coomassie blue-stained SDS-polyacrylamide gel of nickel affinity purified protein, PfPREXpol, as indicated by the arrow (lanes 1 and 2).

4.2.3. Functional characterisation of PfPREXpol

To determine whether the PfPREXpol protein has DNA polymerase activity, the expressed and purified recombinant protein was used in DNA polymerase activity assays (Section 2.8.3). In addition, the recombinant protein was also checked for the presence of 3'-5' exonuclease activity (Section 2.8.4) because it was observed to possess key structural motifs for that activity (Section 3.2.3).

4.2.3.1. DNA polymerase activity

DNA polymerases catalyse the template-directed incorporation of dNTPs into DNA by addition at the 3'OH termini of primer strands. Polymerase activity is assayed by measuring the incorporation of radiolabelled dNTPs into DNA. Primed single-stranded M13 DNA is frequently used as the template and is the template of choice for the DNA polymerase assays carried out (Section 2.8.3).

DNA polymerase activity was detected in the purified PfPREXpol protein (Figure 4.5a). Different conditions known to influence DNA polymerase activity were investigated. Control experiments were also carried out. Commercial Klenow (*E. coli* DNA polymerase I) (Promega) was used as a positive control for template elongation while a negative control using just distilled water was used as a blank to determine background counts from radiolabel that non-specifically reached the product collection tube. Another negative control was carried out using the recombinant PfPREXprihcli. This control helped to check if the presence of DNA polymerase activity is specific to the recombinant PfPREXpol and not from contaminating proteins in *E. coli*. Typical radioactivity counts obtained from the polymerase activity assays were as such:- the negative controls counts per minute (cpm) range from 747.17 to 1402.76, positive
controls gave cpm of greater than 30,000 while PfPREXpol samples gave a cpm range from 2209.78 to 26,750.

First the effect of pH on the enzyme activity was measured using buffers calibrated to different pH values between 6.0 and 9.0. The buffers used were 50 mM MOPS-NaOH (pH 6.0 -6.5), 50 mM Tris-HCl (pH 7.0 to 9.0). The enzyme has an optimum catalytic activity at pH 7.0 but is inhibited at acidic pH. At alkaline pH, its activity also decreased drastically (Figure 4.5b). Experiments were also performed to determine the effect of temperature on the polymerase activity between 25°C and 90°C. The optimum activity of the polymerase was observed at a surprising 75°C and decreases thereafter (Figure 4.5c). Interestingly, the PfPREXpol protein has most similarity to a DNA polymerase from the thermophilic bacteria *Aquifex aquaticus* (Table 4.2)

DNA polymerases require divalent cations for activity as the cations are involved in stabilising of the transition state of the enzyme and facilitating phosphoryl transfer (Steitz, 1998). Therefore, the effect of Mg²⁺ concentration on the polymerase was tested. The optimal Mg²⁺ concentration for activity was at 5 mM (Figure 4.5d). The effect of KCl concentration on the polymerase activity was also investigated because when salt concentration is high, it affects the dynamics of protein-DNA complex interaction. This results in electrostatic interactions which can contribute to the stability of the complex. For PfPREXpol, the optimum concentration of KCl for activity is 50 mM. Higher concentrations of KCl were inhibitory to the DNA polymerase activity (Figure 4.5e). Its range of activities is comparable to other recombinant DNA polymerases from other organisms (Table 4.2).

123

Reference	Blöndal et al, 2001	Gaikwad <i>et al</i> , 2002	Huang & Levin, 2001	Huang & Levin, 2001	Seto <i>et al</i> , 1998	Choi et al, 1999	Pons et al, 1991	Chang et al, 2001	
Optimum Mg ²⁺ conc (mM)	5	8	10-20	10-20		4	6.5	4-5	5
pH optima	7	7.5	7.5	7.5	7.0	8.4-9.0	N.A	6.8-7.2	7.0
Temperature optima	55°C	N.A	N.A	N.A	37°C	70-72.5°C	37°C	70°C	75°C
DNA Polymerase Source	Rhodothermus marinus	Pisura sativura	E. coli	Spodoptera littoralis nucleopolyhedrovirus	Cauliflower inflorescence	Thermus filiformis	Sireptococcus pnemoniae	Aquifex aeolicus	PfPREXpol P. falciparum

Table 4.2. Characteristics of other DNA polymerase I-like proteins from other sources of organisms. N.A stands for not available.

.

. .

1

. .

mins









(A)

Figure 4.5. Characterisation of DNA polymerase activity of PfPREXpol and the effects of different conditions on its activity. The activity is expressed as relative activity (100%) with respect to different conditions. The experiments were carried out in triplicate over three individual experiments (n=9) and the average was taken to draw the graphs using the Prism software (Graphpad) (a) Time course experiment with PfPREXpol, (b) pH profile of PfPREXpol (c) Temperature profile of PfPREXpol, (d) Effect of Mg²⁺ ions on PfPREXpol activity and (e) Effect of KCl concentration of PfPREXpol activity. No error bars are shown due to the wide range of CPM obtained from each experiment, instead the results are represented as mean relative activity.

Aphidicolin, an inhibitor of most eukaryotic DNA polymerases was also tested for its effect on the recombinant polymerase. Aphidicolin was added to the reaction mixture at various concentrations up to 160 mg/ml but no inhibition was seen, indicating that the recombinant DNA polymerase is not sensitive to aphidicolin. A control experiment was also carried out to test aphidicolin on *E. coli* Klenow (Promega) and it is also insensitive to aphidicolin. Other prokaryotic DNA polymerases and the mitochondrial DNA polymerase of eukaryotes are also aphidicolin-insensitive while most types of eukaryotic DNA polymerases are aphidicolin-sensitive. Unfortunately, we were not able to obtain any eukaryotic DNA polymerases to use as a control to verify whether the batch of aphidicolin used here was functional.

4.2.3.2. 3'-5' exonuclease activity

The domain that has been classified as the DNA polymerase domain of PfPREX (PfPREXpol) also has significant homology towards its N-terminus with 3'-5' exonucleases. The three key structural motifs common among these enzymes can all be found within PfPREXpol. The 3'-5' exonuclease activity of the PfPREXpol was investigated (Section 2.8.4). 3'-5' exonuclease activity was detected, however it was slow-acting over a period of time. Our time course experiments were conducted for up to 2 hours (Figure 4.6), during which the activity of the 3'-5' exonuclease did not reach a plateau. This could be due to the non-optimal conditions used for the assay. For example, the temperature optimum obtained for DNA polymerase activity was at 75°C but the temperature used for this assay was 37°C. Its pH profile was also investigated but further experiments established that there was a spontaneous degradation of the double stranded fragments as the pH of the reaction buffer was increased. Therefore, no further characterisation of the 3'-5' exonuclease was carried out.

127



Figure 4.6. Time course experiment of 3'-5' exonuclease domain of PfPREXpol. The activity is expressed as relative activity (100%). The experiments were carried out in triplicate and the average was taken to draw the graphs using Prism software (Graphpad). No error bars are shown because of the large range of CPM obtained from each experiment, instead the results are represented as mean relative activity.

4.2.4. Expression of PfPREXpol in P. falciparum

RT-PCR and Western blotting were used to investigate the expression pattern of the *PfPREX* gene and its product in *P. falciparum*. The PfPREX protein is predicted to be 235.8 kDa in its entirety. Whether the PfPREX is indeed translated as a single polypeptide has implication for the topological aspects of the molecule. Production of such a large protein could pose a problem when it comes to ensuring the proper folding and function of the final molecule.

To investigate the transcription of the *PfPREXpol* gene in *P. falciparum*, **RT-PCR** was performed. Northern blotting using total RNA from asynchronous culture

failed to reveal a signal. Therefore, total RNA from an asynchronous culture of *P. falciparum* was obtained and RT-PCR primers for the PfPREXpol region (Table 2.1) were used (Figure 4.1) to amplify cDNA obtained from *P. falciparum* (Section 2.4.9.2). The results of the RT-PCR (Figure 4.7) showed a band of the predicted size of 400 bp. From the result obtained, it appears that the *PfPREXpol* domain is transcribed. Controls using no reverse transcriptase confirmed that there was no genomic DNA contamination in the RT-PCR samples.



Figure 4.7. RT-PCR amplification of the *PfPREX* transcript from *P. falciparum* 3D7 total RNA. PCR was carried out for 40 cycles; + and – indicate reactions wih and without reverse transcriptase, respectively as a control to check for genomic DNA contamination. PCR amplification was carried out on total RNA obtained from *P. falciparum* 3D7, using primers specific for DNA polymerase region RTPolF1 and RTPolR1. The expected size of the band is approximately 400 bp.

Western blots were also carried to determine whether the PfPREXpol is translated in P. falciparum. The recombinant purified PfPREXpol (Section 2.7.1) was used to generate anti-serum made in rabbits (Section 2.7.3). The anti-serum was used to probe blots of parasitised rbc lysates and non-parasitised samples. All Western blots were performed at least three times to check for reproducibility of results. Since PfPREX comprises a single ORF which translates to a predicted protein size of approximately 2016 amino acids, the size of the PfPREX protein was estimated at 235.8 kDa. Anti-sera to PfPREXpol, however, identified a single band of approximately 91 kDa in infected but not uninfected rbc on the Western blot (Figure 4.8). The Western blot was repeated at least three times which showed that the result obtained was reproducible. Thus, this result suggests that the PfPREXpol is cleaved from the precursor PfPREX protein. The size of the identified band coincides with just the polymerase domain. It may be that the PREX protein is transported to the apicoplast, and then the C-terminal end (PfPREXpol domain) is cleaved post-translationally. Alternatively, cleavage could occur prior to apicoplast entry, or differential splicing could occur. However, it has not been possible to distinguish these different scenarios during the course of this thesis.



Figure 4.8. Western Blot of parasitised rbc (3D7) and non-parasitised rbc lysates probed with anti-PfPREXpol (1:2000) anti-sera. A single band was observed in the 3D7 lane of approximate size of 91 kDa.

4.2.5. Homology modelling of PfPREXpol

Homology modelling was carried with the PfPREXpol amino acid sequence. Due to the similarity of PfPREXpol to the Klenow fragment, we were able to send the PfPREXpol amino acid sequence to the Swiss Model First Approach (Schwede *et al*, 2003) website (http://swissmodel.expasy.org) to obtain a putative 3D structure of PfPREXpol based on the known crystal structure of Klenow fragment of *E. coli* (PDB no:- 1KFD) (Beese *et al*, 1993).

4.2.5.1 A modelled 3D structure of PfPREXpol protein

Like the Klenow fragment, the engineered protein representing the DNA polymerase I, PfPREXpol, consists of a 3'-5' exonuclease domain at its N-terminal end and the DNA polymerase domain at its C-terminal end. PfPREXpol shares 38.1%

similarity and 24.5% identity with the *E. coli* Klenow fragment which shows that they share a high degree of conservation. The alignment of PfPREXpol alongside the *E.* coli Klenow fragment (Figure 3.5) shows the degree of conservation especially in the key motifs involved in both 3'-5' exonuclease activity and DNA polymerase activity. The Swiss Model service of the Expasy server (Guex & Peitsch, 1997; Ollis *et al*, 1985; Peitsch, 1996) was used.

The modelling result suggests that there is a high degree of conservation of structure between PfPREXpol and *E. coli* Klenow fragment. The key thumb-finger-palm structure, characteristic of the Klenow fragment, is also observed in the predicted PfPREXpol 3D structure alongside the 3'-5' exonuclease domain (Figure 4.9).



Figure 4.9. 3D structure of Klenow fragment from *E. coli* and PfPREXpol which was modelled on the Klenow fragment. The two structures look strikingly similar consisting of key characteristic features of DNA polymerase I. It consists of a thumb-finger structure and a 3'-5' exonuclease domain.

4.2.5.2. Comparison of PfPREXpol and E. coli Klenow fragment 3D structures

Figure 4.9 shows the putative 3D structure of PfPREXpol and the solved 3D structure of *E. coli* Klenow. As in Klenow polI, the structure of PfPREXpol consists of three domains (the thumb, the palm and the fingers). When the amino acid sequence entered into the Swissmodel First Approach, it only contained the predicted PfPREXpol domain. Only a partial 3D structure was obtained. Only when the hinge domain is included in the amino acid sequence used for prediction of 3D structure, was a full 3D structure of PfPREXpol was obtained. Therefore, we hypothesised that the hinge domain of PfPREXpol may have a role to play to ensure the proper folding of the DNA polymerase domain.

In addition, it was considered of interest to compare the amino acid residues that are shown to be important in the various aspects of DNA polymerase activity in Klenow fragment with amino acids occupying the same positions in the modelled structure of PfPREXpol. Previous experiments, which dealt with dNTP binding to the polymerase site, revealed that incoming dNTPs interact with Arg 754, Arg 682 and Lys 758 (Basu & Modak, 1987). These residues are highly conserved in the family A of DNA polymerases and interact with the phosphate backbone of incoming dNTPs as these residues are all positively charged. When the same positions were checked on PfPREXpol, they were found to comprise the same residues. The corresponding residues are conserved between the *E* coli and *P. falciparum* except for Arg 682 in Klenow is replaced conservatively as a Lys 205 in PfPREXpol. Essentially in this case, the positive charge remains in place for interaction with dNTPs. Figure 4.10 shows the storcodiagrams with these sites highlighted. The sites in both enzymes are in similar positions for the interaction with dNTPs.







(C) Residues involved in primer attachment





Figure 4.11. Comparison of the structures of Klenow and PfPREXpol. Superimposed stereodiagrams of the structures of both enzymes. Klenow is shown with the blue line while PfPREXpol is shown with the red line.

Previous experiments and X-ray crystallography data have shown that Tyr 766, Arg 841, Asn 845, Asp 882, Asp 705, Głu 883 and His 881 (Joyce & Steitz, 1994; Pandey *et al*, 1987; Polesky *et al*, 1990) are all involved in binding and directing the incoming dNTP. The corresponding residues in PfPREXpol are also well conserved (Table 4.3). Asp 882 (Asp 482) and Asp 705 (Asp 301) correspond to the residues in motif C and A respectively which are shown to be involved in dNTP binding. In addition, they have been shown be involved in the binding of divalent metal cations (Joyce & Steitz, 1994). Other residues which are required for primer attachment and catalysis are observed to be well conserved between Klenow and PfPREXpol. The palm domain contains the catalytic centre where the residues important for catalysis are Asp 882 (Asp 470), Arg 668 (Lys 263) and Gln 849(Gln 439) (Polesky *et al*, 1990). The thumb domain contains highly conserved residues that are involved in primer attachment. These residues are well conserved across the PolI family. Arg 631 (Arg 228), Lys 635 (Arg 235), Asn 675 (Asn 271) and Asn 678 (Asn 275). Table 4.3 shows all the conserved residues important in DNA polymerase activity.

Sites Involvement	Klenow residues	PfPREXpol residues	References
dNTP interaction	Arg 754	Arg 350	Basu & Modak, 1987
	Arg 682	*Lys 250	Pandey et al, 1990
	Lys 758	Lys 355	Beese et al, 1993a
dNTP binding	Tyr 766	Tyr 362	Pandey <i>et al</i> , 1987
	Asp 882	Asp 470	Joyce & Steitz, 1997
	Asp 705	Asp 301	Beese et al, 1993a
	Glu 883	Glu 471	
	His 881	His 469	
	Arg 841	*Lys 432	
	Asn 845	Asn 435	
Catalysis	Gln 849	Gln 439	Polesky et al, 1990
	Arg 668	*Lys 263	
	Asp 882	Asp 470	
primer attachment	Glu 611	Glu 208	Beese et al, 1993b
	Asn 678	Asn 275	
	Asn 675	Asn 271	
	Lys 635	*Arg 232	
	Arg 631	Arg 228	
	1		

Table 4.3. Comparison of key residues on Klenow and PfPREXpol. The key residues involve in dNTP interaction, dNTP binding, primer attachment, catalysis and template attachment for both Klenow and PfPREXpol. Asterisk next to the residue represents a conservative replacement of amino acids.

4.2.6. Subcellular localisation of the *PfPREXpol* gene product

With reference to the presence of the plastid-targeting sequence in the Nterminus of PfPREX, we wanted to confirm the subcellular localisation of PfPREXpol domain. To do this, we used the PfPREXpol anti-serum and performed some indirect immunofluorescence (IDF) and some *in situ* hybridisation using a plastid specific DNA probe to give us co-localisation data. This first approach was taken to use a plastidspecific DNA probe to target the plastid DNA in *P. falciparum* using *in situ* hybridisation. PCR primers specific for the *ssurRNA*, a gene found on the plastid DNA in *P. falciparum*, were used to generate a PCR product (Section 2.9.2.1). The PCR product had biotinylated dUTP incorporated which will bind to Streptavidin-Texas red for fluorescent detection under phase contrast microscopy. The protocol was in addition to IDF which was carried out after the *in situ* hybridisation step. The protocol could not be optimised to give satisfactory results. No red fluorescence could be observed, apparently the DNA probe was either not binding to the target DNA as expected or it was being washed off during the subsequent steps carried out in the protocol.

Therefore, another approach was taken. Instead of targeting the plastid DNA, we used a stain, Mitotracker[®] Red CMXRos (Molecular Probes) that is taken up by the mitochondrion. This was carried out to check if PfPREX is localised to the mitochondrion. The stain was used alongside DAPI (Sigma) which stains the nucleus and immunofluorescence using the anti-PfPREXpol anti-sera. To optimise the IDF protocol, various dilutions of the anti-sera were used, ranging from 1:50 to 1:1000. The optimum dilution of the anti-sera to be used was 1:500 alongside 1:400 dilution of the secondary antibody (anti-rabbit FITC, Diagnostics Scotland) (Figure 4.13). From the figure, there appears to be no overlay between the mitotracker and the FITC (anti-

PfPREXpol). This suggested that the PfPREXpol did not localise to the mitochondrion in *P. falciparum*. However, it was not clear whether the anti-sera localised to the plastid. The DAPI stains both nuclear and plastid DNA, but it is difficult to see the plastid under the highest available magnification. Therefore, another approach was undertaken in order to show whether PfPREX localised to the plastid. A collaboration was set up with Iain Wilson and Shigeharu Sato (NIMR, MRC) to make green fluorescent protein (GFP) constructs using the PfPREX leader sequence to check for targeting. This work is beyond the context of the thesis itself, since these localisation experiments were the work of Dr Sato. However, the results of Dr Sato's work will be discussed briefly in chapter 7.

÷



Figure 4.12. Subcellular localisation of PfPREXpol using immunofluorescence technique. The first diagram is a phase contrast microscopy picture of a red blood cell infected with a ring-form parasite, followed by the same cell viewed under fluorescence using the Texas red filter of an excitation and emission peak of 560 and 630 nm respectively. The three picture shows the same cell viewed under fluorescence using the FITC filter of an excitation and emission peak of 484 and 510 nm respectively, followed by DAPI filter of 360 and 460 nm respectively and finally with all pictures merged together. The mitotracker stain shows the locality of the mitochondrion DNA, FITC shows where the anti-PfPREXpol antisera have bound and DAPI stains specifically nuclear and apicoplast DNA. Scale bar, 5 µm.

4.3. Discussion

The experimental findings and sequence analysis suggested that PfPREX possesses a DNA polymerase functional domain. The C-terminal domain resembles a DNA polymerase domain (Chapter 3) homologous to family A DNA polymerases. Homology searches identified the three key structural motifs conserved among family A DNA polymerases. Furthermore, the presence of the three 3'-5' exonuclease motifs also suggested the presence of that enzymatic activity. Therefore, attempts were made to clone and express the PfPREXpol domain. This was successful, as purified PfPREXpol was obtained. Functional characterisation was carried out to show that the PfPREXpol is a functional DNA polymerase. It has characteristics similar to other family A DNA polymerases (Table 4.2) including its Mg²⁺ requirements and pH optimum. One interesting point to note is that its temperature optimum was high (75°C). The temperature optimum of PfPREXpol is similar to thermophilic Aquifex aquaticus DNA polymerase at 70°C. It is noteworthy that the enzyme closest in sequence similarity to PfPREXpol is also the same DNA polymerase of Aquifex aquaticus. This raises the possibility that the plasmodial DNA polymerase may be a recent acquisition through horizontal gene transfer from a thermophilic bacterium closely related to the Aquifex family.

Transcription of the PfPREXpol domain was also shown in *P. falciparum*, although failure to identify the transcript by Northern blot analysis indicates it is in very low abundance. Recent transcriptome analysis (Le Roch *et al*, 2003) has shown *PfPREX* to be expressed predominantly in replicative time points of the life cycle in mammalian blood. In addition, experimental findings with the Western blots carried out, suggests that PfPREX may be post-translationally cleaved. The PfPREXpol domain was detected

as a 91 kDa band on a Western blot of parasite lysates probed with anti-PfPREXpol anti-sera, rather than a larger band of 235.8 kDa, which is the size of the full polypeptide. The size of the band 91 kDa, coincides approximately with the predicted size of the PfPREXpol domain. This result seems to suggest that the PfPREXpol domain is cleaved. Topologically, it may be necessary to cleave from the full polypeptide to ensure proper folding and for function.

Homology modelling of PfPREXpol structure was carried out based on its similarity to the Klenow fragment of *E. coli*. The 3D structure obtained from PfPREXpol had a similar structure to the Klenow fragment of *E. coli* which is composed of 3 regions (thumb, finger and palm). We were able to compare the key residues and their positions within the structures of Klenow and PfPREXpol. The results obtained suggests that PfPREXpol is indeed similar to Klenow in terms of its structural composition.

Unfortunately, the results obtained from the subcellular localisation studies here were inconclusive. The *in situ* hybridisation experiments were designed such that the biotinylated plastid-specific DNA probe would bind to the plastid DNA which in turn binds to streptavidin Texas red, so that it could be used to give some co-localisation data. If the anti-PfPREXpol antisera localised to the plastid, the two fluorescence signals would overlay on each other. It seemed though that the DNA probe was not binding to the plastid. Instead, when viewed under texas-red filter under confocal microscopy, no red fluorescence was observed. It could be due to the fault of the DNA probe that it is not a good probe or that the protocol used had too many washing steps such that the DNA probe had been washed away. Due to time constraints, the protocol could not be fine-tuned and new probes were not made due to the time limitations in efforts to get better results.

Based on the results obtained with the immunofluorescence experiments, we cannot conclude on the localisation of PfPREXpol. Careful analysis of the FITC staining pattern obtained with the anti-PfPREXpol (Figure 4.12) suggested a staining pattern typical of compartmental localisation. The green fluorescence was found in spots around the infected red blood cell. If PfPREXpol is located in the cytoplasm, a uniform green fluorescence pattern would be expected. The pattern observed has a spot-like characteristic. Dr Sato (NIMR, Mill Hill) then used GFP technology to help find out the exact location of PfPREX and concluded a plastid localisation (discussed later in Chapter 7).

147

Chapter V

Functional analysis and subcellular localisation of the DNA primase/helicase (PfPREXpriheli) domain of the *PfPREX* gene

ŝ

5.1. Introduction

DNA helicases are ubiquitous in nature. Multiple DNA helicases, each with different functions, are present within a single organism. All DNA helicases are NTPases and they share common biochemical properties including nucleic acid binding, NTP binding and NTP hydrolysis. Common amino acid motifs can identify putative DNA helicases. We have detected the presence of multiple motifs that are known to be important for DNA helicase function in PfPREX suggesting that *PfPREX* may encode a functional DNA helicase. In addition to this DNA helicase, *Plasmodium* may also possess other helicases (*E. coli* has at least 11 DNA helicases). Therefore, a GO function search, similar to the one carried out for DNA polymerase in chapter 4, was carried out to search for other helicases in *P. falciparum*. Table 5.1 shows the DNA helicases that have been detected in the *P. falciparum* genome database. At least nine putative helicase genes have been annotated in the *P. falciparum* genome but none of these correspond to the DNA helicase on PIPREX.

DNA primases are often associated with DNA helicases (Ilyina *et al*, 1992). This may be the case with PfPREX, which also contains a region with similarity to a DNA primase. DNA primases are responsible for synthesising short RNA strands for DNA polymerase to use to elongate DNA during DNA replication. DNA primases also possess key motifs required for activity; these include a zinc motif, and two DxD (aspartate) dyad motifs that can be located in PfPREX. In addition, three other motifs important for DNA primase activity were detected on PfPREX which suggest that it may possess a functional DNA primase activity. A GO function search was also carried out to check for the presence of other DNA primases in *P. falciparum* (Table 5.1). Only

two other DNA primases are found, neither of which correspond to the DNA primase identified on PfPREX.

	gene	location	description		
1	PFB0730w	pfal_chr2: 658636-664629	DNA helicase, putative		
2	A Martin	chr6: 248857-254350	DNA helicase, putative		
3	PF10910w	pfal_chr9: 764838-767024	DNA helicase, putative		
4	MAN THE T	chr13_1: 1690520- 1695332	DNA helicase, putative		
5	PF13_0308	chr13_1: 2189979- 2192699	DNA helicase		
6	The Assessor	chr13_1: 2386358- 2387809	ATP-dependent DNA helicase, putative		
7	PF14_0081	chr14: 313974-317456	DNA repair helicase, putative		
8	CONTRACTOR IN	chr14: 1175723-1180045	ATP-dependent DNA helicase, putative		
9	PF08_0100	chr8: 841194-843300	ruvB like DNA helicase, putative		
10	PF10530c	pfal_chr9: 496141-497718	DNA primase, large subunit, putative		
11	Partie Coble	chr14: 1563530-1566794	small subunit DNA primase		

Table 5.1. Putative genes found using a GO function search through the *Plasmodium falciparum* genome using DNA helicase and DNA primase function searches in Plasmodb version 4.1. Genes 1 to 9 shows the putative DNA helicases found and genes 10 and 11 shows the putative DNA primases found through GO function search.

The GO function search for both DNA helicase and DNA primase failed to detect the putative DNA helicase or DNA primase of PfPREX. This could be due to the nature of algorithms of the search programme which were not able to pick up the key amino acid motifs found on each of them.

In this chapter, the cloning and expression of the DNA helicase and DNA primase domains together is described. The two domains were chosen to be cloned and expressed as a single entity because they showed most similarity to the phage T7 gene 4 which is a combined DNA primase-DNA helicase. Therefore, the possibility exists that the DNA primase and helicase need to be together to fold or function correctly.

5.2. Results

5.2.1. Cloning of the *PfPREXpriheli* domain of the *PfPREX* gene

The DNA primase/helicase domain of the *PfPREX* gene spans a region of approximately 2,721 bp over the N-terminal domain of the gene. It was amplified using gene-specific primers, designed specifically for use in the Gateway Cloning system (Invitrogen). The primers used were GWHeliSignalF1 as the forward primer and GWHeliHingeR1 as the reverse primer (Table 2.1). The reverse primer was designed to overlap partially with the hinge domain (~ 400 bp) of the gene while the forward primer precedes the primase domain at the extreme N-terminus within the predicted plastid-targeting domain. Figure 5.1 shows the relative positions of the primers in accordance to the PfPREXpriheli domain. Therefore, the expected size of the PCR product was 3.2 kb (Figure 5.2). The primers used contained the *att* sites for ease of recombination with the Gateway Cloning system (Invitrogen) that was used for expression (Section 2.5.1).



Figure 5.1. Schematic diagram showing the location of the primers used on the gene *PfPREX*. The black arrows show the primers used in the PCR amplification of the fragments used for cloning. The red arrows show the primers used for RT-PCR experiments.

152



Figure 5.2. Cloning of recombinant His₆-tagged PfPREXpriheli DNA helicase/primase. The PCR amplification product was analysed on a 0.8% agarose gel.

In addition, attempts to clone and express fragments of the DNA helicase, DNA primase and the hinge domains as fusion proteins, separately for the purpose of generating anti-sera for localisation studies were carried out. The primers used for all three domains are listed in Table 2.1. The expected sizes of the three PCR products from the DNA primase, helicase and hinge domains were 500 bp, 420 bp and 600 bp respectively (Figure 5.3).



Figure 5.3. PCR amplification products of the fragments from the DNA primase, DNA helicase and hinge domains of *PfPREX*. Fragments of the three domains were amplified using gene-specific primers as detailed in Table 2.1. The PCR products were analysed on a 0.8% agarose gel. The PCR product of the helicase domain has an expected size of 500 bp, the primase domain - 450 bp and the hinge domain - 600 bp.

The PCR products obtained were then cloned into pDONR201 donor vector before they were transferred into destination vectors (expression vectors). The PfPREXpriheli PCR product was transferred into a pDEST15 vector which added an Nterminal GST-tag to the protein to facilitate purification. The other three PCR products were also subcloned into pDEST15.

5.2.2. Expression and purification of DNA primase/helicase and hinge recombinant proteins

The *E. coli* BL21si strain was used to host the expression of the recombinant proteins. Induction of expression of the recombinant proteins was performed at 16°C overnight and soluble proteins were obtained from *E. coli*. The calculated mass of the PfPREXpriheli recombinant protein is approximately 120 kDa. Since the expressed recombinant protein contains an N-terminal GST-tag, it was purified as detailed in Section 2.7.2 in chapter 2. A significantly lower quantity of recombinant protein was obtained with the PfPREXpriheli than with the PfPREXpol protein. The purified protein was analysed on an SDS-PAGE but it ran on the gel as a band at 97 kDa (Figure 5.4), significantly lower than its predicted molecular weight. This could be due to the basic nature of the protein where its predicted pI is 9.13. The three GST-fusion proteins analysed on SDS-PAGE, ran at their expected sizes: - primase domain (22 kDa), helicase domain (20 kDa) and hinge domain (23kDa) (Figure 5.5).







Figure 5.5. Coomassie Blue-stained SDS-polyacrylamide gel. (A) pDEST15pri, the recombinant primase protein. Lanes (1) extract from induced cells, (2) flow-through fraction from GST purification column and (3) purified protein; (B) pDEST15Hinge, the recombinant hinge protein. Lanes (M) marker protein, (1) extract from uninduced cells, (2) extract from induced cells and (3) purified hinge protein; (C) pDEST15Heli, the recombinant helicase protein. Lanes (1) total induced cell lysate and (2) purified protein.

5.2.3. Functional characterisation of PfPREXpriheli

To determine whether the PfPREXpriheli protein has DNA primase and DNA helicase activities, the expressed, purified recombinant protein was used in activity assays (Section 2.8).

5.2.3.1. DNA primase activity

DNA primase activity was detected in the purified PfPREXpriheli protein (Figure 5.6). Different conditions that can affect the DNA primase activity were investigated (Section 2.8.1). In addition, control experiments were carried out. First the purified recombinant PfPREXpol was used as a negative control in both the DNA primase and DNA helicase assays. This was to check for the presence of contaminant *E. coli* proteins that possess either DNA primase or DNA helicase activity so that if either activity was observed, it was due to PfPREXpriheli and not an *E. coli* contaminant

protein. Typical radioactivity counts obtained from the DNA primase activity assays were as such :- the negative controls gave average cpm of 723.73 and PfPREXpri samples gave cpm ranging from 1637.26 to 26,206.53.

Different concentrations of PfPREXpriheli were added to the enzyme assays to see the effect of increased quantity of enzyme on the activity. There was a proportional increase of DNA primase activity with increase in the quantity of enzyme added. Next, a time course experiment with the recombinant protein was carried out. The rate of activity stayed constant, yielding increasing quantities of primer-product over time. The effect of Mg^{2+} on DNA primase activity was also studied. The enzyme activity was at its optimum when the Mg^{2+} concentration was between 4.0 and 6.0 mM (Figure 5.6c). Finally, the effect of pH on primase activity was studied. It appears that the enzyme is most active at alkaline pH (pH 7.5-9.0). Unfortunately, time constraints and the limiting quantity of protein prevented us from trying higher pH values to see if the primase activity is inhibited by more alkaline conditions. Moreover, the other parameters were measured at pH 7.5, which is apparently suboptimal for activity. Notwithstanding this, the activity is clearly present, dose dependent and also dependent on Mg^{2+} concentration and pH.



Figure 5.6 Characterisation of the DNA primase activity of PfPREXpriheli and the effects of different conditions on its activity. The activity is expressed as relative activity with respect to different conditions. The experiments were carried out in triplicate (n=9) and the average was taken to draw the graphs using Prism software (Graphpad). The experiment was carried to checked for DNA primase activity based on different conditions: - (A) increasing quantities (μ g) of PfPREXpriheli, (B) time course experiment, (C) effect of Mg²⁺ ions on DNA primase activity and (D) pH profile of PfPREXpriheli.

5.2.3.2. DNA helicase activity

Using the SPA helicase assay (Amersham), we were able to determine the presence of unwinding activity in the recombinant PfPREXpriheli. Control experiments were carried out. The purified PfPREXpol recombinant protein was used as a negative control to check for any contaminating *E. coli* proteins and to confirm that any presence of unwinding activity was due to the the recombinant PfPREXpriheli. In addition, *E. coli* DnaB helicase, a gift from E Biswas (University of Medicine & Dentistry of New Jersey) was used as a positive control. Very little unwinding activity was observed in either the positive control or PfPREXpriheli. Therefore, a second positive control was
obtained to check if the problem was with the kit or our enzymes. The new positive control was PerA (a DNA helicase from *Bacillus stearothermophilus*) (Cambio Science). Typical radioactivity counts obtained from the DNA helicase assays were as such :- the negative controls average cpm were 312.9, PerA positive controls average cpm were 2801.14 and PfPREXheli samples cpm range from 459.3 to 1081.3.

First, different amounts of the recombinant enzymes were added to the assays which yielded an increase in unwinding activity (Figure 5.7a). This was followed by a time course experiment where we saw an initial time lag in the rate of unwinding activity. The rate of unwinding activity increased after one hour of incubation at 33°C and thereafter (Figure 5.7b). Other conditions that affected the unwinding activity of PfPREXpriheli were investigated too. The Mg²⁺ concentration that was optimum for its unwinding activity was found to be between 2.0 to 3.0 mM (Figure 5.8a) while the optimal temperature is at 37°C (Figure 5.8b). Finally, the pH at which the enzyme activity is optimal at pH 6.8 (Figure 5.8c).



Figure 5.7. Characterisation of DNA helicase activity of PIPREXpriheli and the effects of enzyme quantities on its activity and its time course. The activity is expressed as relative unwinding. The experiments were carried out in triplicate over three individual experiments (n=6) and the average was taken to draw the graphs using Prism software (Graphpad). The experiment was carried:- (A) Increased quantities of PfPREXpriheli, (B) Time course experiment.



Figure 5.8. Characterisation of DNA helicase activity of PfPREXpriheli and the effects of different conditions on its activity. The activity is expressed as relative unwinding with respect to different conditions. The experiments were carried out in triplicate over three individual experiments (n=6) and the average was taken to draw the graphs using Prism software (Graphpad). The following conditions were tested against the activity of PfPREXpriheli: - (A) effect of Mg^{2+} ions on DNA helicase activity, (B) temperature profile of PfPREXpriheli and (C) the pH profile of PfPREXpriheli.

ŝ

5.2.3.2.1. Nucleotide Hydrolysis Assay

A nucleotide hydrolysis assay was carried out to analyse the NTPase activity of the enzyme (Section 2.8.5). All helicases are NTPases as they require the hydrolysis of NTPs to energise the unwinding of DNA. Whether the helicase displays a preference for the use of a particular NTP was analysed, as it has been previously shown that many helicases have preferences for the NTP that they hydrolyse. For example, T7 DNA helicase prefers TTP and can utilise all NTPs except CTP while T4 DNA helicase can use all NTPs and dNTPs (Matson & Kaiser-Roger, 1990). The results obtained indicated seem to suggest that PfPREXpriheli has no preference of nucleotide used for hydrolysis (Figure 5.9). However, an important negative control with another recombinant protein to check for the presence of a contaminanting NTPase was not carried out during the course of the experiment. Therefore, a full interpretation of the data could not be carried out.



Figure 5.9. Nucleotide hydrolysis assay of PfPREXpriheli. Different NTPs and dNTPS were used as the substrate to check for NTP preference of PfPREXpriheli. The experiment was carried out in triplicate and the averages \pm standard errors of four individual experiments (n=12) were used to draw the graph using Prism (Graphpad) software. The change in absorbance was measured as an indication of the release of the phosphate by the absorption of a 'molybdate blue' complex. The more phosphate is released the more complex is formed which results in the increase in absorbance at A₆₉₀.

5.2.4. Expression of PfPREXpriheli in *P. falciparum*

To determine whether *PfPREXpriheli* in *P. falciparum* is transcribed, RT-PCR was performed. Total RNA from an asynchronous culture of *P. falciparum* was obtained and RT-PCR primers designed in the region of PfPREXpriheli were used (Figure 5.2). Two sets of primers were used. One set (RTHeliF2 and RTHeliR2) of primers was used to check if the DNA helicase domain was transcribed and the other set (RTHingeF1 and RTHingeR1) was used to check for the hinge domain. The results of the RT-PCR (Figure 5.10a) showed bands of the predicted sizes of 453 bp and 420 bp respectively. Thus, we can conclude that both the helicase and hinge domains are transcribed. In addition, we had controls with no reverse transcriptase added to confirm that we have no genomic DNA contamination in the RNA samples.

However, these results are unable to tell us if the whole *PfPREX* is transcribed in a single transcript. Therefore, different combinations of RT-PCR primers were used to investigate this, where overlapping amplicons would suggest that *PfPREX* is transcribed as a single transcript. RTHeliF1 (within the primase region) was used in combination with RTHingeR1 which covers the whole of the DNA primase and helicase domain and into the hinge. Another set of primers used was RTHeliF2 and RTPolR1. This set of primers covers the region that overlaps the hinge region of the previous primer set and into the polymerase region. The results of the RT-PCR (Figure 5.10b) showed bands of the predicted sizes of 3.9 kb and 3.2 kb. The two bands had a region of 360 bp that overlapped (hinge domain). This suggests that *PfPREX* is transcribed as a single transcript. However, these results do not indicate the size of the



Figure 5.10. RT-PCR amplification of the *PfPREX* transcript from *P. falciparum* 3D7 total RNA. PCR was carried out for 40 cycles; + and - indicate reactions with and without reverse transcriptase, respectively as a control to check for genomic DNA contamination. PCR amplification using (A) primers specific for the DNA helicase region (Lane 1, 2), for the hinge region (Lane 3, 4). (B) primers specific for the region between primase and hinge region (Lane 1) and between the hinge and polymerase region (Lane 2). PCR products encompassing the primase and hinge and the polymerase and hinge overlap, indicating that the *PfPREX* gene is transcribed as a whole fragment.

PfPREX transcript and multiple attempts to get results from Northern Blots were unsuccessful indicating that the stable transcript is not abundant and is below a level detectable by Northern blot analysis in asexual stages of the life-cycle. Data from elsewhere indicates that in *P. chabaudi PfPREX* is transcribed as a 7 Kb mRNA in asexual bloodstream form stages (C. Janssen, personal communication).

Western blots were also performed to yield more information about whether the PfPREXpriheli is translated in P. falciparum. Antisera were made against the DNA helicase, primase and hinge domains separately as fusion proteins (Section 2.7.3). The antisera were used to probe blots of parasitised red blood cell lysates and nonparasitised samples. All Western blots were performed at least three times to check for reproducibility of results. Since PfPREX represents a single ORF encoding a predicted protein of approximately 2016 amino acids, the estimated size of the PfPREX protein is 235.8 kDa. If PfPREX is translated as a whole polypeptide, a single band of approximately 235.8 kDa would be expected on the blot using any of the antisera. Instead, the anti-PfPREXpol antisera, indicated a post-translational cleavage of PfPREX (Section 4.2.4). The anti-PfPREXhinge revealed no bands on the Western blots, while the anti-PfPREXpri antisera bound to a band of approximately 88 kDa, slightly smaller than the band obtained with anti-PfPREXpol. Several bands were obtained on the blot probed with anti-PfPREXpri antisera. There were two other high molecular weight bands observed in both the 3D7 and red blood cell lysates. The bands could be a crossreaction between the polyclonal antisera and a host polypeptide found within the red blood cell lysates. Using the anti-PfPREXheli antisera to probe the Western blot, several bands were observed on the blot. There was a low molecular weight band on the red blood cell lysate which was not present in the 3D7 lysate. This could be a crossreaction with a protein found in just red blood cells but not in infected red blood cells. On the 3D7 lane, multiple bands were observed. There were an approximately 150 kDa band, followed by a 80 kDa and a 50 kDa band.

The result obtained from the anti-PfPREXpri together with the result obtained from the anti-PfPREXpol, both suggest that PfPREX may be posttranslationally cleaved to give separate DNA polymerase domain from the DNA primase/helicase domain. The size of the band (88 kDa) obtained with the anti-PfPREXpri antisera is consistent with the primase/helicase domain being together (as seen for the bacteriophage T7). However, the size detected with the anti-PfPREXheli antisera disagrees with that. Both antisera should identify the same band if the DNA primase/helicase are together. The different size bands obtained with the anti-PfPREXheli suggest that the some of the DNA primase/helicase are also cleaved into several different fragment sizes. Further experiments, possibly with additional antisera, would be required to resolve the ultimate processing pathway of PfPREX.



Figure 5.11. Western Blot of parasitised rbc (3D7) and non-parasitised rbc lysates probed with anti-PfPREXpri and anti-PfPREXheli. (A) Western blot probed with anti-PfPREXpri (1:1000) and (B) Western blot probed with anti-PfPREXheli (1:1000).

5.2.5. Subcellular localisation of PfPREXpriheli gene product

With reference to the presence of the plastid-targeting sequence in the Nterminus of PfPREX, experiments to identify the subcellular localisation of PfPREXpriheli domain were performed. The anti-PfPREXpri, anti-PfPREXheli and anti-PfPREXhinge were used in indirect immunofluorescence microscopy experiments. Since with the previous studies using anti-PfPREXpol with *in situ* hybridisation using plastid DNA specific probes did not give satisfactory results, that approach was not used here. We also used the Mitotracker stain (Molecular Probes) to stain the mitochondria in *P. falciparum* and DAPI to counterstain the nucleus.

To optimise the IDF protocol, various dilutions of the anti-sera were used, ranging from 1:50 to 1:1000. When higher dilutions are used, no fluorescence could be observed with any of the three anti-sera and when lower dilutions (below 1:100) were used, ubiquitous fluorescence was observed. Finally, the optimum dilution of the anti-sera to be used for anti-PfPREXheli, anti-PfPREXpri and anti-PfPREXhinge was determined to be 1:100 alongside 1:400 dilution of the secondary antibody (anti-rabbit FITC, Diagnostics Scotland) (Figure 5.13). The immunofluorescence detected in the infected red blood cells observed under the microscope. For the anti-PfPREXheli and anti-PfPREXheli and anti-PfPREXpri results, there appears to be no overlay between the mitotracker and the FTTC. This suggested that the PfPREXpriheli did not localise to the mitochondrion in *P. falciparum*. However, it was not clear whether either anti-sera localised to the plastid. The DAPI stains both nuclear and plastid DNA, but it is difficult to see the plastid under the highest available magnification.



infected with a ring-form parasite, and stained with different fluorescence markers outlined in the text. The second picture shows the same cell viewed under fluorescence using the Texas red filter with an excitation and emission peaks of 560 and 630 nm respectively. The third picture shows the same cell viewed under fluorescence using the FITC filter with an excitation and emission peaks of 484 and 510 nm respectively, followed by the DAPI filter with an excitation and emission peaks of 360 and 460 nm respectively and the last set of image were all previous images merged together. The mitotracker stain shows the locality of the mitochondrion, FITC shows where the anti-Figure 5.12. Subcellular localisation of PfPREXpol using immunofluorescence technique. The first diagram is a phase contrast microscopy picture of a red blood cell PfPREXpol antisera have bound and DAPI stains specifically nuclear and apicoplast DNA. (A) shows the immunofluorescence results obtained with the anti-PfPREXpri antisera and (B) shows the immunofluorescence results with the anti-PfPREXheli antisera. Scale bars, $5 \,\mu m$.

5.3. Discussion

The experimental findings and sequence analysis suggested that PfPREX consists of a DNA primase and DNA helicase functional domains. This chapter provides a functional confirmation of these activities. In addition, the reaction conditions may not be optimal for the helicase activity of PfPREX. Therefore, the DNA helicase was unwinding at suboptimal conditions, which may have been why the time lag occurred. However, one positive control, E. coli DnaB (a gift from Dr E Biswas) showed low unwinding activity (<30% unwinding activity). In comparison, the same quantity of the PcrA achieved unwinding activity of up to 60% over the same time course. In addition, Amersham also claimed that using E. coli DNA helicase I, the highest unwinding activity achieved in their labs was 75%. Therefore, the low rate of unwinding activity observed with DnaB and PfPREXpriheli was not due to a fault with the kit. Both PcrA and E. coli DNA helicase I, unlike DnaB, are not hexameric DNA helicases. Perhaps, the reaction conditions of the kit are not suited for hexameric DNA helicases like DnaB and possibly for PfPREXpriheli, which is most similar to DnaB helicase family and thus likely to take on a hexameric form to function. Currently we could not come up with any function reasons that could be put forward forth to explain why the assay did not work as well. Therefore, it is postuated that the conditions required for these helicases to form the correct structure for DNA unwinding were not provided, which may explain the lower than expected unwinding activity.

The NTP requirements of the DNA helicase is also investigated. It appears that PfPREXheli does not have major preference over the type of NTP it utilises for hydrolysis. This trait can also be observed with the T4 DNA helicase and T7 DNA helicase (Matson & Kaiser-Roger, 1996). The RT-PCR results suggest that PfPREX is transcribed as a single transcript. Unfortunately, we were not able to obtain any Northern blot data to give the size of the PfPREX transcript. However, in *P. chabaudi*, a 7 kb transcript of the PcPREX homologue has been detected (C Janssen, personal communication). Towards the end of this study, Le Roch *et al* (2003) carried out a detailed transcriptome analysis of *Plasmodium falciparum*. In this study, it has shown that *PfPREX*'s transcription is upregulated (data available in Appendix B) when the ring stages are developing into trophozoites and continues into schizogony , which corresponds to the initiation of DNA synthesis as observed (Inselburg & Banyal, 1984).

The Western blot results obtained with the anti-PfPREXheli and anti-PfPREXpri anti-sera supported the conclusions of the result obtained with anti-PfPREXpol, that PfPREX is post-translationally cleaved. It was postulated that the PfPREX would be cleaved into two fragments, to give one that contains the DNA primase and DNA helicase domains and the other to contain the DNA polymerase domain. The result obtained with anti-PfPREXpol suggested that the DNA polymerase domain is on a separate fragment. If the DNA primase and helicase domains are together on a single fragment, the Western blot results obtained with both individual anti-sera would show similar size bands. However, this was not what was observed. The Western blot probed with anti-PfPREXpri anti-sera gave a band of the expected size of approximately 88 kDa, but the blot probed with PfPREXheli anti-sera gave a multitude of different bands and nothing at the 88 kDa mark. On the same blot, the anti-PfPREXheli appeared to be cross-reacting with a polypeptide present only in the uninfected red blood cell lysate. It could be that the fusion protein sent for antisera production is not pure and could be contaminated with some other proteins. In addition, the anti-PfPREXheli antisera have not been checked against the recombinant PfPREXheli to ensure that it contains antibodies specific against the protein. Nonetheless, the other Western blot results do suggest that the post-translational cleavage of PfPREX does occur.

Finally, the subcellular localisation of PfPREXpriheli was carried out to help reveal the location of the protein *in vivo*. The immunofluorescence results obtained revealed similar conclusion as with the anti-PfPREXpol results. Both sets of results agree that PfPREX does not localise to the mitochondrion but the results were not able to tell us if PfPREX localised to the apicoplast. Due to the fact that the apicoplast is in close association with the nucleus within *Plasmodium* (Bannister *et al*, 2000), when DAPI (which stains both nuclear DNA and apicoplast DNA) is used to stain the cells, it is difficult to tell the apicoplast DNA and nuclear DNA apart. However, a similar staining pattern was observed with FITC fluorescence, where both anti-PfPREXpri and anti-PfPREXheli gave results that indicated the compartmentalisation of the products that these antiscra were bound to. Chapter VI

The Evolution of *PfPREX*

6.1. Introduction

The discovery of a gene sequence, *PfPREX* that encodes three functional domains, where each may have a role to play in the replication/repair of the plastid DNA in *P. falciparum* raises a number of questions regarding its origins and evolution. To help answer these questions, phylogenetic analysis on each of the individual domains was carried out. This analysis has shed light on the evolutionary pathway that led to the derivation of PfPREX in *P. falciparum*.

P. falciparum belongs to the phylum of parasites called the Apicomplexa. All members of the group are characterised by a complicated apical complex found at the anterior of the cell. In recent years, they have been found to harbour a relict plastid – termed the apicoplast. According to Gray (1992), all plastids ultimately derive from cyanobacterial-like, prokaryotic organisms that were retained by a phagotrophic eukaryote. They established an 'endosymbiotic' relationship whereby the 'engulfed' cell is retained in a mutually beneficial relationship with the host. In the apicomplexa, the apicoplast too was derived from such an event, but instead of the eukaryote engulfing a cyanobacterial prokaryote, it acquired a eukaryote that already harboured a plastid. This is referred to as a secondary endosymbiotic event (Delwiche & Palmer, 1997).

The fact that many plastid genes and biochemical pathways are of prokaryotic origin, agrees with the serial endosymbiosis theory of the origin of the plastid. For example, *P. falciparum* contains a type II pathway for fatty acid biosynthesis, typical of bacterial-like systems and also utilised by plants. This indicates that this pathway was inherited with the plastid in both *P. falciparum* and plants

(Harwood, 1996). Given the plastid localisation of PfPREX, it was of interest to determine whether the DNA polymerase, DNA helicase and DNA primase were derived from a prokaryotic source, particularly a cyanobacterium. Therefore, a detailed phylogenetic analysis was performed to test this hypothesis. Since PfPREX contains three functional domains, the domains were analysed separately.

6.2. Phylogenetic analysis of DNA polymerases

6.2.1. DNA polymerase families

In recent years, with the boom in genome sequencing, many gene sequences have been gathered and a large number of those encode DNA polymerases. These DNA polymerases have been studied closely and phylogenetic analyses have been carried out. They have been classified into various families and subgroups. The most cited work was carried out by Braithwaite and Ito (1993) where they have placed DNA polymerases into four distinct families (A, B, C and X). Recently, a new class (class Y) has been added (Ohmori *et al*, 2001). Braithwaite and Itos' (1993) phylogeny provided the framework for the comparative analysis reported here. Filée *et al* (2001) have attempted to analyse the evolution of DNA polymerase families and have suggested that there were multiple gene exchanges between cellular and viral proteins. These will be discussed in later sections within the chapter.

BLAST searches reported earlier (Chapter 3) indicated that the DNA polymerase domain of PfPREX (PfPREXpol) has highest similarity to the *Aquifex aeolicus* DNA polymerase (Genbank accession no: - AAC07735.1) (more recently, *Aquifex pyrophilus* DNA polymerase [Genbank accession no: AAO15360.1]) and other prokaryotic type bacterial DNA polymerase I sequences, all of which belong to the family A DNA polymerases. This preliminary data suggests that PfPREXpol groups

within the family A group of DNA polymerases to which these bacterial DNA polymerase I sequences belong (Braithwaite & Ito, 1993). Therefore, the phylogenetic analysis was carried out with DNA polymerase I sequences from members belonging to this family.

6.2.2. Alignment of DNA polymerases

The "type"-member representative of DNA polymerase family A is *E. coli* DNA polymerase I which consists of a 5'-3' exonuclease domain, a 3'-5' exonuclease domain and a DNA polymerase domain. PfPREXpol does not have a typical 5'-3' exonuclease domain. In order to produce alignments of optimal phylogenetic content, DNA polymerase I sequences that possess 5'-3' exonuclease domains used in the alignments were culled to remove their 5'-3' exonuclease sequences.

Several DNA polymerase I sequences (some are selected based on their recognition by PfPREXpol in BLAST searches) from various organisms were used for the alignment. They included bacterial homologues, DNA polymerase I sequences from bacteriophages and mitochondrial DNA polymerases gamma. A novel eukaryotic nuclear DNA polymerase/helicase (Harris *et al*, 1996), the N-terminal part of which is a prokaryotic DNA polymerase I-like sequence, was also used. This protein, Mus308 (NCBI accession number:- NP_851315) was first isolated and characterised in *Drosophila melanogaster* and is also called DNA polymerase eta (Harris *et al*, 1996). *Mus308* orthologues are found in other eukaryotes including *Homo sapiens* to *Mus musculus*. Filée suggested that the eukaryotic Mus308 could be of mitochondrial origin. Therefore, it was of interest to include this sequence in our phylogenetic analysis.

All of the sequences used in the alignment are listed and tabulated in Table 6.1. The table shows the full names of the organisms from which the DNA polymerase sequences were taken and their database accession numbers. The alignment comprises 24 sequences alongside the PfPREXpol sequence.

The sequences were aligned first using the DIALIGN version 2 programme (Morgenstern et al, 1996) using the default settings (Section 2.11). DIALIGN relies on the comparison of whole segments of sequence rather than single residues. Pairwise, as well as multiple, alignments are constructed from gap-free pairs of equal length segments, which are called diagonals. Thus this method is well suited to detect local similarities on otherwise unrelated sequences. This preliminary alignment was carried out to enable us to determine which part of the sequences should be culled in order to retain the most phylogenetically informative sequence. Essentially the 5'-3' exonuclease domains of some of the sequences were removed as indicated in Table 6.1. The culled alignment was realigned using ClustalX1.81 (Thompson et al, 1994) using the settings shown in Table 6.2. Clustal X is a global alignment program for nucleotide or amino acid sequences. It uses a progressive alignment algorithm to align sequences. In other words, it aligns more similar sequences before the distant sequences. So the pairwise and multiple parameters were set as shown. The general rule of thumb for choosing the parameters are:- the higher the gap opening (GO), the less frequent the gaps and the higher the gap extension (GE), the shorter are the gaps. Therefore, the settings were chosen to optimise our particular alignment. The full alignment is shown in Appendix D.

Phyla	Enzyme	Organism	Shortened Name	Length of sequence	Genbank/ embl accession	Reference
			1 4611110	(a.a)	no,	
Bacteria	poli	Mycobacterium tuberculosis	Myctu	904	AAB46393.1	Huberts & Mizrahi , 1995
	poll	Thermogata maritima	Thermogata	893	AAD36686.1 ⁺	Nelson <i>et al</i> , 1999
	poll	Aquifex aeolicus	A.aeolicus	574	AAC07735.1	Deckert <i>et</i> <i>al</i> , 1998
	poll	Aquifex pyrophilus	A. pyrophilus	574	AAO15360.1	Direct submission
	polI	Escherichia coli	E.coli	928	САЛ23607.1*	Joyce <i>et al</i> , 1982
	poli	Salmonella typhimurium	S.typhimurium	928	AAG43170.1 ⁺	Direct submission
	poll	Thermus aquaticus	T.aquaticus	832	AAA27507.1*	Lawyer <i>et</i> <i>al</i> , 1989
	poll	Synechocystis sp	Synechocystis	986	BAA10748.1*	Kaneko et al, 1995
	poll	Rhizobium leguminosarum	Rhizobium	1016	AAD45559.1*	Huang <i>et al</i> , 1999
Virus	poll	SP6 phage	SP6	849	AAP48753.1	Direct submission
	ροΪΪ	T7 phage	T7	704	CAA24412.1	Dunn & Studier, 1983
	polI	T3 phage	T3	704	CAC86283.1	Pajunen <i>et</i> al, 2002
	poll	phiYe03-12 phage	phiYe03-12	704	CAB63614.1	Pajunen et al, 2001
	polI	Cyanophage P60	Cyanophage	587	AAL73268.1	Chen & Lu, 2002
Yeast	poly	Saccharomyces cerevisiae	Mip1p	1254	CAA89977.1	Direct submission
Plants	poll	Arabidopsis thaliana	A.thaliana	1049	AAM13892.1	Direct submission
	poll	Oryza sativagi	Rice	976	BAB40805.2	Kimura et al, 2002
Fungi	poll	Dictyostelium discoideum	Dictyostelium	1369	AAO12053.1*	Direct submission
Insect	poly	Drosophila melanogaster	Fly	1145	AAC47290,1	Direct submission
Mammal	poly	Mus musculus	Mouse	1238	AAA98977.1	Direct submission
	ροίγ	Homo sapiens	Human	1239	AAC50712.1	Direct submission
	polN	Homo sapiens	polN	900	AAN52116.1	Direct submission
	թօհղ	Homo sapiens	eta	1154	AAD05272.1	Direct submission
	polN	Mus musculus	MmpolN	866	AAN39837.2	Direct submission

Table 6.1. The DNA polymerase sequences used for multiple alignment. The phyla, name of organism, number of amino acids, type of polymerase and the accession number of each DNA polymerase sequence is shown. The shortened name of each sequence used in the alignment is also shown. The + signs indicate those sequences whose 5'-3' exonuclease domains are removed from the alignment.

Step	Sequences	Pairwise	Multiple	a.a matrix
		parameters	parameters	
1	All polymerases	GO 10	GO 10	Gonnet
	excluding 5'-3' exonuclease	GE 0,1	GE 0.1	
	domains			l
2	Step 1 sequences excluding 3'-5'	GO 10	GO 10	Gonnet
	exonuclease domains (motifs A-C).	GE 0.1	GE 0.1	
	Large gaps are removed.			
	DNA primase sequences	GO 10	GO 10	Gonnet
		GE 0.1	GE 0.1	
· · · · ·	DNA helicase sequences	GO 10	GO 10	PAM250
		GE 0.1	GE 0.1	L

Table 6.2. Various parameters used in ClustalX 1.81 for the alignment of the DNA polymerase, DNA helicase and DNA primase sequences. The sequences used, pairwise and multiple parameters are shown. GO – Gap opening penalty, GE – Gap extension penalty. The amino acid substitution matrix used is shown in the right most column.

The alignment revealed that the 3'-5' exonuclease domains of the DNA polymerases were relatively poorly conserved. Conservation was only observed within the three Exo motifs. Therefore, it was decided to use only the polymerase domains, which spans the three key polymerase motifs A- C (Section 1.4.3.1.3). The alignment was checked through by eye first before it was culled and realigned for a third time using ClustalX 1.81. Positions that could not be aligned unambiguously were excluded as they carried no phylogenetically useful information. Olsen (1988) also argued that these parts of the sequences should be excluded from phylogenetic analyses. All of the sequences retained the three key polymerase motifs. The alignment was checked through manually and adjusted as necessary to improve the alignment. This final alignment was then used for later phylogenetic analysis (Figure 6.1).

Mouse Human Fly Miplp E.coli S.typhimurium Rhizobium Myctu Thermotoga T.aquaticus Synechocystis polN MmpolN eta A.thaliana Rice Dictyostelium A.aeolicus A.pyrophilus PfPREXpol T7 T3	1 DEEGHYGAILPQVVTA TTTRRAVEPTWITASARPDRVSELMAMVQAPTGY 1 DEEGLYGAILPQVVTA TTTRRAVEPTWITASARPDRVSELMAMVQAPPGY 1 CQPIAYGAICPQVVAC TTTRRAMEPTWITASARPDRVSELMAMVQAPPGY 1 CQPIAYGAICPQVVAC TTTRRAMEPTWITASARPDRVSELMAMVQAPPGY
Cyanophage	1BSCVLN-INTCPOAHMRPNIACVPSHEYPSLEKPSENHL
SP6	1PUDLAGKEK
Mouse Human Fly Miplp E.coli S.typhimurium Rhizobium Myctu Thermotoga T.aquaticus Synechocystis polN MmpolN eta A.thaliana Rice Dictyostelium A.aeolicus A.pyrophilus PfPREXpol T7 T3 phiYeO3-12 Cyanophage SP6	53VLVGALVDSCEIW ALVIGDAHLAGMHGCT FGWMTLQGRKSRGTD HSKTAATVRI 53TLVGALVDSCEIW ALVIGDAHFAGMHGCT FGWMTLQGRKSRGTD HSKTATTVGI 53RINGALVDSCEIW ASVIGDAYACGEHGATPLGWMTLSGKSNGSD HSITAKAVGI 53CFYGALVDSCEIW ASVIGDAYACGEHGATPLGWMTLSGKSNGSD HSITAKAVGI 53CFYGALVDSEIW ASVIGDAYACGEHGATPLGWMTLSGKSNGSD HSITAKAVGI 53CFYGALVDSEIW ASVIGDAYACGEHGATPLGWMTLSGKSNGSD HSITAKAVGI 54VIVS NIVSUFELFWAHSRIKGETAFAEDKOTHRAFAEWFG PLESVTSE 44VIVS NIVSUFELFWAHSRIKGETAFAEDKOTHRAFAEWFG PLESVTSE 55ELSTALVOIGHMAHNSGEGGTFEAFNTEDISFVSRAFGVPIGVEGPGG-E 45WIVSUFELFWAHSGEGGTFEAFNTEDISFVSRAFGVPIGVEGPGG-E 45WIVSUVSUFELFWIAHNSGEGGTFEAFNTEDISFVSRAFGVPIGVEGVPGG-E 44VIVSUFELFWIAHNSGEGGTFEAFNT
010	
	Motif A
Mouse Human Fly	110 HEEHATIFUTCRIYGA QSFAERLLMQFNHRLTROFAB KAQQMNAVTKGIRRAR 110 SEHATIFUTCRIYGA QPFAERLLMQFNHRLTQOFAB KAQQMHAATKGIRWAR 110 SEDHATVTIYAR QLFAETLLROFNPTFSAS KAKAMMMASTTKKKUWAR
Miplp	109 SENERTIFUKER MEANERFASO LLKRFNPSLTDE TKKIANSLVENTKE
E.coli S.tvphimurium	97 ORSALAU 2011 YEMSN FG 2ROUNT PRK DACKYODL YEERY PSYLEYM 97 ORSALAU 2015 FG 2ROUNT PRK DACKYODL YEERY PSYLEYM
Rhizobium	97 VERRANATURUTIYUIS FG MNQUSTERS MGYUKKWEREPEREPERED
Myctu	98 LERRVIANSYGLATCLSTYG SQQUKESTEMANNOVDAY HAREGEVEDYL
Thermotoga T aquaticus	98 MURACHMUNISI OVVTPYGISVRIGWPVKDWEKMIVNA VLWPKWRDMI 97 MURACHMUNISI OVVTPYGISCFISSUPYEDWEFMINGSWDKWRDMI
Synechocystis	95 EENLGET LIFEVING ORFERETGISAVREFIDRINGTONL
polN	104 DEEQTHKWWWAVAYNGA KER AACIGUPIQMAAQFHESELQKNKKHKDDA
MmpolN	104 DECTHKVVISWARA KER AACIGUTVLMTHFIERLQKNKK KDA
A.thaliana	98 ERKAMMUSSAACKTIVG SRDWKWSTKDOTTUDLUMNDROEWRKOO
Rice	98 EERKAMMATESTATCKTAVGISWDWKWSVRDARTTKLWTRDRKEVSAMO
Dictyostelium	98 ERREATTINESTATCKTHEGSQDWGVTLNZAKETTINEVXEDRPEVLVWQ
A. pyrophilus	91 EDOLARAUTECHTYCTSAKGLAEVARTGYGVENSEEBARTERNEREVARVARVARV
PfPREXpol	93 DEHIATATUEGI YOMNYVN KNYANTYYGINASLD CLYFYNSEEHWKAYKM
Т7	101 - DNATTFICFLIGAEDEKICQINGAGKEREKELKKKELENTPATAALR
T3	101 DNALTFIKEFLIKEADEKLEQIVGAGKEREKELKKKELENTPALAALR
Cvanophage	101 - DAWLTTTTTTTTTTTTALDEKICQIVGAGKERCKELKKMLENTPATAALR 85 DEKSELGUTCCLUEGENHKKELTAGASKSSASPKEOTPGATMOGLSEFADIN
SP6	85 KEDMANTFILEFILLES IAN AAVCGWTEDMKEVWARFEIELPSTARLR

Motif B

Mouse	165	AWTGGTESEMENKUESTAMSDTPRTPULGCCISRALEPSVVOGEFITSRUNWVKAS
Human	165	AWKGGTESEMENKUESTATSDIPRTPULGCCISRALEPSAVOEEFMTSRVWWWWSS
Fly	165	-WOGGTESAMENREFERATGSOPRTPFLGGRLSRALEADEOEORFLPTRINWVW.SG
Minin	159	-WYGGSESTLENKUESTAEOETPKTPULGCGITYSLMKKLRANSFLPSRHWARGS
E coli	147	BTRACHEORY
S tunhimurium	147	BTRACHTECKY DEWERHERT, YHEDIKSSNAAR
Phizobium	147	SPRAMA DREYWEINFRUHTNYEF BSSNDSVEAFNERAA MADT DES
Muctu	148	RAVUERASKO VTSHUL HER VEELDSSNROVEEAFEAA HAADHAB
Thermotoga	148	OPUUSEANERKY WRITERIA DIA OFMARDRNTOAEGERTATETOR
Tamaticus	147	CTLEE BODDY
Supechocustis	145	STARLENTAKEYWTUTUTEEYENFETEALROLECKMNYNDAOLLUSAADAFTOCS
bynechocyscis	154	PAATAOCHOTEC
MmpolN	154	OTVICOCESALYWTGHT, EDEEPERPERCAODOOLEAOAELOAVERVAOGS
ota	152	TUKN-CEPDEF
A thaliana .	148	MERKENTED YWILLITT HSERTASKSRAOSNHTORAMITTEVERS
Rico	140	KKOKAFOLEKCE
Dictuostelium	148	
DictyOscerium Declique	147	PUKKET NEKETWKCHTLIC PERANTEND
A. acorrous	147	WUKKEI DEKU
PEPPEYpol	149	
m7	150	STOOTLVESSOWVACEOOKKWEEKWEKCEDCRKVHVE
m2	150	
15 phiVo03-12	150	BATOOST UNSSAUT COFORMON WERE CODOR VINUE
Cuapophago	130	
cyanophage	135	ANUTAANNERYODDCUWCDEMSCOFT
M2	133	
	100	
Mouse	222	
Human	222	WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW
Fly	221	THE MAN SARWINGS
Minln	215	AVALUTICS MANTERKYN LEAR CUS HOADRAT SEK KYRAAMANOT SNEWT
E coli	196	AND THE REAL AWAR OAE OPRVING WHILE WERE H-KEDVE AVAKO HOL VEN
S typhimurium	196	WART THE WHAT AWOAF OPRVENTION HOLE WEEK H-KEDLEAVAKENHOLEN
Bhizobium	196	NADWIER MARKIERA VEVGLADRVEN OWNER FAVE-DODVERAMEW VSVEN
Myctu	197	AND AND MANUAL NEAOLAS PAIN OF HIM ALCONG A POER BY AND
Thermotoga	197	AND THE PREFERENCES
T. aquaticus	196	AADTMALE MURIFPRICELGARMALOVIDELAT AAP-KURABRVAAUAKEVNEG
Synechocystis	201	SADI HELAMWKHAKI JESYOTEMIA OVUDDALED AP-PREWEELAPINONTHEO
polN	203	AAUACHIAMHHWETAWAASH-TI.TARIWAOUUULAATWE-DPOTPECAAUARBTHES
MmpolN	203	AADACHAMTRUSTAWATSP-TLTARWAOLUOFLUMENE-DTOVPEFAALWRRTVES
eta	200	AADT VELATION OKOLET FHSTEKSHGEFT OFHDELLY EVA-EDVVOVAOLI KNEVES
A.thaliana	195	AADVAMC WEETSINOOLKKLGWEYWKONHDEVELEGP-IESAEIAKDWVDCUSK
Rice	197	AADVAMCAMETERNARLKELGWBHDRVHLEGP-TESAFEAKTHVECUSK
Dictvostelium	196	MAD MMKAMMI THEDNKRLKELGERLEUGER-LEGE-ECHAREARS MMNLMSN
A.aeolicus	186	CADAMNAVALFIAN OKKGIDAR WNLWHOL WVICE-KEKABEVKE EKSEKT
A.pyrophilus	186	GADILE LAVELF AEAKKKKLDAKLVNLVHDELVVECR-KEVANOVKEVJEKANKO
PfPREXpol	185	CNDIBILALVDIYDN KDINGRIILCVIDERIIEVN-KKFOREALKIIVOSVEN
т7	200	CALICKIWIIKTEEMIVEKG-LKHGWDGYMAWVHDEIQVGCRTEEIAOVVIETAOEANRW
Т3	200	GALICUWIIKTEEM VEKG-LKHGWDGY AWDUDELOVACRTEEIAKTVIEVAOEA RW
phiYeO3-12	200	CALICE WITKTEEM IEKG-LKHGWDGOMGWLHDEVOIACRTEEVAKVVVEVAOEA RW
Cyanophage	174	GATIC WY TRTHELLOEAG-IDYPLAFVHDPOOLSVR-ABOAMAAOUTTLAKKD
SP6	178	SELCONYALVKAFAVERREG-VALDNLGGVANVHDELQMEVPEORRYHRAGHT ADAVTW
M2	175	FREFERENCY ROKTY ODIYWEVICKER CSPDEATTTKESERCAGETD

Motif C

Figure 6.1. Alignment of deduced amino acid sequences of DNA polymerase domains of DNA polymerases from family A. Details of sequences used are shown in Table 6.1, alongside PfPREXpol. These sequences were aligned using ClustalX 1.81. Black shading indicates identical amino acids; grey shading indicates conserved amino acid substitutions. The numberings on the left hand side show the relative position of the amino acid in the respective sequences. Motifs A, B and C are shown with underlined text.

6.2.3. Generation of phylogenetic trees for the DNA polymerase domains

Phylogenetic trees for the family A DNA polymerases were generated using MEGA (Molecular Evolution Genetic Analysis) software version 2.1 (Kumar et al, 2001). The trees were created from the alignment (Figure 6.1) which comprised the three key motifs found in family A polymerases. Four different types of trees were generated with the alignment set. All trees generated were unrooted. Three trees were generated using distance methods: Unweighted Paired Group Method with Arithmetric Mean (UPGMA), Neighbour-Joining (NJ) and Minimum Evolution (ME) and one from a character-based method, Maximum Parsimony (MP). Distance methods convert aligned sequences into a pairwise distance matrix and then input that matrix into tree building while character-based methods use a specific criterion to build a tree. This criterion is then used to assign to each tree a score which is the function of the relationship between tree and data. The distance trees inferred from the amino acid sequences were computed on the basis of the Poisson correction in the MEGA programme. Poisson correction is calculated as the distance between two amino acids based on the assumption that the rate of amino acid substitution at each site follows a Poisson distribution.

The four different methods were chosen to determine the stability of the tree topologies. All four trees (Figure 6.2) showed six distinct clusters: - bacteria, plant, mammal, bacteriophage, mitochondria (γ) and *Aquifex*. The mitochondrial DNA polymerases (γ) form their own cluster and are not part of the family A cluster. However, this differs from the conclusion of Filée *et al* (2001) who noted in their trees that DNA polymerase γ was only distantly related to bacterial DNA polymerase I, (also

observed in our trees) but that it branched with the T7/T3 DNA polymerases. This was not observed with the trees obtained in our phylogenetic studies.

PfPREXpol clusters very tightly with the Aquifex species of DNA polymerase sequences and sits more closely to the bacterial cluster of DNA polymerases than to the eukaryote cluster. The bootstrap support for this particular branching is very strong, with values ranging from 86 to 100 for all four trees. Throughout the four trees, the topology whereby the PfPREXpol sits within the Aquifex cluster was maintained, which showed the stability of this branching. Aquifex is a thermophilic bacterium. The similarity between PfPREXpol and Aquifex is without precedent. Remarkably, in addition to the extraordinary similarity in sequence between the P. falciparum sequence and the Aquifex protein, the Plasmodium protein also encodes a thermophilic DNA polymerase with maximal activity as 75°C (Section 4.2.3.1). The optimum temperature for the activity of PfPREXpol was similar to that of the Aquifex aeolicus (Chang et al, 2001) at 75°C. These findings suggest that the origin of PfPREXpol might be from a thermophilic relative of Aquifex which was quite unexpected and indicative of a relatively recent horizontal gene transfer. It is also noteworthy that a DNA polymerase I analogue believed to be involved in the regulation of chloroplast DNA replication in plants (Oryza sativagi) is clearly more distantly related to PfPREXpol than are the Aquifex polymerases. This indicates that DNA polymerases involved in the apicoplast and plant chloroplast DNA replication have separate origins.



.

(a)







Figure 6.2. Phylogenetic trees generated by MEGA derived from the DNA polymerase alignment (Figure 6.1). The bootstrap values (100 resamplings) are indicated on the branches. Unrooted trees were generated using the distance method UPGMA (a), the NJ method (b), ME method (c) and MP method (d).

6.3. Phylogenetic analysis of the DNA helicase and DNA primase domains of PfPREX

÷.

6.3.1. DNA helicases and DNA primases

6.3.1.1. DNA primases

DNA primases are enzymes that are essential for DNA replication and their activity is required at the DNA replication fork. The activity involves the catalysis of synthesis of short oligoribonucleotides used as primers for DNA polymerases (Kornberg, 1980; 1982). Most DNA primases can be divided into 2 classes: - class I comprise bacterial and bacteriophage primases (Section 1.4.3.1.2) and class II contains heterodimeric eukaryotic type primases. The prokaryotic primases are found to be associated with replicative DNA helicases while the cukaryotic primases form a complex with DNA polymerase alpha. All DNA primases share several key motifs, all of which are found to be present in PfPREX (Section 3.3.2.1).

6.3.1.2. DNA helicases

Like the DNA primases, DNA helicases also play essential roles in DNA replication. They are responsible for the unwinding of double stranded nucleic acids before DNA replication, repair or recombination. This reaction is dependent on energy derived from nucleoside-5⁺-triphosphate (NTP) hydrolysis. All DNA helicases share several common motifs as discussed in Chapters 1 and 3. These motifs are also observed in PfPREX (Section 3.3.3). There are several classes of helicases including superfamily (SF) 1 to 3 and the DnaB-like group. Along with the DNA primases, I have also aimed to explore the phylogenetic relationship of PfPREXheli alongside other helicases in the hope of shedding light on their origin and on the overall evolution of the extraordinary structure of PfPREX.

6.3.1.3. DNA Helicase/Primase Organisation

Bacteria possess separately encoded replicative DNA primase and DNA helicase. In *E. coli*, they are termed DnaG and DnaB, although their protein products join in multimeric complexes. In bacteriophages, there are two distinct types of DNA primase and DNA helicase system. One group is similar to the bacterial system (DNA primase and DNA helicase are encoded separately e.g. P22 and T4 bacteriophages). The second group comprise the T7(T3) and P4 bacteriophages in which their DNA primase and DNA helicase are found on a single polypeptide (Figure 6.3). Ilyina *et al* (1992) investigated these one and two component DNA primase/helicase systems. Pronounced



Figure 6.3. Organisation of DNA primase/helicase systems in bacteria and phages. (Adapted from Ilyina et al, 1992).

sequence homology was observed between the latter group and PfPREXpriheli in both the DNA primase and also the DNA helicase domains. Given also that the primasehelicase is fused and that *Plasmodium* has a tendency to fuse genes, it was of interest to determine the relatedness of this protein with the others.

6.3.2. Alignment of DNA primases

Several DNA primases from various organisms based on best homologies in BLAST searches were included in the alignment. The DNA primase sequences ranged from examples of the single component DNA primase/helicase systems of both bacteria and phages and the two component ones from phages (Table 6.3). A sequence from *Arabidopsis* was also included as it had homologues to this group. A total of 13 sequences were used in the DNA primase alignment, alongside the PfPREXpri sequence.

The sequences were aligned first using the DIALIGN version 2 programme (Morgenstern *et al*, 1996) with the default settings (Appendix D), as with the DNA polymerase alignment. The alignment was checked manually and large gaps were removed. The culled alignment was realigned using ClustalX 1.81 (Thompson *et al*, 1994) using the settings as indicated in Table 6.2 and was checked and adjusted manually.

From the alignment, a strong conservation at the N-terminus, of two pairs of Cys/Cys or Cys/His residues was observed. This conserved motif is the zinc-binding motif (Section 3.3.2). The zinc-binding motif is present throughout the bacterial and phage primases. In addition to the zinc-binding motif, the DNA primase alignment possesses another five conserved sequence motifs (Figure 6.4). This alignment was used for the generation of phylogenetic trees.

Phyla	Organism	Shortened Name	Length of sequence (a.a)	Genbank/embl accession no.	Reference
Virus	P4 phage	P4	777	CAD06954.1	Parkhill <i>et al</i> , 2001
	T7 phage	17	566	CAA24405.1	Dunn & Studier, 1983
5	T3 phage	T3	566	CAA35135.1	Direct submission
	T4 phage	. T4	475	AAA32554.1	Direct submission
	phiYe03-12	phiYe03-12	566	CAB63608.1	Pajunen <i>et al</i> , 2001
	Cyanophage P60	RP	531	AAL73266.1	Chen & Lu, 2002
	Pseudomonas phage gh-1	gh-1	544	AA073154.1	Direct submission
Eukaryote	Arabidopsis thaliana	Arabidopsis	709	NP_849735.1	Direct submission
Bacteria	E. coli	E.coli	581	CAA23531.1	Smiley <i>et al</i> , 1982
	Salmonella typhimurium	Salmonella	581	AAL22084.1	McClelland et al, 2001
	Bacillus subtilis	B.subtilis	603	CAB14464.1	Kunst <i>et al</i> , 1997
	Synechocystis sp	Syne	635	BAA18229.1	Kaneko <i>et al</i> , 1995
	Pseudomonas putida KT2440	Pseudomonas	544	AAN67883.1	Nelson <i>et al</i> , 2002
	Aquifex aeolicus	Aquifex	498	AAC07430.1	Deckert <i>et al</i> , 1998

Table 6.3. The DNA primase sequences used for multiple alignment. The phyla, name of organism, number of amino acids, type of polymerase and the accession number of each DNA primase sequence is shown. The shortened name of each sequence used in the alignment is also shown.

6.3.3. Alignment of DNA helicases

For the DNA helicase alignment, a total of 20 sequences were included alongside PfPREXheli (Table 6.4). These sequences include two sequences, one from a *Pseudomonas* phage gh-1 and a sequence from *Pseudmonas putida*, both of which were hits in recent BLAST results obtained with PfPREXheli. Mus308 (NCBI accession number:- NP_851315) (Harris *et al*, 1996), which was used in the DNA polymerase alignment possesses both a DNA polymerase and also a DNA helicase domain. The DNA helicase domain is called HEL308 and has been determined to be most similar to the DNA helicases of the superfamily 2 (SF2) (Marini & Wood, 2002). Therefore, it was of interest to include this sequence in our phylogenetic analysis for use as an outgroup. The helicase moiety of a DNA primase-helicase reported in *Arabidopsis* was also included.

Phyla	Organism	Shortened Name	Length of sequence	Genbank/cmbl accession no.	Reference
			(a.a)		
Virus	P22 phage	P22	458	AAM81428.1	Direct submission
	T7 phage	T7heli	566	CAA24405.1	Dunn & Studier, 1983
	T3 phage	T3heli	566	СЛА35135.1	Direct submission
	T4 phage	T4heli	475	AAD42466.1	Direct submission
	phiYe03-12	phiYe03-12	566	CAB63608.1	Pajunen et al, 2001
	Cyanophage P60	RP	531	AAL73266.1	Chen & Lu, 2002
	Pseudomonas phage gh-1	gh-1	544	AAO73154.1	Direct submission
Bacteria	E. coli	E.coli	471	ΛΛΛ23689.1	Nakayama <i>et al</i> , 1984
	Salmonella typhimurium	S.typhimurium	471	AAL23070.1	McClelland et al, 2001
	Bacillus subtilis	B.subtilis	454	BAA05176.1	Moriya <i>et al</i> , 1985
	Pseudomonas putida KT2440	P.putida	544	AAN67883.1	Nelson <i>et al</i> , 2002
	Aquifex aeolicus	Aquifex	530	AAC07803.1	Deckert <i>et al</i> , 1998
Eukaryote	Arabidopsis thaliana	Arabidopsis	709	NP_849735.1	Direct submission
	Drosophila melanogaster	D.melanogaster	395	AAM27521.1	Direct submission
	Caenorhabditis elegans	C.clegans	566	AAA81397,1	Direct submission
	Homo sapiens	HuTwinkle	685	AAK69558.1	Spelbrink et al, 2001
	Mus musculus	MmTwinkle	684	AAL27647.1	Direct submission
	Homo sapiens	HuHEL308	1101	AAL85274,1	Direct submission
	Mus musculus	MmHEL308	527	AAL85275.1	Marini & Wood, 2002
	Saccharomyces cerevisiae	S.cerevisiae	542	CAA88166.1	Direct submission

Table 6.4. The DNA helicase sequences used for multiple alignment. The phyla, name of organism, number of amino acids, type of polymerase and the accession number of each DNA helicase sequence is shown. The shortened name of each sequence used in the alignment is also shown.

The sequences were aligned first using DIALIGN and subsequently the same process was carried out as with the DNA primases (Appendix D). Large gaps were removed from the alignment and then realigned using ClustalX1.81 using the settings indicated in Table 6.2. Five conserved motifs (Ilyina *et al*, 1992) can be found among the group of DNA helicases in the alignment. Motif 1 and 2 correspond to the Walker A and B motifs (Section 3.3.3) found in all NTP-binding proteins (Walker *et al*, 1982). The alignment obtained was used for the generation of phylogenetic trees (Figure 6.5).

6.3.4. Generation of phylogenetic trees for the DNA primases

Phylogenetic trees for the DNA primase sequences were generated using MEGA (Molecular Evolution Genetic Analysis) software version 2.1 (Kumar *et al*, 2001). The trees were created from the alignment (Figure 6.4). As with the DNA polymerases, four different types of trees were generated with the alignment set to give an idea of the stability of the tree topology based on the given alignment.

All of the four trees (Figure 6.6) showed two distinct clusters:- T7/T3 group versus the bacteria/T4 group. PfPREXpri appeared to branch with the T7/T3 group with 100 bootstrap support for the maximum parsimony method while for the distance based method, bootstrap support was 74 to 100. PfPREXpri is most closely related to the cyanophage P60 (RP/Podoviridae) DNA primase (Chen & Lu, 2002). For the T4/bacteria group, the members within the cluster include the other bacterial homologues of DnaG although some of the bootstrap support within the group is not strong. Most importantly, the PfPREXpri node is well supported which implied that the relationship of PfPREXpri with the T7/T3 group is stable. It was of interest too to determine whether the PfPREXheli sequence also branches with the T7/T3 group of

DNA helicases whereby the T7/T3 bacteriophages also belong to the same family, Podoviridae as cyanophage P60.

6.3.5. Generation of phylogenetic trees for the DNA helicases

For the DNA helicases, the same four methods of tree building were used, based on the alignment shown in Figure 6.5. Phylogenetic analysis revealed that the eukaryotic T7-like Twinkle DNA helicases form a sister group branch with the T7/T3 group of DNA helicases with strong bootstrap support in all four trees. Filée *et al* (2003) had previously observed this, where the Twinkle-like sequence, including PfPREXheli sequence as a sister group with the T7/T3 phage sequences. These eukaryotic T7-like DNA helicases have been proposed to be of mitochondrial origin. Filée *et al* (2003) postulated that as mitochondria were derived from a α proteobacterium, the Twinkle-like DNA might be derived from a horizontal gene transfer event from a T7/T3 phage. This is reminiscent of the situation observed with the mitochondrial RNA polymerase (Hedtke *et al*, 1997).

Both the DNA primase and DNA helicase of PfPREX cluster with the T7phage like proteins. This indicates a common origin. It is possible that the DNA primase domain has been secondarily lost in other Twinkle-like DNA helicase of eukaryotic mitochondria, but is retained in *Plasmodium*. Possible evolutionary scenarios will be discussed in the next section.

1HIP ENWEST DANSLFSDEHTYCYACENWTPEDEQKAEQLSNERYS
1HAPEENEESEDGNSVYSDEHEWEPVEHRVPENEEREAKLSNERYS
1HIP DN CSEDGNSLFSDCHTELWVEKWTACNEDTKERASNCRYS
1HVPEN CSEDANSLFSDCHMENDALDEYVKEDGESSGERSQCRFQ
1HLPHEN CSEDANSIYTDNHQWEITA EHYVHEDGETPISCEFK
1HEPMEN PSEDALARYSDEGAMES CYSERCLECDFH
1CHEFHNEKTPSFTVNGEKQFYHERGACAHENAIDELMNYDAKPARD
1CHEFHNEKTPSFTVNGEKQFYHENEKAHENHIDELMNYDATSARQ
1CHEFHKEKTPSFTVSPDKQFYYCHCHCAGENALGEWMDHDRKEAVN
1NCH FHPDDTPSFYVSPSKQIFK CCCVGCDQIKTVSLYEASE
1LOFFHEETPSFSVSPDKQIFHERCCAGENVFSELRQMEGQEALD
1LEFFHDEKSPSFSVSPAKQMYYCFGCCAGENAFNFLMELGQEALTY
1HSGLIGITTEEGGNSGEKSLSIAPDGSSATWNCFRGKCGLKGGVRADCDEIQDY
1 LKNRHQP WWWGG DRFRFDREGRETWWWWQGAGDGLKLVEKVFGTGTGN-
1 TLKY FPPHKYKYDNMYKHEIFKNTCNSKEHRKCYKESFYDEKLKENARNYL-
1 WWGDWKTDQNKARGYYGDNNEWIHWNWNYHAPIGIWWKEFEPDLY

Zinc-binding motif (1)

phiYeO3-12	47 DITARCHERDICOMARYWLMKVRMYQGADYRIONESIWSCKVRDKD-IN
Т3	47 DETARCESKERCOMACYWLEKVRMYQWADYREQNESEWSOKVEDKD-
т7	47 ANTARCINKE COMACYWINKVVMYQWADYR ONENIWSOKVEDKD-
gh-1	47 DEPARFNQESICROYCYWVERVVMAQWANYYDAQENNTSOKVEDAT-
P.putida	44 PEMKEKLTLE CREAGYFVEEVRLVQAPYFINS VEVACKLEDQD-C
RP	39 RUPS NETER CREANWRVNDGQFPYYUNSRQUWGYKHEGKE-
S.typhimurium	47 Y QKIGH ABIIQK FALCE PPPNS-DNKALLLDA MAYNE-EQGST-Y
E.coli	47 YEKUGLEHEVIAKIALEFEPPNP-ENRQSLIDAEMAWTN-DQGRS-Y
P.aeruginosa	47 YEKGROLUGEIARD GLOEF PPDN-LQLKAMLDAGLEVENSDTGKR-Y
Aquifex	44 YEKS G DPKVARKEDLEY PSEN-DELEAYLETKNELSP-TEGVY-E
B.subtilis	47 YELSHEFEKELINE QICYELDRG-FSEAQMEKAELGIRREDGSGY-F
Synechocystis	47 LDQKECKSAHIQE QLCYMPAQKRYPVAAVEQACLEKARQSGTGY-Y
Arabidopsis	54 -FAARAISRKTLEENRYMQKRIDEIVHAFTYWQREENVSCKYBSLT-
P4	52 - T- KGFPGRECEMLT THRAGAGDLVVPLYDDSEE WNLQLISADGE
PfPREXpri	55V-BKINIELKKELKEFEVMEKIRSEKDKGYMRLYPNVRNEME
т4	50 - FEIRKEKGKSRPIEKPKELPPSCVRLDKLAEDHP

	2	3
phiYeO3-12	95 KTTGSHKSDALFLK	TEGET DALTVMELQDCKYPVVSLG
T3	95 KTTGSHKSDALFLKHIWS-GENKIVV	TREE INALTWMELQDCKYPVVOLCH
Т7	95 KTTGSHKSDALFGKHIWN-GGKRIVV	DISCIDION MELQDCKYPVVSLCH
gh-1	95 PETAGAHDKDALFGROWS-CORE	TEGHTICLINVAQLQGGKYPVVSIGH
P.putida	92 AILGDGAK TLFGQN WAS GCKKIVV	TEGALEAMSVSQVQNNKWPVVSLPN
RP	83 RWQGKNPDNQLFGQQWWC-KBA	TECH FRAMSWFQARKN-WPVVSVPN
S.typhimurium	93 RERNRVMFPERDKRGRVICFCF-RILLV	VSCYMOVVALAQYDINYAVAS
E.coli	93 RERERVMFPERDKRGEVICFCC-RELV	Vac YMDWYARAQYGINYAVAS
P.aeruginosa	94 RERDRVMFPERDSRGRIIAFE -EINV	VECYNWTALAQQGIRNAVAT
Aquifex	90 LELRRVVIPIKDPRGEVICFSCPFALL	VECYFILM REFSEGIRNVVIE
B.subtilis	94 RERNRVMFPEHDHHGAWVAFSE-BAVE	FSCFADVYNAVSSDVKESIAT
Synechocystis	95 OFRHRLMIPERDVQGETLAFS-EAD	VECYFEV A HESGIKQTVA
Arabidopsis	101 DOERKTRR LYGLDD EKTSEVIT	VEGET KLAVEEAGFLNCVSVPDGA
P4	100 RTLKGGQVRGTCHTLEGQNQACKRIWI	ASCYATAMTWHHLTGETVMVALSSV
PfPREXpri	98 PKNVRNEMKLFFFGDH IKNSEE	TEGEIRAN
т4	98 KEKYLWFTTEWPKLVNSIAPETYEKEI	SEPREVEPIYNANGKAES

phiYeO3-12	146	GASMAK TCAANYENFDOEEQ
тз	146	GASAAK TCAANYENFDAHEQ MANNA MUDAGHKWVEEAAQW PA
T7	146	GASMAKETCAANYENFDOREQUITIME MEAGHKEVEEDAQWLPA
gh-1	146	GSKMAKETCASNYENFDTEDELELMERMEDAGEAMSQEMAEVI PP
P.putida	144	GAPMAR AIQRNETLESTEEVELAND MEPHERAQECAEFFSP
RP	133	GAQSAK ALAAQTILLNDEE ALAAD NIDAGIQAAEECVS FPA
S.typhimurium	140	LGTSTTADHMHMFRATENVICCM GERACEDAWRALETAMPYM
E.coli	140	LGTSTTADHIQLFRATINWCCWGRAGEDAWR LETALPYM
P.aeruginosa	141	LGTATSEEHIKRFRLVPSINFCHIGIOAGHKAAWRALESVLPNL
Aquifex	138	LGTALTQNQANLSKFTKKWYILYIG DAGEKAMKSAIPHALS
B.subtilis	141	MGTSLTDDHVKIWRRNVEEWRRNVEEWWW.SWKAQYEWTLKWSEWWQK
Synechocystis	142	LCIPLSEDQVQSMRFSSSKQLEFNELAKACINGTQRAIQE EPLV
Arabidopsis	151	PAKVSSEIPSEDKDTKYKFLWN-CNDYLKKASRIVIAT GEGPEQAMAEEIARRIGKER
P4	152	NLLELASLARQKHPACQIVLAADRD SGDGGKKAAAAADACBGGGEATRKAIYDALRPPA
PfPREXpri	133	-TISQETKYPAISLPNGSKSLPIYL PULEREKKUH WL FIKAGASSVFNFVNKUGLG-
Τ4	144	QCRULK DAPQKYITIEAN PEATKI GVERVKDGDVYVLEGP DS
phiYeO3-12	191	GKVAVAVAT-YZDANCHIMGED-
---------------	-----	---------------------------------
T3	191	GKWEVEVEL-CHEMECHIMGED-
т7	191	GKWEVAVINE-CEDANECHLNGHD-
gh-1	191	GKWEINVIN -FIGUNNCCWEQGNA-
P.putida	189	GKCEINTS-MEDANELEVAGRE-
RP	178	DR FT FDGY DAS AWSAKDY-
S.typhimurium	185	TDGRQ-VEFMFHDGE POTLVEKEGK-
E.coli	185	TDGRQ-INFMF DGE PETLWEKEGK-
P.aeruginosa	186	QDGKR-WEFLFEGE PESLWEAEGE-
Aquifex	181	-AGVE-WYPVY EGY POPFICEFGK-
B.subtilis	184	-KGCK-WEVAMIEDGLEPDEYIEKFGG-
Synechocystis	189	YSGQVNEF ILN HAGNDADE FILLSSAEN
Arabidopsis	210	CWRVKWPEKEEDEHF DAN VIMSKGP-
P4	212	ESPFDTOSEAEFSAMSTSEKAMFIYEHY
PfPREXpri	191	RTNVITOANVHYLNPDVF
Τ4	189	LFEENCIAITGGQLELEEVPFKD-

Figure 6.4. Alignment of deduced amino acid sequence of DNA primases. Details of sequences used are shown in Table 6.3 alongside PfPREXpri. These sequences were aligned using ClustalX 1.81. Black shading indicates identical amino acids; grey shading indicates conserved amino acid substitutions. The numberings on the left hand side are the relative positions of the amino acid in the respective sequences. Motifs 1-6 are underlined with text.

тз	1	AVGLLFDGCQGLMDRTLEARGEEVVMVTSGSCMCKSTFVRAQALAWGKRMG
phiYeO3-12	1	AVGLLFDGCQGLNDRTLCALGGDVINVISGCCMCHSH5VRQQALAWGKRMG
т7	1	SVGLLFSGCTGINDKTLIANGGNVAIOVAISGSCMOLSIPWRCQALQWGTAMG
gn-1	1	DAEVDEAAHERKAMIKDOGGAAVITAUSGGESBEGAAVARANVINDERNES
Mmtwinkle	1	ACVEWSEEPD WELLK HEK TWEEPUSSE WESSELLDLCTOGV
Hutwinkle	1	AGLRWSRFPD RILK HEKEN TWFNCPACS HARDSEYALDLCSOGV
D.melanogaster	1	NGVKWKRFPV MKLLKEH REPUT THE PHOES SEAN SEYSLDLAMOGV
PfPREXheli	1	NGVKSKTIPS WKYLYCL MELSIW CSDOVELT LLSCLSLDYCIQGV
C.elegans	1	VVGFGQWKRFAVILKYLGELLPGDWTVLHCGRCNHLTHFLCEYSLDLFTQGV
E.coli	1	TGYDDDUKKTACLQPSDILLWAARPSMSKTTEAMULVENAAMLQD
S.typhimurium	1	TGYQDUNKKTACLQRSDDDUWAARPSMHRADBAMALCENAAMLQD
B.subtilis	1	TGFTELLRMTACFQRNDIGLWAPRPSVBMTAPALLIAQNVATKTD
P22	1	TGIEEMAITGOMNAEMAARPMMIMELALKIAEGVASRVIPGSGVR
HUHEL308	1	LVFCPSKKNCENMAEMICKFLSK YMKHKEKEKCEVIKN KKIGNGNLCPVLKR
MMHEL308	1	
Arabidoneis	1	HEVCUSTCWKNI.DNI.YSVUPED TWY OF DNSH SEEDDAMLCNI.NHSVC
T4	1	KITKGGAETGTI.NVI.MAEVNWGKSI.GLCSI.AADVOLGHNVI.YI.SME
A.aeolicus	1	RSMAKWIKLOEKKKELYEKAKKE EKANNKILSROVVCTTUSTAGSEVLON
S.cerevisiae	1	ILKKYGIPYOKLKSOPDWMDDL IOF LDLLKVCSLASDEKHNREFNTGDKWOSNF
		Walker A (H1)
		walkel A (III)
Т3	52	-KRVGLAMLEESVEDTIQDMMGLNNKVRLRQSDEVKKAIAED
phiYeO3-12	52	-KRVGLAMLEESVEDTIQDVMGLNNKVRLRQSDEVKKAIAED
Т7	52	-KKNGLAMLEESWEETAEDWIGLHNRVRLRQSDSLKREIIEN
gh-1	50	-IPCGVAMLEEAWEETVQDUVGLHIGARVRQNPDETTE
RP	48	-QRWGYIALEESWKRTGLRIMTVAANKPLHLNNELPT
Mintwinkle	51	
Hutwinkie D. molanogastor	51	
D.Meranogaster DfDRFYholi	51	-STIWGSFEIRNIRLAAILIRQIVGIP
C.elegans	53	-RTTFCSFEMPEKKTLHWM VOYAGYDDLVFFLFPSHLRSLARTNSYKNGTKLPLHRVEY
E.coli	46	KPVI IFSLEMPSEOIMMRS ASLSRVDOTKIRTG-
S.typhimurium	46	KPVHIFSLEMPGEOIMMRMHASLSRVDOTRIRTG-
B.subtilis	46	ESVAIFSLEMGAEQLVMRM CAEGNINAQNLRTG-
P22	52	RGVIIFSMEMSATQVVERGLAGAGMMSVSVLRNPS
HuHEL308	55	TIPFGVAYHHSGUTSDERKULEEAYSTGVLCLFT
MmHEL308	55	TVPFGIAYHHSGUTSEERKULEEAYSTGVLCLLT
P.putida	48	-ERVGTIFLEQKPTETAKRVAGKIAGKRFHVPKDTAGW
Arabidopsis	51	-WKFALCSMENKWRDHARKWLEKHIKKPFFDADYGRSVQ
T4	48	-MAEEVCAKRIDANMLDVSMDDIDDGHISYAEY
A.aeolicus	53	LNFDVVIIDEATQATEPSCHIPLIKGKKLIMAGDHKQLP
S.Cerevisiae	57	SILWIMSALKGIGDASIQAWIKACSLKNLSIWKILIMVPNFEWP
		Hla
m 3	02	
nhiYe03-12	93	GREDWUDELEGDDTEHLYDSFAEADADRLHAKHAYMPTHIGCDVHVLHHSIWVS
T7	93	GKEDOWEDELEGNDTEHLYDSFAEADTDRLWAKWAYMRSGLGCDVUUUHHISIWUS
gh-1	87	EVFDRAGDEIFESDKI FUYDAFAEAAEDRLIAKI AYMVEAEGCRVI VIII HUSIWVS
RP	84	DELRTANDSTLGTGRWFTRDGFGSVTPDSLINDTRFLTKAHEVOWLVHHISTLS
Mmtwinkle	81	LDKYDEWADRFEDLPLYFMTFHGQQSIRSVIDT QHAVYVYDVCHVVII NLQFMMG
Hutwinkle	81	LDKYTHWADRFEDLPTYFMTFHGQQSIRTVIDTWQHAVYVYDICHVIIINDQFMG
D.melanogaster	81	LHEFNHWAAEFERLPHYFMTFHGQQPLKPVHEAHEHASYVHDVMHVLLHNWQFMG
PfPREXheli	81	IELFUIYADKFELLPUKFLKFHGSTNIDQVUDAWDYAVYAYDVKHIIIIWNQFULN
C.elegans	112	SNSINSALDRFERSSSALTMLDSEEFMEKSINETAIRIHVENSGIQHVVIINIQFUIN
E.coli	80	QLDDEDWARISGTMGTLLEKRNIYIDDSSGLTPTEVRSRMRRIAREHGGTGLIMTDYLQL
S.typhimurium	80	QLDDDDWARISGTMGULLEKRNMYILDSSGLTPTEVRSRARRIFREHGGLSLIMIDYLQL
B.subtilis	80	NLTEEDWGKLTMAMGSLSNS-GIYIEDTPGERVSEIRAKCRRLKQESG-EGWILEDYFQL
P22	87	RMDDIGWAKVASGMKHLAEL-DVWVVDASRISVEEIRSISERHKQEHPNISIIMADYIGL
MmHEL308	89	CTSTLAAGVNLFAKKUIIKAPYVAKDFLKKNQYKQMIGROGRAGIDTIGESILIIQEKDK
P. putida	09	TOFFLDARUDALCENHUWYDAFCETTEDIUVBYDAUCECTVI WYDAUCEC
Arabidonsis	80	RMSVDEKDEGKKWINDTFYPIRCEMISI.PSIDWULERAKMAVI.PVCTPCIVILEDV
T4	80	KGKMKKREKSTLGR TWKOYPTGGADANTERSILNELKIKKNEVPTLAVDYLCCK
A.aeolicus	92	PTVLSOENOEALSYTEFERLLDLYGEEIYEELREOYRMNKKIMEFSNKMFYEGKETADKS
S cerevisiae	101	LOT ST VVVMENVTENI VENTENDOULOT DEDMET FVUX STANNI NI NETVEOS SDA OS

Walker B (H2)

Т3	149	ASE-ESDERKMIERIMTKLKGFAKSTGVULVVICHIKNPE
phiYeO3-12	149	ASE-ESDERKMIDRIMTKLKGFAKSTGVULVVICHIKNPE
T 7	149	ASG-ESDERKMINNMTKLKGFAKSTGVULVVUCHUKNPD
gh-1	143	AMDGDQDERKTIERMMTKIKAFAKTKNVAVFVICHKNPD
RP	140	GNE-STDERKMILITMTKLRSFVEETGIGMILISHIRRNQ
Mmtwinkle	137	HEQLSSERMAAQDYIWGAFRKFATDNSCHWTLWI
Hutwinkle	137	HEQLSTBRNAAQDYI GVFRKFATDNNCHVTLVI
D.melanogaster	137	VSTFRGEKFFEQDSI AAFRSFATKHNVI VTLVM
PfPREXheli	137	INKFSUIYELQNIAUDKFRSFSTNKNVUUTLVV
C.elegans	170	QGMMADEKSSGLERFHLQDRFWGYMRQLATQNQI HTMWV
E.coli	140	MRVPALS-DNRTLE AEISRS KALAKE NVPWVA SQUN
S.tvphimurium	140	MRVPSLS-DNRTLEHAEISRSHKALAKEHQVPWVALSQUN
B.subtilis	138	IOGSGRSKDNROOEWSEISRELKSIARELOVPWIALSOLS
P22	146	IEKPKAERNDLAHAHISGS KAMAKD KTPUIS SQUS
HuHEL308	149	OOVLELITKPLENCYSHLVOEFTKGIOTUFLSLIGUKIATNLDDIYHFMNGTFFGVQQKV
MmHEL308	149	OOVLELISGPLETCCSHLVEEFTKGIOAUFLSHIG KEKG
P. putida	141	TADEKGSLEOMMKEMAG ANELGI ITE SHITTPE
Arabidopsis	149	ORTPROTETEYVSONLTKIKRESOHHSCHVWEVALPKONO
T4	138	SCRIRVYSENSYTTWKAIAEEURALAVETETVWWTAAOWG
A apolicus	152	VENHTIKDLINPEKEKEIPEPEKSULEPEKVVWFINVRGK
S cerevisiae	161	SLEFKTHLOFMAOVWKVSKSNKPPGISFWKWFWFTYFDOT
0.0010110100	101	UID III IIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
		H3
m 2	188	
15 phiVe02-12	100	
phileos-12	100	
17 mb 1	100	
gn-1	170	
RP Method a bla	179	
Mmtwinkle	1/1	
Hutwinkle	171	
D.melanogaster	1/1	PRKERQEDELTTSSWFTBSKATLEALNWEILODK
PIPREXhel1	170	
C.elegans	210	
E.coli	179	RSLEQRADKRPVNebingEngShEnDAllLindFiryBD
S.typhimurium	179	BSLEQRADKRPVNSDLPESTERIDARLINGFTYPD
B.subtilis	178	BGVEQRQDKRPM03DIRESGSIEDDARIWAFTYRD
P22	184	DVEKRPNKRPTNA DHESHE DANSI MAYEE
HuHEL308	209	LLKEKSLWEITVESLRYLTEKGLLQMDTIYKSEEEVQYNFHITKLGRASFKGTHDLAYCD
MmHEL308	189	LLQKDSCGINEGIECHFRITKLGQASFKGALDLAYCD
P.putida	177	GIPHENGGRUTHRHFIXGNRAMGFWSYFOFGHERD
Arabidopsis	189	HWEGGAPNLYDISHMAHFIRKCING INHEN
Τ4	178	QAWESSDWNWSDUAE GUPATAE FOLAWIET
A.aeolicus	192	EKQRRGSTSFYNEEEARVWVKEVEYLMKEGLRSEH
S.cerevisiae	201	MVFHQSQQALQTTGPGTVK STIHSAKGLEFP FTTNG
		H4
		114
T3	223	QQGDMPNLVLVRLLLCRFTCDTGIAGYNEWNRETGWLEPSSYTGGEGEGDTGWTEQD
phiYeO3-12	223	QQGDMPHLVLWRLLLCLETEDTGIAGYMENNETGWLEPSSYTGEEGEGDTGWTEQD
T7	223	QQGDMPNLVLWRIINCHETCOTGIAGY@EWNKETGWIEPSSYSGEEESHSESTDWSN
gh-1	218	QQGAFPHIILFRVLIC FTCETGVAGFWRWDKATGRLEPMPEGWKPEDTSGADEAWK
RP	213	DISKGDNRSQURVILIN IN QTGPAGGUSUSVETGROIEALDFEDDKPTAPTDYG
Mmtwinkle	206	KLVTGPGKRYLQVS NEDCDVGVFPLEFNKNSTTFSIPPKSKARLKKIKDDNG
Hutwinkle	206	KLVTGPGKRYLOVSUNDOCDVGVFPLEFNKNSUTFSIPPKNKARLKKIKDDTG
D.melanogaster	206	RLTSVRGKKYHQUAHNDWSCDLGIOPLEFDKDGHSYSTQIQNAKRKREKTPSEN
PfPREXheli	204	VSKTNETVFFIDIKINFIKCSLGKIPULYNKENTIKEISINNFNEHVVSNTYL
C.elegans	245	RDDRDRSKFRK-FYTLINEYY RRVESDOLEMVFNPSTYSHTVVEFPNVIIOFYFLNIF
E.coli	214	EVYHENSDLKG-IAELIIGKORNGPIGTVRUTENGOWSRFDNYAGPOYDDE
S.typhimurium	214	EVYHENSDLKG-IAELIIGKORNGPIGTVRITENGOWSRFDNYAGPOYDDE
B.subtilis	213	DYYDKETENKN-I EI HANORNGPVGTVSMAHVKEYNKFVNLERRFDDAGVPPGA
P22	219	AVYDENSSAAP-FAEL VTENREGSLGTVYOR CNGHEVACDODEAROICTASNAPAGRR
HuHEL308	269	ILYRDLKKGLE-GWESLLHLIYLTTPYDWSOCNPDWETYFROFSOLSPAFONVAATL
MmHEL308	226	TLYRDLKKGLE-GRVNESLLHLIYLTTPYDDAAOSEPDWWVYFKOVTFOVSWAFTSLDFV
P. putida	211	OOAEDPVVROTTTFRILLDEWT OATGEVLYLAVDRDTGL.SLTEAPEDSSPEKDESEF-
Arabidonsis	220	RDENAGPLDLVOIGVRKVRNKVHGOIGDAYHCHDRTTGSVSDSDUTDCMDEDBSDVPV_
T4	211	EELAAAEOOLIKOWKSRYGDKNKWNKFLMGOKGNOKWVEIEODSTPTEVNEVAGSOOTO
A.aeolicus	227	IGVISPYEDOVNELEELKDEEVEVKTVDGFOGREKEVT ISFVRSNEKGFIGELKDVPP
and a first har and and har had had		and a second

Figure 6.5. Alignment of deduced amino acid sequence of DNA helicases. Details of sequences used are shown in Table 6.4 alongside PfPREXheli. These sequences were aligned using ClustalX 1.81. Black shading indicates identical amino acids; grey shading indicates conserved amino acid substitutions. The numberings on the left hand side are the relative positions of the amino acid in the respective sequences. Motifs H1-H4 are underlined with text.



Figure 6.6. Phylogenetic trees generated by MEGA derived from the DNA primase alignment (Figure 6.4). The bootstrap values (100 resamplings) are indicated on the branches. Unrooted trees were generated using the distance method UPGMA (a), the NJ method (b), ME method (c) and MP method (d).



;



Figure 6.7. Phylogenetic trees generated by MEGA derived from the DNA helicase alignment (Figure 6.5). The bootstrap values (100 resamplings) are indicated on the branches. Unrooted trees were generated using the distance method UPGMA (a), the NJ method (b), ME method (c) and MP method (d).

ie.

6.4. Discussion

Work in this chapter aimed to seek the evolutionary origins of PfPREX - agene that encodes several proteins likely to be involved in the replication of plastid DNA. *Plasmodium* has a tendency to fuse genes of related functions (Section 1.4.1) e.g. DHFR-TS, ODC-AdoMetDC, G6PD-6PGL etc. In the case of the DNA primase and DNA helicase domains, there are precedents where these two proteins are fused. This is the case in T-odd bacteriophages which belong to the Podoviridae group. In PIPREX, the phylogenetic analysis points to both the DNA primase and DNA helicase domains branching with similar proteins from this group of T-odd bacteriophages. Therefore, it seems likely that the DNA primase and DNA helicase domains were fused prior to their appearance in PfPREX. Intriguingly, the T7 DNA helicase protein is also found in many eukaryotes - where in some of them, this protein localises to mitochondria (Spelbrink et al, 2001). Leipe et al (2000) looked in detail at the evolution of the DnaB proteins in Eukaryota. In that study, they observed that the DnaB found in chloroplasts are highly similar to the bacterial ones. Therefore, it was proposed that these DnaB sequences had been vertically inherited via the bacterial endosymbiont – the origin of the plastids. In addition, their phylogenetic analysis found that eukaryotic DnaB homologues grouped with the T-odd bacteriophage proteins. This is the same scenario observed in this chapter (Figure 6.7). However, unlike the results reported by Leipe et al (2000), this relationship observed in our phylogenetic trees, obtained strong statistical support (Figure 6.7). However, the study of Leipe et al (2000) was done prior to the discovery of the other T7-like mitochondrial associated eukaryotic sequences. Filée et al (2003) carried out a similar study. In their study, it was observed that these T7-like eukaryote DNA helicases form a sister group to the T7/T3 bacteriophage DNA helicases. Incidentally, Filée et al (2003) also included the P. chabaudi PomI sequence

201

that corresponds to the PfPREXheli sequence in their phylogenetic analysis and the same branching was observed whereby PfPREXheli sat within the T7-like sequences (Twinkle), forming a sister group with T7/T3 bacteriophage sequences. However, the Twinkle-like sequences are found to be mitochondrial and as mitochondria were derived from a α -proteobacterium, mitochondrial DNA helicase might be expected to more closely related to the bacterial ones if they were derived from the DNA helicase of the endosymbiont. The results obtained indicated that these eukaryote Twinkle sequences originated from a T7/T3 bacteriophage rather than a bacterium. This scenario is similar to that of mitochondrial RNA polymerase which also has a bacteriophage origin (Hedtke *et al*, 1997).

One sequence found in the databases is of particular interest. An *Arabidopsis* sequence that Leipe *et al* (2000) also included in their analysis, has the same domain architecture as with PfPREXpriheli and the phage proteins, whereby the DNA primase and DNA helicase domains are fused together. Leipe *et al* (2000) saw this conservation of architecture and proposed that it was due to a horizontal gene transfer event via a bacteriophage. Like the PfPREXpriheli sequence, the *Arabidopsis* sequence contains an N-terminal target sequence and is predicted by various programmes ranging from ChloroP, MitoProt and Predotar to have potential chloroplast and potential mitochondrial localisation. We were interested to find out if the DNA primase and DNA helicase trees (Figure 6.7), the *Arabidoposis* sequence belonged to the T-odd bacteriophage clade whereas, PfPREXheli clustered together with the cukaryotic T7-like DNA helicases. The DNA primase domains of both, on the other hand, clustered together with the T-odd DNA primase sequences, but with weak statistical support (Figure 6.6). These

results seem to suggest that the DNA helicase domain of both *Arabidopsis* and *PfPREX* (and possibly the primase domains) were obtained at different points in evolution. Both sequences appear to have originated from T-odd bacteriophages.

The fact that the Twinkle proteins are associated with the mitochondria in most species in which they are found, indicates that the DNA primase-helicase is ancestrally mitochondrial. The fact that the targeting sequence at the N-terminus of PfPREX directs proteins to the plastid, indicates that a role in plastid DNA replication was acquired secondarily in apicomplexans. Genes originally involved in mitochondrial functions, that later evolved separate functions are common and a T7-like RNA polymerase initially involved in mitochondrial transcription is also known to have taken a secondary role in plastid transcription in plants (Hedtke et al, 2000). These data seem to give rise to two possible hypotheses for the presence of such sequences in eukaryotes. The first h is that a horizontal gene transfer occurred between a phage and either the mitochondrion or the nucleus of a eukaryotic cell that was the ancestor of all of those species carrying Twinkle-like helicase today. Given that it appears that T7-like genes have entered eukaryotic cells on more than one occasion (hence the divergence between the Arabidopsis helicase/primase and PfPREX), it seems possible that horizontal gene transfer from bacteriophages to eukaryotic cells has occurred more than once. A second hypothesis is that the α -proteobacterium that initially gave rise to the mitochondrion, was itself infected with a bacteriophage whose remnants can now be traced in these fossils. In the case of the apicomplexans these genes secondarily gained plastid-targeting sequences and are now involved in the plastid DNA replication instead.

It seems probable that uniquely in the Apicomplexa, the DNA primase-DNA helicase domains fused into a single gene with a DNA polymerase. Extraordinarily, the DNA polymerase domain is most closely related to bacterial DNA polymerase I, and most closely to thermophilic DNA polymerase I of *Aquifex spp*. The *Plasmodium* DNA polymerase is also rather thermophilic (Section 4.2.3.1) and like the *Aquifex* proteins, lacks 3'-5' exonuclease domain. It seems probable that this gene entered the ancestral apicomplexans by horizontal gene transfer (the same DNA polymerase is found in other apicomplexans e.g. *Toxoplasma* and *Theileria*) – although the point in time and mechanisms of this event remains elusive.

í.

Chapter VII

General Discussion

1000

A gene encoding a plastid DNA replicase, PfPREX, from Plasmodium falciparum has been cloned and expressed in E. coli. The full-length protein is composed of three functional domains, a DNA primase, a DNA helicase and a DNA polymerase. The Cterminal domain encodes a DNA polymerase, with 25-30% identity to prokaryotic DNA polymerases from family A. Homology searches identified the three motifs, A (DFKQIELK), B (AINFGLIYGM) and C (LCVHDE) conserved among these family A DNA polymerases (Delarue et al, 1990) in PfPREXpol. In addition, it possesses a 3'-5' exonuclease domain common to this class of enzyme, but no 5'-3' exonuclease domain is found on the same enzyme. The N-terminal domain encodes both a DNA primase and a DNA helicase. The DNA helicase part of PfPREX has the highest similarity to a protein called 'Twinkle', which is a eukaryotic mitochondrial T7-like DNA helicase. This domain contains five key motifs (Ilyina et al, 1992), including the Walker A and B motifs, which are found in all DNA helicases and NTP-utilising enzymes. The DNA primase domain revealed less homology to other DNA primases, however all of the key DNA primase motifs identified elsewhere (Ilyina et al., 1992) can be located within the PfPREXpri domain.

Since the *PfPREX* gene was over 6 kb in length and possessed a very high A-T content, it was foreseen that attempts to clone the gene in its entirety in *E. coli* would prove difficult, if not impossible. Therefore, the gene was cloned in two separate parts, one containing the DNA helicase and DNA primase domains and the other containing the DNA polymerase domain. Initial attempts to clone the DNA polymerase domain into an *E. coli* expression vector, frequently used in the laboratory, pET-16b (Novagen), failed. Next, the

DNA polymerase domain was cloned into a Strep-tag vector, pASK-IBA (Sigma). The reason for choosing to use this expression system was because of the success of another *Plasmodium* protein (ornithine decarboxylase, s-adenosylmethionine decarboxylase) using this system (Müller et al, 1999). Various approaches were taken e.g. introducing the streptag as both an N-terminal and a C-terminal extension, in the hope that either would allow the expression of the recombinant protein. This did not work, thus another approach was tested, whereby an additional plasmid was introduced into the E. coli expression strain. This plasmid (RIG) (Baca & Hol, 2000) carries the genes that encode three tRNAs (Arg, Ile, Gly) cloned from E. coli. These genes direct the constitutive expression of tRNAs that recognise the codons AGA/AGG (Arg), ATA (IIe) and GGA (Gly). Since these codons are used more frequently in *Plasmodium*, the increased levels of the three tRNAs may help E. coli to better translate the *Plasmodium* mRNA and thus increase the yield of the recombinant *Plasmodium* protein. However this approach did not increase the yield of the expressed protein. Finally, the Gateway Cloning system (Invitrogen) was used for cloning *PfPREXpol.* The DNA polymerase domain was successfully cloned and expressed to high levels using the pDEST17 vector (Chapter 4). The DNA helicase/DNA primase domain was also functionally expressed using the pDEST17 construct but only a low yield of the recombinant protein was obtained with this system. The construct was moved into pET-DEST42 vector which introduced a C-terminal His-tag as opposed to a N-terminal His-tag, and a pDEST15 vector which had a N-terminal GST-tag. Overall, the pDEST15 construct gave the highest yield of protein and was the final construct used for the production of the recombinant PfPREXpriheli.

207

PCR experiments using a variety of oligonucleotides from within the sequence of each domain confirmed the identity of each construct. Attempts were also made to sequence each construct. However, with the Gateway vectors, the presence of the att recombination sites which contain large inverted repeats, posed a problem for sequencing and the reason being that the inverted repeats tend to form large loops, which prevents sequencing polymerase from progressing. More than ten sequencing attempts on each clone were made to try to sequence the final constructs. The constructs were even shuttled from one vector to another vector (as recommended by Invitrogen) in efforts to eliminate the sequencing problems, but to no avail. Finally, only the pDEST15priheli construct was successfully sequenced and its sequence was shown to be identical to that in the database. The pDESTpol could only be partially sequenced (approximately 2 kb), and the sequence obtained was as predicted. I also tried to re-amplify the DNA polymerase domain from the pDEST17pol construct to eliminate the att sites and subclone into pGEM-T vector (Promega) for sequencing. The problem encountered with this construct was the same as the Gateway vectors, whereby only the forward primer gave any sequence reads, while the reverse primer was read back into the vector sequence. Regrettably, we ran out of time to sequence the remaining constructs although PCR amplification consistently indicated that they contained the expected inserts.

The study was able to demonstrate that *PfPREX* does encode for functional DNA polymerase, DNA helicase and DNA primase activities. All of these activities were characterised from the corresponding recombinant proteins. First, the PfPREXpol was shown to possess both DNA polymerase activity and 3'-5' exonuclease activity, while the

PfPREXpriheli recombinant protein was found to possess both DNA primase and helicase activity. The characterisation of all activities was limited and future work should focus on characterising each activity in more detail. Given the probable essential role of each protein in replication of the *Plasmodium* plastid DNA – an ultimate goal would be to seek specific inhibitors of the proteins that could represent lead compounds in chemotherapy of malaria.

All of the Plasmodium homologues of PfPREX found in other Plasmodium species possess putative plastid targeting sequences, which suggests that the PREX protein is targeted to the plastid. Therefore, experiments were designed to confirm this. Unfortunately, the indirect immunofluorescence method could not be optimised to give a clear answer to the localisation of the PREX protein in Plasmodium. DAPI was used to stain the nuclear DNA and also the plastid DNA, but due to the close proximity of the nuclear DNA to the plastid DNA, it was impossible to differentiate these two different DNA structures. An *in situ* hybridisation method was used to localise the plastid DNA and to see if the anti-sera that binds to PIPREX protein would give a staining pattern that overlaps the staining pattern of the plastid DNA. Therefore, due to the limitations of both methods, no conclusions could be drawn from the data generated. Hence, a collabration was set up with Dr Shigeharu Sato (NIMR, Mill Hill) who devised a method using GFP technology (Sato, et al, 2003) to target both the mitochondrion and the plastid in Plasmodium. He used the same method to make GFP constructs containing the plastidtargeting sequence of *PfPREX* (the first 120 amino acids) to see if the sequence does localise GFP to a plastid location. Figure 7.1 shows the results obtained in his experiments.



Dr Sato also carried out two sets of control experiments. In the first of these, the leader sequence from the HSP60 protein from Plasmodium was cloned into a GFP construct and transfected into Plasmodium. Figure 7.1A shows the result of that experiment. One can observe the overlapping of two fluorescence signals, the red belonging to the Mitotracker and the green to the GFP, which merged to yield a yellow hue. This was indicative that the HSP60 leader sequence was directing GFP into the mitochondrion. The second set of transfectants contained the GFP constructs with the ACP leader sequence (Figure 7.2B). This time, the red and green fluorescence were observed as separate entities. In addition, the green fluorescence resembled that of an organellar localisation. Furthermore, ACP has been previously shown to localise to the plastid (Waller et al, 1998). The transfectants containing the GFP constructs with PtPREX leader sequence were observed with the same fluorescence pattern as those with the ACP leader sequence. Therefore, the PfPREX leader sequence is concluded to direct GFP into the plastid as with the ACP. In addition, Dr Sato has obtained double transfectants containing GFP with the ACP leader sequence and red fluorescent protein (DsRed) with PfPREX leader sequence, which shows that PfPREX does co-localise with ACP and goes to the plastid (data not shown).

The functional characterisation of the different subunits and the presence of the plastid-targeting sequence at the N-terminus of PfPREX, indicate that *PfPREX* encodes several key components of the replication machinery of the plastid. One might assume that since these replication proteins are associated with the plastid, they could have been derived from the primary endosymbiont from which the plastid is evolved. The primary

endosymbiont is generally accepted to have a eubacterium. Therefore, we would expect the DNA helicase, DNA primase and DNA polymerase of PfPREX to have originated from a eubacterial source. However, the results of the phylogenetic analysis indicate otherwise and have cast an interesting light on an alternative origin of the PfPREX gene. Not only was the origin of the DNA polymerase domain different from that of the DNA helicase and DNA primase domains, the results indicated a rather complex origin of the gene, which might have involved more than one gene transfer event. First, the DNA polymerase domain was shown to be most closely related to the Aquifex DNA polymerase. Aquifex is a thermophilic bacterium belonging to the Archaebacteria group. Not only is the sequence most closely related to the Aquifex protein, both also lack the 5'-3' exonuclease domain and the *Plasmodium* protein is optimally active at high temperature (75°C) as is that of the Aquifex protein. The likely explanation of this relation is that a horizontal gene transfer (HGT) event had occurred such that the Aquifex-like DNA polymerase was introduced into the apicomplexans and replaced the original plastid DNA polymerase by non-orthologous gene displacement. This event must have happened relatively early on in the evolution of the apicomplexan ancestor as the PREX homologues containing this DNA polymerase domain. can be found in several other apicomplexans and among *Plasmodium* species. However, it is not possible to pinpoint where and when exactly the HGT occurred. Also noted is that plants e.g. Oryza (rice), appear to possess a plastid DNA polymerase of different origin as shown in Figure 6.2. The Oryza plastid DNA polymerase though related to the family A DNA polymerase did not cluster with PfPREXpol.

Even more interesting, the phylogenetic data obtained with the DNA helicase and DNA primase domains of *PfPREX* suggest that these two appear not to be typical eubacterial descendents. Instead, both the DNA primase and DNA helicase were shown not to cluster with the bacterial homologues, but to cluster most closely with the T-odd bacteriophages. This situation is similar to the case of mitochondrial RNA polymerases (Hedtke et al, 1997). Therefore, it was proposed that the original mitochondrial RNA polymerase was replaced by the bacteriophage one (Forterre et al, 1999). Furthermore, the T-odd bacteriophage-like RNA polymerase was observed to be widespread throughout the eukaryotic lineage (Cermakian et al, 1996). Other recent studies have also noted the close relationship between some eukaryotic DNA primases and DNA helicases with these T-odd bacteriophage sequences (Filée et al, 2003, Leipe et al, 2000), which was also observed in this study. As discussed in Chapter 6, the phylogenetic data obtained support the proposal that along with these eukaryotic sequences, PfPREXpriheli was derived from a T-odd bacteriophage origin. Two hypotheses can be put forward (Section 6.4) to try to explain the presence of such sequences in *Plasmodium*. The first involves a HGT event involving a phage and the eukaryotic ancestor. The second suggests that the original α -proteobacterium that gave rise to the mitochondrion harboured a T-odd bacteriophage, which left behind the DNA primase and DNA helicase as molecular fossils. In this latter case, the T7-like RNA polymerase may have been part of the same phage that infected the mitochondrion. The fact that the DNA helicase domain is also found in many other eukaryotic mitochondria, as with the Twinkle protein, does support an ancient introduction of the phage gene. However, a DNA primase-helicase also related to T7-like primase-helicase but clearly distinct from the PfPREX protein and the Twinkle-like helicases is also present in the Arabidopsis genome and may also play a role in organellar DNA replication. This demonstrates that the T7-like genes have entered eukaryotes on more than one occasion which weakens the second argument that the protomitochondrion carried an infecting phage. Further arguments against this theory include the fact that T-odd phages are known to infect a select group of proteobacteria; but they are not known to infect rickettsial-type α -proteobacteria (Hausmann, 1988). Therefore, it seems unlikely that the protomitochondrion which originated from α -proteobacterium was indeed infected by a T-odd phage. The second argument is T-odd phages are virulent, lytic phages; they infect a host and multiply, which leads to the destruction of the cubacterial host cell. Thus, it is unlikely that a eubacterial endosymbiont harbouring such a phage could have been present in the host long enough to allow it to be integrated into a mitochondrion. Hence, the overall scenario does not support the above theory. However, irrespective of its origin, the PfPREXpriheli had somehow gained a plastid targeting sequence and is now involved in the plastid replication machinery.

Due to the nature of *Plasmodium*'s penchant for combining related genes together, we have here an example, whereby *Plasmodium* has stuck three genes encoding DNA replication proteins altogether. This study has established that due to topological constraint, it is most likely that the PfPREX protein is post-translationally cleaved, in order for each of the replication enzymes to work at its optimum. So far, there are other such multifunctional proteins found in *Plasmodium*, as discussed in Section 1.4.1. Interestingly, Lipps *et al* (2003) recently discovered another such multifunctional DNA replication protein in a thermoacidophile archaeon, *Sulfolobus islandicus*. The ORF904 encodes an

ATPase, DNA primase and a DNA polymerase, functionally similar to PfPREX. They proposed that the ORF904 could be a relict of an evolutionary old but independent DNA replication protein. Another example of a multifunctional DNA replication enzyme is the Mus308 in cukaryotes which possesses both a DNA helicase and a DNA polymerase domain. Given time, one may find more of these types of multifunctional replication proteins spread throughout the various kingdoms of organisms.

Most importantly, the discovery of this gene, *PfPREX* which encodes a plastid DNA replication machinery, has shed some light on the complex evolutionary path that has led to the modern apicomplexan order. The study has also opened the possibility that this protein can be used as a drug target. By disabling the plastid DNA replication machinery, one can kill the parasite and prevent it from continuing its life cycle in the host. It is hoped that this protein will provide a specific target for future drugs used in combating the discases caused by apicomplexan parasites.

Bibliography

Abdel-Monem, M., & Hoffmann-Berling, H. 1976, "Enzymic unwinding of DNA. 1. Purification and characterization of a DNA-dependent ATPase from Escherichia coli", *Eur.J.Biochem*, vol. 65, no. 2, pp. 431-440.

Abdel-Monem, M., Taucher-Scholz, G., & Klinkert, M. Q. 1983, "Identification of *Escherichia coli* DNA helicase I as the tral gene product of the F sex factor", *Proc.Natl.Acad Sci.U.S.A*, vol. 80, no. 15, pp. 4659-4663.

Albà, M. 2001, "Replicative DNA polymerases", Genome Biol., vol. 2, no. 1, p. REVIEWS3002.

Altincicek, B., Hintz, M., Sanderbrand, S., Wiesner, J., Beck, E., & Jomaa, H. 2000, "Tools for discovery of inhibitors of the 1-deoxy-D-xylulose 5-phosphate (DXP) synthase and DXP reductoisomerase: an approach with enzymes from the pathogenic bacterium *Pseudomonas aeruginosa*", *FEMS Microbiol.Lett.*, vol. 190, no. 2, pp. 329-333.

Aravind, L., Leipe, D. D., & Koonin, E. V. 1998, "Toprim--a conserved catalytic domain in type IA and II topoisomerases, DnaG-type primases, OLD family nucleases and RecR proteins", *Nucleic Acids Res.*, vol. 26, no. 18, pp. 420burger5-4213.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., & Sherlock, G. 2000, "Gene ontology: tool for the unification of biology. The Gene Ontology Consortium", *Nat.Genet.*, vol. 25, no. 1, pp. 25-29.

Astatke, M., Grindley, N. D., & Joyce, C. M. 1995, "Deoxynucleoside triphosphate and pyrophosphate binding sites in the catalytically competent ternary complex for the polymerase reaction catalyzed by DNA polymerase I (Klenow fragment)", *J.Biol.Chem.*, vol. 270, no. 4, pp. 1945-1954.

B

Baca, A.M., & Hol, W.G. J. 2000, "Overcoming Codon Bias: A Method for High-Level Overexpression of *Plasmodium* and AT-rich Parasite Genes in *Escherichia coli*", *International Journal for Parasitology*, vol. 30, pp. 113-118.

Bannister, L. H., Hopkins, J. M., Fowler, R. E., Krishna, S., & Mitchell, G. H. 2000, "A brief illustrated guide to the ultrastructure of *Plasmodium falciparum* asexual blood stages", *Parasitol.Today*, vol. 16, no. 10, pp. 427-433.

2

Basu, A. & Modak, M. J. 1987, "Identification and amino acid sequence of the deoxynucleoside triphosphate binding site in *Escherichia coli* DNA polymerase I", *Biochemistry*, vol. 26, no. 6, pp. 1704-1709.

Basu, S., Basu, A., & Modak, M. J. 1988, "Pyridoxal 5'-phosphate mediated inactivation of *Escherichia coli* DNA polymerase I: identification of lysine-635 as an essential residue for the processive mode of DNA synthesis", *Biochemistry*, vol. 27, no. 18, pp. 6710-6716.

Beese, L. S. & Steitz, T. A. 1991, "Structural basis for the 3'-5' exonuclease activity of *Escherichia coli* DNA polymerase I: a two metal ion mechanism", *EMBO J.*, vol. 10, no. 1, pp. 25-33.

Beese, L. S., Friedman, J. M., & Steitz, T.A. 1993a, "Crystal structures of the Klenow fragment of DNA polymerase I complexed with deoxynucleoside triphosphate and pyrophosphate", *Biochemistry*, vol. 32, no. 51, pp. 14095-14101.

Beese, L. S., Derbyshire, V., & Steitz, T. A. 1993b, "Structure of DNA polymerase I Klenow fragment bound to duplex DNA", *Science*, vol. 260, no. 5106, pp. 352-355.

Bernad, A., Blanco, L., Lazaro, J. M., Martin, G., & Salas, M. 1989, "A conserved 3'----5' exonuclease active site in prokaryotic and cukaryotic DNA polymerases", *Cell*, vol. 59, no. 1, pp. 219-228.

Bernstein, J. A. & Richardson, C. C. 1988, "A 7-kDa region of the bacteriophage T7 gene 4 protein is required for primase but not for helicase activity", *Proc.Natl.Acad.Sci.U.S.A*, vol. 85, no. 2, pp. 396-400.

Blakey, R.L., Benkovic, S.J. 1984, Folates and Pteridines, pp. 191-243, Wiley, New York.

Blondal, T., Thorbjarnardottir, S. H., Kieleczawa, J., Hjorleifsdottir, S., Kristjansson, J. K., Einarsson, J. M., & Eggertsson, G. 2001, "Cloning, sequence analysis and functional characterization of DNA polymerase I from the thermophilic eubacterium *Rhodothermus marinus*", *Biotechnol.Appl.Biochem*, vol. 34, no. Pt 1, pp. 37-45.

Boehmer, P. E., & Emmerson, P. T. 1992, "The RecB subunit of the *Escherichia coli* RecBCD enzyme couples ATP hydrolysis to DNA unwinding", *J.Biol.Chem.*, vol. 267, no. 7, pp. 4981-4987.

Bradford, M. 1976, "A Rapid and Sensitive Method for the Quantitation of Microgram Quantities of Protein Utilizing the Principle of Protein-Dye Binding" *Anal. Biochem.* vol. 72, pp. 248-254.

Braithwaite, D. K., & Ito, J. 1993, "Compilation, alignment, and phylogenetic relationships of DNA polymerases", *Nucleic Acids Res.*, vol. 21, no. 4, pp. 787-802.

Bremen, J. 2001, "The ears of the hippopotamus: manifestations, determinants, and estimates of the malaria burden", *Am.J.Trop.Med.Hyg*, vol. 64, no. 1S,2S, pp. 1-11.

Brennan, C. A., & Dombroski, A. J. P. T. 1997, "Transcription termination factor rho is an RNA-DNA helicase", *Cell*, vol. 48, no. 6, pp. 945-952.

Bruce, B. D. 2000, "Chloroplast transit peptides: structure, function and evolution", *Trends Cell Biol.*, vol. 10, no. 10, pp. 440-447.

Burgers, P. M. 1998, "Eukaryotic DNA polymerases in DNA replication and DNA repair", *Chromosoma*, vol. 107, no. 4, pp. 218-227.

Burgers, P. M., Koonin, E. V., Bruford, E., Blanco, L., Burtis, K. C., Christman, M. F., Copeland, W. C., Friedberg, E. C., Hanaoka, F., Hinkle, D. C., Lawrence, C. W., Nakanishi, M., Ohmori, H., Prakash, L., Prakash, S., Reynaud, C. A., Sugino, A., Todo, T., Wang, Z., Weill, J. C., & Woodgate, R. 2001, "Eukaryotic DNA polymerases: proposal for a revised nomenclature", *J.Biol.Chem.*, vol. 276, no. 47, pp. 43487-43490.

Bzik, D. J., Li, W. B., Horii, T., & Inselburg, J. 1987, "Molecular cloning and sequence analysis of the *Plasmodium falciparum* dihydrofolate reductase-thymidylate synthase gene", *Proc.Natl.Acad Sci.U.S.A*, vol. 84, no. 23, pp. 8360-8364.

<u>C</u>

Carlton, J. M., Angiuoli, S. V., Suh, B. B., Kooij, T. W., Pertea, M., Silva, J. C., Ermolaeva, M. D., Allen, J. E., Selengut, J. D., Koo, H. L., Peterson, J. D., Pop, M., Kosack, D. S., Shumway, M. F., Bidwell, S. L., Shallom, S. J., van Aken, S. E., Riedmuller, S. B., Feldblyum, T. V., Cho, J. K., Quackenbush, J., Sedegah, M., Shoaibi, A., Cummings, L. M., Florens, L., Yates, J. R., Raine, J. D., Sinden, R. E., Harris, M. A., Cunningham, D. A., Preiser, P. R., Bergman, L. W., Vaidya, A. B., van Lin, L. H., Janse, C. J., Waters, A. P., Smith, H. O., White, O. R., Salzberg, S. L., Venter, J. C., Fraser, C. M., Hoffman, S. L., Gardner, M. J., & Carucci, D. J. 2002, "Genome sequence and comparative analysis of the model rodent malaria parasite *Plasmodium yoelii yoelii*", *Nature*, vol. 419, no. 6906, pp. 512-519.

Carucci, D.J., Winey, A.A., Muhia, D.K., Warhurst, D.C., Schaap, P., Meima, M., Li, J.L., Taylor, M.C., & Kelly, J.M. 2000, "Guanylyl cyclase activity associated with putative bifunctional integral membrane proteins in *Plasmodium falciparum*", *J.Biol.Chem*, vol. 275, no. 29, pp. 22147-22156.

Caruthers, J.M., Johnson, E.R., & McKay, D.B. 2000, "Crystal structure of the yeast initiation factor 4A, a DEAD-box RNA helicase", *Proc.Natl.Acad Sci.U.S.A*, vol. 97, no.24, pp. 13080-13085.

Caruthers, J.M., & McKay, D.B. 2002, "Helicase structure and mechanism", *Curr.Opin.Struct Biol.*, vol. 12, no.1, pp. 123-133.

Cermakian, N., Ikeda, T. M., Cedergren, R., & Gray, M. W. 1996, "Sequences homologous to yeast mitochondrial and bacteriophage T3 and T7 RNA polymerases are widespread throughout the eukaryotic lineage", *Nucleic Acids Res.*, vol. 24, no. 4, pp. 648-654.

Chang, J. R., Choi, J. J., Kim, H. K., & Kwon, S. T. 2001, "Purification and properties of Aquifex aeolicus DNA polymerase expressed in *Escherichia coli*", *FEMS Microbiol.Lett.*, vol. 201, no. 1, pp. 73-77.

Chavalitshewinkoon, P., de Vries, E., Stam, J. G., Franssen, F. F., van der Vliet, P. C., & Overdulve, J. P. 1993, "Purification and characterization of DNA polymerases from *Plasmodium falciparum*", *Mol.Biochem.Parasitol.*, vol. 61, no. 2, pp. 243-253.

Chen, F., & Lu, J. 2002, "Genomic sequence and evolution of marine cyanophage P60: a new insight on lytic and lysogenic phages", *Appl. Environ. Microbiol.*, vol. 68, no. 5, pp. 2589-2594.

Choi, J. J., Jung, S. E., Kim, H. K., & Kwon, S. T. 1999, "Purification and properties of *Thermus filiformis* DNA polymerase expressed in *Escherichia coli*", *Biotechnol.Appl.Biochem*, vol. 30 (Pt 1), pp. 19-25.

Chomczynski, P. & Sacchi, N. 1987, "Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction", *Anal.Biochem*, vol. 162, no. 1, pp. 156-159.

Clarke, J. L., Scopes, D. A., Sodeinde, O., & Mason, P. J. 2001, "Glucose-6-phosphate dehydrogenase-6-phosphogluconolactonase. A novel bifunctional enzyme in malaria parasites", *Eur.J.Biochem.*, vol. 268, no. 7, pp. 2013-2019.

Cloonan, N., Fischer, K., Cheng, Q., & Saul, A. 2001, "Aldolase genes of *Plasmodium* species", *Mol. Biochem. Parasitol*, vol. 113, no. 2, pp. 327-330.

D

De Vries, E., Stam, J. G., Franssen, F. F., van der Vliet, P. C., & Overdulve, J. P. 1991, "Purification and characterization of DNA polymerases from *Plasmodium berghei*", *Mol.Biochem.Parasitol.*, vol. 45, no. 2, pp. 223-232.

Deckert, G., Warren, P.V., Gaasterland, T., Young, W.G., Lenox, A.L., Graham, D.E., Overbeek, R., Snead, M.A., Keller, M., Aujay, M., Huber, R., Feldman, R.A., Short, J.M., Olsen, G.J., & Swanson, R.V. 1998, "The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus*", *Nature* vol. 392, no. 6674, pp. 353-358.

Delarue, M., Poch, O., Tordo, N., Moras, D., & Argos, P. 1990, "An attempt to unify the structure of polymerases", *Protein Eng*, vol. 3, no. 6, pp. 461-467.

Delwiche, C. F., & Palmer, J. D. 1997, "The origin of plastids and their spread via secondary symbiosis", *Plant Syst. Evol.*, vol.11, pp 53-86.

Derbyshire, V., Pinsonneault, J. K., & Joyce, C. M. 1995, "Structure-function analysis of 3'-->5'-exonuclease of DNA polymerases", *Methods Enzymol.*, vol. 262, pp. 363-385.

Dingman, C.W. 1974, "Bidirectional chromosome replication: some toplogical considerations." *J.Theor.Biol.*, vol. 43, no.1, pp. 187-195.

Dominguez, O., Ruiz, J. F., Lain, d. L., Garcia-Diaz, M., Gonzalez, M. A., Kirchhoff, T., Martinez, A., Bernad, A., & Blanco, L. 2000, "DNA polymerase mu (Pol mu), homologous to TdT, could act as a DNA mutator in eukaryotic cells", *EMBO J.*, vol. 19, no. 7, pp. 1731-1742.

Doublie, S., Tabor, S., Long, A. M., Richardson, C. C., & Ellenberger, T. 1998, "Crystal structure of a bacteriophage T7 DNA replication complex at 2.2 A resolution", *Nature*, vol. 391, no. 6664, pp. 251-258.

Dunn, J.J., & Studier, F.W. 1983, "The nucleotide sequence of bacteriophage T7 DNA and the locations of T7 genetic elements", J. Mol. Biol., vol. 166, no. 4, no. 477-535.

F

Feagin, J. E., Gardner, M. J., Williamson, D. H., & Wilson, R. J. 1991, "The putative mitochondrial genome of *Plasmodium falciparum*", *J.Protozool.*, vol. 38, no. 3, pp. 243-245.

Feagin, J. E. 1992, "The 6-kb element of *Plasmodium falciparum* encodes mitochondrial cytochrome genes", *Mol. Biochem. Parasitol.*, vol. 52, no. 1, pp. 145-148.

Fichera, M. E., & Roos, D. S. 1997, "A plastid organelle as a drug target in apicomplexan parasites", *Nature*, vol. 390, no. 6658, pp. 407-409.

Filée, J., Forterre, P., Sen-Lin, T., & Laurent, J. 2001, "Evolution of DNA polymerase families: evidences for multiple gene exchange between cellular and viral proteins", *J.Mol.Evol.*, vol. 54, no. 6, pp. 763-773.

Filée, J., Forterre, P., & Laurent, J. 2003, "The role played by viruses in the evolution of their hosts: a view based on informational protein phylogenics", *Res. Microbiol.*, vol. 154, no. 4, pp. 237-243.

Forterre, P. 1999, "Displacement of cellular proteins by functional analogues from plasmids or viruses could explain puzzling phylogenies of many DNA informational proteins", *Mol. Microbiol.*, vol. 33, no. 3, pp. 457-465.

÷

Foth, B. J., Ralph, S. A., Tonkin, C. J., Struck, N. S., Fraunholz, M., Roos, D. S., Cowman, A. F., & McFadden, G. I. 2003, "Dissecting apicoplast targeting in the malaria parasite *Plasmodium falciparum*", *Science*, vol. 299, no. 5607, pp. 705-708.

Frick, D.N., & Richardson, C.C. 2001, "DNA primascs", Annu. Rev. Biochem., vol. 70, pp. 39-80.

Friedberg, E. C., Feaver, W. J., & Gerlach, V. L. 2000, "The many faces of DNA polymerases: strategies for mutagenesis and for mutational avoidance", *Proc.Natl.Acad Sci.U.S.A*, vol. 97, no. 11, pp. 5681-5683.

Fry, D. C., Kuby, S. A., & Mildvan, A. S. 1986, "ATP-binding site of adenylate kinase: mechanistic implications of its homology with ras-encoded p21, F1-ATPase, and other nucleotide-binding proteins", *Proc.Natl.Acad Sci.U.S.A*, vol. 83, no. 4, pp. 907-911.

Fuks, B. & Schnell, D. J. 1997, "Mechanism of Protein Transport across the Chloroplast Envelope", *Plant Physiol.*, vol. 114, no. 2, pp. 405-410.

Fuller, R. S., & Kornberg, A. 1983, "Purified DnaA protein in initiation of replication at the *Escherichia coli* chromosomal origin of replication", *Proc.Natl.Acad Sci.U.S.A*, vol. 80, no. 19, pp. 5817-5821.

Fuller, C. W., & Richardson, C. C. 1985, "Initiation of DNA replication at the primary origin of bacteriophage T7 by purified proteins. Site and direction of initial DNA synthesis", *J.Biol.Chem.*, vol. 260, no. 5, pp. 3185-3196.

G

Gaikwad, A., Hop, D. V., & Mukherjee, S. K. 2002, "A 70-kDa chloroplast DNA polymerase from pea (*Pisum sativum*) that shows high processivity and displays moderate fidelity", *Mol. Genet. Genomics*, vol. 267, no. 1, pp. 45-56.

Garcia-Diaz, M., Dominguez, O., Lopez-Fernandez, L. A., de Lera, L. T., Saniger, M. L., Ruiz, J. F., Parraga, M., Garcia-Ortiz, M. J., Kirchhoff, T., del Mazo, J., Bernad, A., & Blanco, L. 2000, "DNA polymerase lambda (Pol lambda), a novel cukaryotic DNA polymerase with a potential role in meiosis", *J.Mol.Biol.*, vol. 301, no. 4, pp. 851-867.

Gardner, M. J., Bates, P. A., Ling, I. T., Moore, D. J., McCready, S., Gunasekera, M. B., Wilson, R. J., & Williamson, D. H. 1988, "Mitochondrial DNA of the human malarial parasite *Plasmodium falciparum*", *Mol.Biochem.Parasitol.*, vol. 31, no. 1, pp. 11-17.

Gardner, M. J., Williamson, D. H., & Wilson, R. J. 1991a, "A circular DNA in malaria parasites encodes an RNA polymerase like that of prokaryotes and chloroplasts", *Mol.Biochem.Parasitol.*, vol. 44, no. 1, pp. 115-123.

Gardner, M. J., Feagin, J. E., Moore, D. J., Spencer, D. F., Gray, M. W., Williamson, D. H., & Wilson, R. J. 1991b, "Organisation and expression of small subunit ribosomal RNA genes encoded by a 35-kilobase circular DNA in *Plasmodium falciparum*", *Mol.Biochem.Parasitol.*, vol. 48, no. 1, pp. 77-88.

Gardner, M. J., Feagin, J. E., Moore, D. J., Rangachari, K., Williamson, D. H., & Wilson, R. J. 1993, "Sequence and organization of large subunit rRNA genes from the extrachromosomal 35 kb circular DNA of the malaria parasite *Plasmodium falciparum*", *Nucleic Acids Res.*, vol. 21, no. 5, pp. 1067-1071.

Gardner, M. J., Preiser, P., Rangachari, K., Moore, D., Feagin, J. E., Williamson, D. H., & Wilson, R. J. 1994, "Nine duplicated tRNA genes on the plastid-like DNA of the malaria parasite *Plasmodium falciparum*", *Gene*, vol. 144, no. 2, pp. 307-308.

Gardner, M. J. 1999, "The genome of the malaria parasite", *Curr.Opin.Genet.Dev.*, vol. 9, no. 6, pp. 704-708.

Gardner, M. J., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R. W., Carlton, J. M., Pain, A., Nelson, K. E., Bowman, S., Paulsen, I. T., James, K., Eisen, J. A., Rutherford, K., Salzberg, S. L., Craig, A., Kyes, S., Chan, M. S., Nene, V., Shallom, S. J., Suh, B., Peterson, J., Angiuoli, S., Pertea, M., Allen, J., Selengut, J., Haft, D., Mather, M. W., Vaidya, A. B., Martin, D. M., Fairlamb, A. H., Fraunholz, M. J., Roos, D. S., Ralph, S. A., McFadden, G. I., Cummings, L. M., Subramanian, G. M., Mungall, C., Venter, J. C., Carucci, D. J., Hoffman, S. L., Newbold, C., Davis, R. W., Fraser, C. M., & Barrell, B. 2002, "Genome sequence of the human malaria parasite *Plasmodium falciparum*", *Nature*, vol. 419, no. 6906, pp. 498-511.

Gibbs, P.E., McGregor, W. G., Maher, V. M., Nisson, P., & Lawrence, C. W. 1998, "A human homolog of the *Saccharomyces cerevisiae REV3* gene, which encodes the catalytic subunit of DNA polymerase zeta", *Proc.Natl.Acad.Sci.U.S.A.*, vol. 95, no. 12, pp. 6876-6880.

Gorbalenya, A. E., Koonin, E. V., Donchenko, A. P., & Blinov, V. M. 1989, "Two related superfamilies of putative helicases involved in replication, recombination, repair and expression of DNA and RNA genomes", *Nucleic Acids Res.*, vol. 17, no. 12, pp. 4713-4730.

Gorbalenya, A. E., Koonin, E. V., & Wolf, Y. I. 1990, "A new superfamily of putative NTP-binding domains encoded by genomes of small DNA and RNA viruses", *FEBS Lett.*, vol. 262, no. 1, pp. 145-148.

Gorbalenya, A. E., & Koonin, E. V. 1993, "Helicases: amino acid sequence comparisons and structure-function relationships", *Curr.Opin.Struct.Biol.*, vol. 3, pp. 419-429.

Gray, M. W. 1993, "Origin and evolution of organelle genomes", *Curr.Opin.Genet.Dev.*, vol. 3, no. 6, pp. 884-890.

Gray, M. W. & Lang, B. F. 1998, "Transcription in chloroplasts and mitochondria: a tale of two polymerases", *Trends Microbiol.*, vol. 6, no. 1, pp. 1-3.

Guex, N. & Peitsch, M. C. 1997, "SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling", *Electrophoresis*, vol. 18, no. 15, pp. 2714-2723.

Η

Hall, M. C., & Matson, S. W. 1999, "Helicase motifs: the engine that powers DNA unwinding", *Mol.Microbiol.*, vol. 34, no. 5, pp. 867-877.

Harris, P. V., Mazina, O. M., Leonhardt, E. A., Case, R. B., Boyd, J. B., & Burtis, K. C. 1996, "Molecular cloning of *Drosophila* mus308, a gene involved in DNA cross-link repair with homology to prokaryotic DNA polymerase I genes", *Mol.Cell Biol.*, vol. 16, no. 10, pp. 5764-5771.

Harwood, J. L. 1996, "Recent advances in the biosynthesis of plant fatty acids", *Biochim.Biophys.Acta*, vol. 1301, no. 1-2, pp. 7-56.

He, C. Y., Shaw, M. K., Pletcher, C. H., Striepen, B., Tilney, L. G., & Roos, D. S. 2001, "A plastid segregation defect in the protozoan parasite *Toxoplasma gondii*", *EMBO J.*, vol. 20, no. 3, pp. 330-339.

Hedtke, B., Borner, T., & Weihe, A. 1997, "Mitchondrial and chloroplast phage-type RNA polymerases in *Arabidopsis*", *Science*, vol. 277, no. 5327, pp. 809-811.

Hedtke, B., Borner, T., & Weihe, A. 2000, "One RNA polymerase serving two genomes", *EMBO Rep.*, vol. 1, no. 5, pp. 435-440.

Hellgren, U., Berggreb-Palme, I., Bergqvist, Y. and Jerling, M. 1997, "Enantioselective pharmacokinetics of mefloquine during long-term intake of the prophylactic dosc", *Br.J. Clin.Pharmacol.*, vol. 44, pp. 119-124.

Hien, T. T., & White, N. J. 1993, "Qinghaosu", Lancet, vol, 341, no. 8845, pp. 603-608.

Horrocks, P., Jackson, M., Cheesman, S., White, J. H., & Kilbey, B.J. 1996, "Stage specific expression of proliferating cell nuclear antigen and DNA polymerase delta from *Plasmodium falciparum*", *Mol.Biochem.Parasitol.*, vol. 79, no. 2, pp.177-182.

Hopkins, J., Fowler, R., Krishna, S., Wilson, I., Mitchell, G., & Bannister, L. 1999, "The plastid in *Plasmodium falciparum* asexual blood stages: a three-dimensional ultrastructural analysis", *Protist.*, vol. 150, no. 3, pp. 283-295.

Howe, C. J. 1992, "Plastid origin of an extrachromosomal DNA molecule from *Plasmodium*, the causative agent of malaria", *J. Theor. Biol.*, vol. 158, no. 2, pp. 199-205.

Huang, Y.P., Downie, J.A., & Ito, J. 1999, "Primary structure of the DNA polymerase I gene of an alpha-proteobacterium, *Rhizobium leguminosarum*, and comparison with other family A DNA polymerases", *Curr. Microbiol.*, vol. 38, no. 6, pp. 355-359.

Huang, J. & Levin, D. B. 2001, "Expression, purification and characterization of the *Spodoptera littoralis* nucleopolyhedrovirus (SpliNPV) DNA polymerase and interaction with the SpliNPV non-hr origin of DNA replication", *J.Gen.Virol.*, vol. 82, no. 7, pp. 1767-1776.

Huberts, P., & Mizrahi, V. 1995, "Cloning and sequence analysis of the gene encoding the DNA polymerase I from *Mycobacterium tuberculosis*", *Gene*, vol. 164, no. 1, pp. 133-136.

Hubscher, U., Nasheuer, H. P., & Syvaoja, J. E. 2000, "Eukaryotic DNA polymerases, a growing family", *Trends Biochem.Sci.*, vol. 25, no. 3, pp. 143-147.

Ī

Ilyina, T. V., Gorbalenya, A. E., & Koonin, E. V. 1992, "Organization and evolution of bacterial and bacteriophage primase- helicase systems", *J.Mol.Evol.*, vol. 34, no. 4, pp. 351-357.

Inselburg, J. & Banyal, H. S. 1984, "*Plasmodium falciparum*: synchronization of asexual development with aphidicolin, a DNA synthesis inhibitor", *Exp.Parasitol.*, vol. 57, no. 1, pp. 48-54.

Ito, J. & Braithwaite, D. K. 1991, "Compilation and alignment of DNA polymerase sequences", *Nucleic Acids Res.*, vol. 19, no. 15, pp. 4045-4057.

Ivanetich, K. M., & Santi, D.V. 1990, "Bifunctional thymidylate synthase-dihydrofolate reductase in protozoa", *FASEB J.*, vol. 4, no. 6, pp. 1591-1597.

Ī

Jeanmougin, F., Thompson, J.D., Gouy, M., Higgins, D.G., & Gibson, T.J. 1998, "Multiple sequence alignment with Clustal X". *Trends Biochem. Sci.*, vol 23, no. 10, pp. 403-5.

Jelenska, J., Crawford, M. J., Harb, O. S., Zuther, E., Haselkorn, R., Roos, D. S., & Gornicki, P. 2001, "Subcellular localization of acetyl-CoA carboxylase in the apicomplexan parasite *Toxoplasma gondii*", *Proc.Natl.Acad.Sci U.S.A*, vol. 98, no. 5, pp. 2723-2728.

.

Jensen, J.B., & Trager, W. 1978, "Some recent advances in the cultivation of *Plasmodium falciparum*", *Isr.J.Med.Sci.*, vol. 14, no. 5, pp. 563-70.

Jomaa, H., Wiesner, J., Sanderbrand, S., Altincicek, B., Weidemeyer, C., Hintz, M., Turbachova, I., Eberl, M., Zeidler, J., Lichtenthaler, H. K., Soldati, D., & Beck, E. 1999, "Inhibitors of the nonmevalonate pathway of isoprenoid biosynthesis as antimalarial drugs", *Science*, vol. 285, no. 5433, pp. 1573-1576.

Joseph, J. T., Aldritt, S. M., Unnasch, T., Puijalon, O., & Wirth, D. F. 1989, "Characterization of a conserved extrachromosomal element isolated from the avian malarial parasite *Plasmodium gallinaceum*", *Mol.Cell Biol.*, vol. 9, no. 9, pp. 3621-3629.

Joyce, C. M., & Grindley, N. D. 1984, "Method for determining whether a gene of *Escherichia coli* is essential: application to the polA gene", *J.Bacteriol.*, vol. 158, no. 2, pp. 636-643.

Joyce, C.M., Kelley, W.S., Grindley, N.D. 1982, "Nucleotide sequence of the *Escherichia coli polA* gene and primary structure of DNA polymerase I", *J.Biol.Chem.*, vol. 257, no. 4, pp. 1958-1964.

Joyce, C. M. & Steitz, T. A. 1994, "Function and structure relationships in DNA polymerases", *Annu. Rev. Biochem*, vol. 63, pp. 777-822.

Joyce, C. M. & Steitz, T. A. 1995, "Polymerase structures and function: variations on a theme?", *J.Bacteriol.*, vol. 177, no. 22, pp. 6321-6329.

Joyce, C. M. 1997, "Choosing the right sugar: how polymerases select a nucleotide substrate", *Proc.Natl.Acad Sci.U.S.A*, vol. 94, no. 5, pp. 1619-1622.

K

Kaneko, T., Tanaka, A., Sato, S., Kotani, H., Sazuka, T., Miyajima, N., Sugiura, M. & Tabata, S. 1995, "Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC6803. I. Sequence features in the 1 Mb region from map positions 64% to 92% of the genome", *DNA Res.*, vol. 2, no. 4, pp. 153-166.

Karam, J. D., & Konigsberg, W. H. 2000, "DNA polymerase of the T4-related bacteriophages", *Prog.Nucleic Acid Res.Mol.Biol.*, vol. 64, pp. 65-96.

Kedzierski, L., Escalante, A. A., Isea, R., Black, C. G., Barnwell, J. W., & Coppel, R. L. 2002, "Phylogenetic analysis of the genus *Plasmodium* based on the gene encoding adenylosuccinate lyase", *Infect.Genet.Evol.*, vol. 1, no. 4, pp. 297-301.

Kelman, Z. & O'Donnell, M. 1995, "DNA polymerase III holoenzyme: structure and function of a chromosomal replicating machine", *Annu. Rev. Biochem*, vol. 64, pp. 171-200.

Kilbey, B. J., Fraser, I., McAleese, S., Goman, M., & Ridley, R.G. 1993, "Molecular characterisation and stage-specific expression of proliferating cell nuclear antigen (PCNA) from the malarial parasite, *Plasmodium falciparum*", *Nucleic Acids Res.* vol. 21, no. 2, pp.239-243.

Killick-Kendrick, R. 1978, "Taxonomy, zoogeography and evolution," in *Rodent Malaria*, R. P. W. Killick-Kendrick, ed., Academic Press, London, pp. 1-53.

Kimura, S., Uchiyama, Y., Kasai, N., Saotome, A., Ueda, T., Ando, T., Ishibashi, T., Namekawa, S., Oshige, M., Furukawa, T., Yamamoto, T., Hashimoto, J., & Sakaguchi, K. 2002, "A novel DNA polymerase homologous to *Escherichia coli* DNA polymerase I from a higher plant, rice (*Oryza sativa L.*)", *Nucleic Acids Res.*, vol. 30, no. 7, pp. 1585-1592.

Klenow, H. & Overgaard-Hansen, K. 1970, "Proteolytic cleavage of DNA polymerase from *Escherichia coli* B into an exonuclease unit and a polymerase unit", *FEBS Lett.*, vol. 6, no. 1, pp. 25-27.

Kohler, S., Delwiche, C. F., Denny, P. W., Tilney, L. G., Webster, P., Wilson, R. J., Palmer, J. D., & Roos, D. S. 1997, "A plastid of probable green algal origin in Apicomplexan parasites", *Science*, vol. 275, no. 5305, pp. 1485-1489.

Kornberg, A. 1960, "Biological synthesis of dcoxyribonucleic acid", *Science*, vol. 131, pp. 1503-1508.

Kornberg, A., & Baker, T. 1992, "DNA replication", 2nd edn. New York. W.H. Freeman and Company.

Kornberg, A. 1984, "Enzyme studies of replication of the Escherichia coli chromosome", *Adv. Exp. Med. Biol.*, vol. 179, pp. 3-16.

Korolev, S., Nayal, M., Barnes, W. M., Di Cera, E., & Waksman, G. 1995, "Crystal structure of the large fragment of *Thermus aquaticus* DNA polymerase I at 2.5-A resolution: structural basis for thermostability", *Proc.Natl.Acad Sci.U.S.A*, vol. 92, no. 20, pp. 9264-9268.

Korolev, S., Hsieh, J., Gauss, G. H., Lohman, T. M., & Waksman, G. 1997, "Major domain swiveling revealed by the crystal structures of complexes of *Escheichia coli* Rep helicase bound to single-stranded DNA and ADP", *Cell*, vol. 90, no. 4, pp. 635-647.

Kowalski, D., & Eddy, M. J. 1989, "The DNA unwinding element: a novel, cis-acting component that facilitates opening of the *Escherichia coli* replication origin", *EMBO J.*, vol. 8, no. 13, pp. 4335-4344.

Krause, T., Luersen, K., Wrenger, C., Gilberger, T. W., Muller, S., & Walter, R. D. 2000, "The ornithine decarboxylase domain of the bifunctional ornithine decarboxylase/Sadenosylmethionine decarboxylase of *Plasmodium falciparum*: recombinant expression and catalytic properties of two different constructs", *Biochem.J.*, vol. 352 Pt 2, pp. 287-292. Kumar, S., Tamura, K., Jakobsen, I. B., & Nei, M. 2001, "MEGA2: molecular evolutionary genetics analysis software", *Bioinformatics*, vol. 17, no. 12, pp. 1244-1245.

Kunst, F., Ogasawara, N., Moszer, I., Albertini, A.M., Alloni, G., Azevedo, V., Bertero, M.G., Bessieres, P., Bolotin, A., Borchert, S., Borriss, R., Boursier, L., Brans, A., Braun, M., Brignell, S.C., Bron, S., Brouillet, S., Bruschi, C.V., Caldwell, B., Capuano, V., Carter, N.M., Choi, S.K., Codani, J.J., Connerton, I.F., Cummings, N.J., Daniel, R.A., Denizot, F., Devine, K.M., Dusterhoft, A., Ehrlich, S.D., Emmerson, P.T., Entian, K.D., Errington, J., Fabret, C., Ferrari, E., Foulger, D., Fritz, C., Fujita, M., Fujita, Y., Fuma, S., Galizzi, A., Galleron, N., Ghim, S.Y., Glaser, P., Goffeau, A., Golightly, E.J., Grandi, G., Guiseppi, G., Guy, B.J., Haga, K., Haicch, J., Harwood, C.R., Henaut, A., Hilbert, H., Holsappel, S., Hosono, S., Hullo, M.F., Itaya, M., Jones, L., Joris, B., Karamata, D., Kasahara, Y., Klaerr-Blanchard, M., Klein, C., Kobayashi, Y., Koetter, P., Koningstein, G., Krogh, S., Kumano, M., Kurita, K., Lapidus, A., Lardinois, S., Lauber, J., Lazarevic, V., Lee, S.M., Levine, A., Liu, H., Masuda, S., Mauel, C., Medigue, C., Medina, N., Mellado, R.P., Mizuno, M., Moestl, D., Nakai, S., Noback, M., Noone, D., O'Reilly, M., Ogawa, K., Ogiwara, A., Oudega, B., Park, S.H., Parro, V., Pohl, T.M., Portetelle, D., Porwollik, S., Prescott, A.M., Presecan, E., Pujic, P., Purnelle, B., Rapoport, G., Rey, M., Reynolds, S., Rieger, M., Rivolta, C., Rocha, E., Roche, B., Rose, M., Sadaic, Y., Sato, T., Scanlan, E., Schleich, S., Schroeter, R., Scoffone, F., Sekiguchi, J., Sekowska, A., Seror, S.J., Serror, P., Shin, B.S., Soldo, B., Sorokin, A., Tacconi, E., Takagi, T., Takahashi, H., Takemaru, K., Takeuchi, M., Tamakoshi, A., Tanaka, T., Terpstra, P., Tognoni, A., Tosato, V., Uchiyama, S., Vandenbol, M., Vannier, F., Vassarotti, A., Viari, A., Wambutt, R., Wedler, E., Wedler, H., Weitzenegger, T., Winters, P., Wipat, A., Yamamoto, H., Yamane, K., Yasunoto, K., Yata, K., Yoshida, K., Yoshikawa, H.F., Zumstein, E., Yoshikawa, H., & Danchin, A. 1997, "The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*", *Nature*, vol. 390, no. 6657, pp. 249-256.

Kusakabe, T. & Richardson, C. C. 1996, "The role of the zinc motif in sequence recognition by DNA primases", *J. Biol. Chem.*, vol. 271, no. 32, pp. 19563-19570.

 $\mathbf{\overline{L}}$

Laemmli, U. K. 1970, "Cleavage of structural proteins during the assembly of the head of bacteriophage T4", *Nature*, vol. 227, pp. 680-685.

Law, A. E., Mullineaux, C. W., Hirst, E. M., Saldanha, J., & Wilson, R. J. 2000, "Bacterial orthologues indicate the malarial plastid gene ycf24 is essential", *Protist.*, vol. 151, no. 4, pp. 317-327.

Lawyer, F.C., Stoffel, S., Saiki, R.K., Myambo, K., Drummond, R., & Gelfand, D.H. 1989, "Isolation, characterization, and expression in *Escherichia coli* of the DNA polymerase gene from *Thermus aquaticus*", *J.Biol.Chem.*, vol. 264, no. 11, pp. 6427-6437.

Leipe, D. D., Aravind, L., Grishin, N. V., & Koonin, E. V. 2000, "The bacterial replicative helicase DnaB evolved from a RecA duplication", *Genome Res.*, vol. 10, pp. 5-16.

Lell, B., & Kremsner, P. G. 2002, "Clindamycin as an antimalarial drug: review of clinical trials", *Antimicrob. Agents Chemother.*, vol. 46, no. 8, pp.2315-2320.

Lemon, K. P., & Grossman, A. D. 1998, "Localisation of bacterial DNA polymerase: evidence for a factory model of replication", *Science*, vol. 282, no.5393, pp. 1516-1519.

Le Roch, K.G., Zhou, Y., Blair, P. L., Grainger, M., Moch, J. K., Haynes, J. D., De La Vega, P., Holder, A. A., Batalov, S., Carucci, D. J., & Winzeler, E. A. 2003, "Discovery of gene function by expression profiling of the malaria parasite life cycle", *Science*, vol.301, no. 5639, pp. 1503-1508.

Lewis, D. L., Farr, C. L., Wang, Y., Lagina, A.T. 3rd., & Kaguni, L. S. 1996, "Catalytic subunit of mitochondrial DNA polymerase from *Drosophila* embryos. Cloning, bacterial overexpression, and biochemical characterization", *J.Biol.Chem.*, vol 271, no. 38 pp 23389-23394.

Lipps, G., Rother, S., Hart, C., & Krauss, G. 2003, "A novel type of replicative enzyme harbouring ATPase, primase and DNA polymerase activity", *EMBO J.*, vol. 22, no. 10, pp. 2516-2525.

M

Marians, K. J. 2000, "Crawling and wiggling on DNA: structural insights to the mechanism of DNA unwinding by helicases", *Structure*. *Fold.Des*, vol. 8, no. 12, p. R227-R235.

Marini, F. & Wood, R. D. 2002, "A human DNA helicase homologous to the DNA crosslink sensitivity protein Mus308", *J.Biol.Chem.*, vol. 277, no. 10, pp. 8716-8723.

Martin, W. & Herrmann, R. G. 1998, "Gene transfer from organelles to the nucleus: how much, what happens, and Why?", *Plant Physiol.*, vol. 118, no. 1, pp. 9-17.

Masutani, C., Araki, M., Yamada, A., Kusumoto, R., Nogimori, T., Maekawa, T., Iwai, S., & Hanaoka, F. 1999, "*Xeroderma pigmentosum* variant (XP-V) correcting protein from HeLa cells has a thymine dimer bypass DNA polymerase activity", *EMBO J.*, vol. 18, no. 12, pp. 3491-3501.

Matesanz, F., Duran-Chica, I., & Alcina, A. 1999, "The cloning and expression of Pfacs1, a *Plasmodium falciparum* fatty acyl coenzyme A synthetase-1 targeted to the host crythrocyte cytoplasm", *J.Mol.Biol*, vol. 291, no. 1, pp. 59-70.

Matson, S. W., & Kaiser-Rogers, K. A. 1990, "DNA helicases", Annu. Rev. Biochem., vol. 59, pp. 289-329.

Matson, S. W. 1991, "DNA helicases of *Escherichia coli*", *Prog.Nucleic Acid Res.Mol.Biol.*, vol. 40, pp. 289-326.

Matsui, M., Oka, A., Takanami, M., Yasuda, S., & Hirota, Y. 1985, "Sites of dnaA proteinbinding in the replication origin of the *Escherichia coli* K-12 chromosome", *J.Mol.Biol.*, vol. 184, no. 3, pp. 529-533.

McClelland, M., Sanderson, K.E., Spieth, J., Clifton, S.W., Latreille, P., Courtney, L., Porwollik, S., Ali, J., Dante, M., Du, F., Hou, S., Layman, D., Leonard, S., Nguyen, C., Scott, K., Holmes, A., Grewal, N., Mulvaney, E., Ryan, E., Sun, H., Florca, L., Miller, W., Stoneking, T., Nhan, M., Waterston, & R., Wilson, R.K. 2001, "Complete genome sequence of *Salmonella enterica serovar Typhimurium* LT2", *Nature*, vol. 413, no. 6858, pp. 852-856.

McConkey, G. A., Rogers, M. J., & McCutchan, T. F. 1997, "Inhibition of *Plasmodium falciparum* protein synthesis. Targeting the plastid-like organelle with thiostrepton", *J.Biol. Chem.*, vol. 272, no. 4, pp. 2046-2049.

McFadden, G. I. & Waller, R. F. 1997, "Plastids in parasites of humans", *Bioessays*, vol. 19, no. 11, pp. 1033-1040.

McFadden, G. I. & Roos, D. S. 1999, "Apicomplexan plastids as drug targets", *Trends Microbiol.*, vol. 7, no. 8, pp. 328-333.

Mendelman, L. V., Beauchamp, B. B., & Richardson, C. C. 1994, "Requirement for a zinc motif for template recognition by the bacteriophage T7 primase", *EMBO J.*, vol. 13, no. 16, pp. 3909-3916.

Meshnik, E. 2002, "Artemisinin: mechanisms of action, resistance and toxicity", *Int.J. Parasitol.*, vol. 32, no. 13, pp. 1655-1660.

Morgenstern, B., Dress, A., & Werner, T. 1996, "Multiple DNA and protein sequence alignment based on segment-to-segment comparison", *Proc.Natl.Acad Sci.U.S.A*, vol. 93, no. 22, pp. 12098-12103.

Moriya, S., Ogasawara, N., & Yoshikawa, H. 1985, "Structure and function of the region of the replication origin of the *Bacillus subtilis* chromosome. III. Nucleotide sequence of some 10,000 base pairs in the origin region", *Nucleic Acids Res.*, vol. 13, no. 7, pp. 2251-2265.

Morrison, A., Bell, J. B., Kunkel, T. A., & Sugino, A. 1991, "Eukaryotic DNA polymerase amino acid sequence required for 3'----5' exonuclease activity", *Proc.Natl.Acad Sci.U.S.A*, vol. 88, no. 21, pp. 9473-9477.

Müller, S., Da'dara, A., Luersen, K., Wrenger, C., Das, G. R., Madhubala, R., & Walter, R. D. 2000, "In the human malaria parasite *Plasmodium fulciparum*, polyamines are synthesized by a bifunctional ornithine decarboxylase, S- adenosylmethionine decarboxylase", *J.Biol. Chem.*, vol. 275, no. 11, pp. 8097-8102.
Nagasawa, K., Kitamura, K., Yasui, A., Nimura, Y., Ikeda, K., Hirai, M., Matsukage, A., & Nakanishi, M. 2000, "Identification and characterization of human DNA polymerase beta 2, a DNA polymerase beta -related enzyme", *J.Biol.Chem.*, vol. 275, no. 40, pp. 31233-31238.

Nakayama, N., Arai, N., Bond, M.W., Kaziro, Y., & Arai, K. 1984, "Nucleotide sequence of dnaB and the primary structure of the dnaB protein from *Escherichia coli*", *J.Biol. Chem.*, vol. 259, no. 1, pp. 97-101.

Nelson, J. R., Lawrence, C. W., & Hinkle, D. C. 1996, "Thymine-thymine dimer bypass by yeast DNA polymerase zeta", *Science*, vol. 272, no. 5268, pp. 1646-1649.

Nelson, K.E., Clayton, R.A., Gill, S.R., Gwinn, M.L., Dodson, R.J., Haft, D.H., Hickey, E.K., Peterson, J.D., Nelson, W.C., Ketchum, K.A., McDonald, L., Utterback, T.R., Malek, J.A., Linher, K.D., Garrett, M.M., Stewart, A.M., Cotton, M.D., Pratt, M.S., Phillips, C.A., Richardson, D., Heidelberg, J., Sutton, G.G., Fleischmann, R.D., Eisen, J.A., & Fraser, C.M. 1999, "Evidence for lateral gene transfer between Archaea and bacteria from genome sequence of *Thermotoga maritima*", *Nature* vol. 399, no. 6734, pp. 323-329.

Nelson, K., Paulsen, I., Weinel, C., Dodson, R., Hilbert, H., Fouts, D.,Gill, S., Pop, M., Martins Dos Santos, V., Holmes, M., Brinkac, L., Beanan, M., DeBoy, R., Daugherty, S., Kolonay, J., Madupu, R., Nelson, W., White, O., Peterson, J., Khouri, H., Hance, I., Lee, P., Holtzapple, E., Scanlan, D., Tran, K., Moazzez, A., Utterback, T., Rizzo, M., Lee, K., Kosack, D., Moestl, D., Wedler, H., Lauber, J., Hoheisel, J., Straetz, M., Heim, S., Kiewitz, C., Eisen, J., Timmis, K., Duesterhoft, A., Tummler, B., & Fraser, C. 2002, "Complete genome sequence and comparative analysis of the metabolically versatile *Pseudomonus putida* KT2440", *Environ. Microbiol.*, vol. 4, no. 12, pp. 799-808.

Nielsen, II., Engelbrecht, J., Brunak, S., & von Heijne, G. 1997, "Identification of prokaryotic and cukaryotic signal peptides and prediction of their cleavage sites", *Protein Eng.*, vol. 10, no. 1, pp. 1-6.

$\underline{\mathbf{0}}$

Ohashi, E., Bebenek, K., Matsuda, T., Feaver, W. J., Gerlach, V. L., Friedberg, E. C., Ohmori, H., & Kunkel, T. A. 2000, "Fidelity and processivity of DNA synthesis by DNA polymerase kappa, the product of the human DINB1 gene", *J.Biol.Chem.*, vol. 275, no. 50, pp. 39678-39684.

Ohmori, H., Friedberg, E. C., Fuchs, R. P., Goodman, M. F., Hanaoka, F., Hinkle, D., Kunkel, T. A., Lawrence, C. W., Livneh, Z., Nohmi, T., Prakash, L., Prakash, S., Todo, T., Walker, G. C., Wang, Z., & Woodgate, R. 2001, "The Y-family of DNA polymerases", *Mol.Cell*, vol. 8, no. 1, pp. 7-8.

Ollis, D. L., Brick, P., Hamlin, R., Xuong, N. G., & Steitz, T. A. 1985, "Structure of large fragment of *Escherichia coli* DNA polymerase I complexed with dTMP", *Nature*, vol. 313, no. 6005, pp. 762-766.

Olsen, G. J. 1988, "Phylogenetic analysis using ribosomal RNA", *Methods Enzymol.*, vol. 164, pp. 793-812.

P

Pajunen, M.I., Kiljunen, S.J., Soderholm, M.E., & Skurnik, M. 2001, "Complete genomic sequence of the lytic bacteriophage phiYeO3-12 of *Yersinia enterocolitica* serotype O:3", *J. Bacteriol.*, vol. 183, no. 6, pp. 1928-1937.

Pajunen, M.I., Elizondo, M.R., Skurnik, M., Kieleczawa, J., & Molineux, I.J. 2002, "Complete nucleotide sequence and likely recombinatorial origin of bacteriophage T3", *J. Mol.Biol.*, vol. 319, no. 5, pp. 1115-1132.

Pandey, V. N., Williams, K. R., Stone, K. L., & Modak, M. J. 1987, "Photoaffinity labeling of the thymidine triphosphate binding domain in *Escherichia coli* DNA polymerase I: identification of histidine-881 as the site of cross-linking", *Biochemistry*, vol. 26, no. 24, pp. 7744-7748.

Pandey, V. N. & Modak, M. J. 1988, "Affinity labeling of *Escherichia coli* DNA polymerase I by 5'-fluorosulfonylbenzoyladenosine. Identification of the domain essential for polymerization and Arg-682 as the site of reactivity", *J.Biol.Chem.*, vol. 263, no. 13, pp. 6068-6073.

Pandey, V. N., Kaushik, N. A., Pradhan, D. S., & Modak, M. J. 1990, "Template primerdependent binding of 5'-fluorosulfonyl-benzoyldeoxyadenosine by *Escherichia coli* DNA polymerase I. Identification of arginine 682 as the binding site and its implication in catalysis", *J.Biol.Chem.*, vol. 265, no. 7, pp. 3679-3684.

Pardoll, D.M., Vogelstein, B., & Coffey, D.S. 1980, "A fixed site of DNA replication in eukaryotic cells", *Cell*, vol 19, no.2, pp. 527-536.

Parkhill, J., Dougan, G., James, K.D., Thomson, N.R., Pickard, D., Wain, J., Churcher, C., Mungali, K.L., Bentley, S.D., Holden, M.T.G., Sebaihia, M., Baker, S., Basham, D., Brooks, K., Chillingworth, T., Connerton, P., Cronin, A., Davis, P., Davies, R.M., Dowd, L., White, N., Farrar, J., Feltwell, T., Hamlin, N., Haque, A., Hien, T.T., Holroyd, S., Jagels, K., Krogh, A., Larsen, T.S., Leather, S., Moule, S., O'Gaora, P., Parry, C., Quail, M., Rutherford, K., Simmonds, M., Skelton, J., Stevens, K., Whitehead, S., & Barrell, B.G. 2001, "Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar Typhi CT18", *Nature*, vol. 413, no. 6858, pp. 848-852.

Peitsch, M.C. 1996, "ProMod and Swiss-Model: Internet-based tools for automated comparative protein modelling", *Biochem.Soc.Trans.*, vol. 24, no. 1, pp. 274-279.

Podobnik, M., McInerney, P., O'Donnell, M., & Kuriyan, J. 2000, "A TOPRIM domain in the crystal structure of the catalytic core of *Escherichia coli* primase confirms a structural link to DNA topoisomerases", *J. Mol. Biol.*, vol. 300, no. 2, pp. 353-362.

Polesky, A. H., Steitz, T. A., Grindley, N. D., & Joyce, C. M. 1990, "Identification of residues critical for the polymerase activity of the Klenow fragment of DNA polymerase I from *Escherichia coli*", *J.Biol.Chem.*, vol. 265, no. 24, pp. 14579-14591.

Polesky, A. H., Dahlberg, M. E., Benkovic, S. J., Grindley, N. D., & Joyce, C. M. 1992, "Side chains involved in catalysis of the polymerase reaction of DNA polymerase I from *Escherichia coli*", *J.Biol.Chem.*, vol. 267, no. 12, pp. 8417-8428.

Pons, M. E., Diaz, A., Lacks, S. A., & Lopez, P. 1991, "The polymerase domain of *Streptococcus pneumoniae* DNA polymerase I. High expression, purification and characterization", *Eur.J.Biochem.*, vol. 201, no. 1, pp. 147-155.

Preiser, P. R., Wilson, R. J., Moore, P. W., McCready, S., Hajibagheri, M. A., Blight, K. J., Strath, M., & Williamson, D. H. 1996, "Recombination associated with replication of malarial mitochondrial DNA", *EMBO J.*, vol. 15, no. 3, pp. 684-693.

R

Ridley, R. G., White, J. H., McAleese, S. M., Goman, M., Alano, P., de Vries, E., & Kilbey, B. J. 1991, "DNA polymerase delta: gene sequences from *Plasmodium falciparum* indicate that this enzyme is more highly conserved than DNA polymerase alpha", *Nucleic Acids Res.*, vol. 19, no. 24, pp. 6731-6736.

Rimchala, P., Karbwang, J., Sukontason, K., Banmairuroi, V., Molunto, P and Na-Bangchang, K.1996. "Pharmacokinetics of quinine in patients with chronic renal failure", *Eur. J. Clin. Pharmacol.*, vol. 49, pp. 497-501. Roos, D.S., Crawford, M. J., Donald, R, G, K., Kissinger, J.C., Klimczak, L, J., & Striepen, B. 1999, "Origin, targeting and function of the apicomplexan plastid", *Curr.Opin.Microbiol.*, vol. 2, pp. 426-432.

Rosenthal, P. J. 1996, "Conservation of key amino acids among the cysteine proteinases of multiple malarial species", *Mol.Biochem.Parasitol.*, vol. 75, no. 2, pp. 255-260.

S

Salas, M. 1991, "Protein-priming of DNA replication", Annu. Rev. Biochem., vol. 60, pp. 39-71.

Salcedo, E., Cortese, J.F., Plowe, C.V., Sims, P.F.G., & Hyde, J.E. 2001, "A bifunctional dihydrofolate synthetase--folylpolyglutamate synthetase in *Plasmodium falciparum* identified by functional complementation in yeast and bacteria", *Mol.Biochem.Parasitol.*, vol. 112, no. 2, pp. 239-52.

Sambrook, J., Fristsch, E. F., & Maniatis, T. 1989, *Molecular cloning: A laboratory manual*, 2nd edn, Cold Spring Habour NY.

Sancar, A. 1996, "DNA excision repair", Annu. Rev. Biochem., vol. 65, pp. 43-81.

Sato, S., Tews, I., & Wilson, R. J. 2000, "Impact of a plastid-bearing endocytobiont on apicomplexan genomes", *Int.J.Parasitol.*, vol. 30, no. 4, pp. 427-439.

Sato, S. & Wilson, R. J. 2002, "The genome of *Plasmodium falciparum* encodes an active delta-aminolevulinic acid dehydratase", *Curr.Genet.*, vol. 40, no. 6, pp. 391-398.

Sato, S., Rangachari, K., & Wilson, R. J. 2003, "Targeting GFP to the malarial mitochondrion", *Mol. Biochem. Parasitol.*, vol. 130, pp. 155-158.

Saul, A., & Battistutta, D. 1988, "Codon usage in *Plasmodium falciparum*", *Mol.Biochem. Parasitol.*, vol. 27, no. 1, pp. 35-42.

Sawaya, M. R., Guo, S., Tabor, S., Richardson, C. C., & Ellenberger, T. 1999, "Crystal structure of the helicase domain from the replicative helicase-primase of bacteriophage T7", *Cell*, vol. 99, no. 2, pp. 167-177.

Schwede, T., Kopp, J., Guex, N., & Peitsch, M. C. 2003, "SWISS-MODEL: An automated protein homology-modeling server", *Nucleic Acids Res.*, vol. 31, no. 13, pp. 3381-3385.

Seto, H., Hatanaka, M., Kimura, S., Oshige, M., Tsuya, Y., Mizushina, Y., Sawado, T., Aoyagi, N., Matsumoto, T., Hashimoto, J., & Sakaguchi, K. 1998, "Purification and characterization of a 100 kDa DNA polymerase from cauliflower inflorescence", *Biochem. J.*, vol. 332 (Pt 2), pp. 557-563.

ŝ

Sharief, F. S., Vojta, P. J., Ropp, P. A., & Copeland, W. C. 1999, "Cloning and chromosomal mapping of the human DNA polymerase theta (POLQ), the eighth human DNA polymerase", *Genomics*, vol. 59, no. 1, pp. 90-96.

Sibley, C. H., Hyde, J. E., Sims, P. F., Plowe, C.V., Kublin, J. G., Mberu, E. K., Cowman, A. F., Winstanley, P. A., Watkins, W. M., & Nzila, A. M. 2001, "Pyrimethamine-sulfadoxine resistance in Plasmodium falciparum: what next? ", *Trends Parasitol.*, vol. 17, no. 12, pp. 582-588.

Singh, D., Chaubey, S., & Habib, S. 2003, "Replication of the Plasmodium falciparum apicoplast DNA initiates within the inverted repeat region", *Mol.Biochem.Parasitol.*, vol. 126, no. 1, pp. 9-14.

Smiley, B.L., Lupski, J.R., Svec, P.S., McMacken, R., & Godson, G.N. 1982, "Sequences of the *Escherichia coli* dnaG primase gene and regulation of its expression", *Proc. Natl. Acad. Sci. U.S.A.*, vol. 79, no. 15, pp. 4550-4554.

Soil, J., & Tien, R. 1998, "Protein translocation into and across the chloroplastic envelope membranes", *Plant Mol.Biol.*, vol. 38, no. 1-2, pp. 191-207.

Soultanas, P., Dillingham, M. S., Velankar, S. S., & Wigley, D. B. 1999, "DNA binding mediates conformational changes and metal ion coordination in the active site of PcrA helicase", *J.Mol.Biol.*, vol. 290, no. 1, pp. 137-148.

Sousa, R. 1996, "Structural and mechanistic relationships between nucleic acid polymerases", *Trends Biochem.Sci.*, vol. 20, no. 5, pp. 186-190.

Spadari, S., Sala, F., & Pedrali-Noy, G. 1984, "Aphidicolin and eukaryotic DNA synthesis", *Adv.Exp.Med.Biol.*, vol. 179, pp. 169-181.

Spadari, S., Pedrali-Noy, G., Falaschi, M. C., & Ciarrocchi, G. 1984, "Control of DNA replication and cell proliferation in cukaryotes by aphidicolin", *Toxicol.Pathol.*, vol. 12, no. 2, pp. 143-148.

Spelbrink, J.N., Li, F.Y., Tiranti, V., Nikali, K., Yuan, Q.P., Tariq, M., Wanrooij, S., Garrido, N., Comi, G., Morandi, L., Santoro, L., Toscano, A., Fabrizi, G.M., Somer, H., Croxen, R., Beeson, D., Poulton, J., Suomalainen, A., Jacobs, H.T., Zeviani, M., & Larsson, C. 2001, "Human mitochondrial DNA deletions associated with mutations in the gene encoding Twinkle, a phage T7 gene 4-like protein localized in mitochondria", *Nat. Genet.*, vol. 28, no. 3, pp. 223-231.

Steitz, T. A. 1998, "A mechanism for all polymerases", Nature, vol. 391, no. 6664, pp. 231-232.

Steitz, T. A. 1999, "DNA polymerases: structural diversity and common mechanisms", *J.Biol.Chem.*, vol. 274, no. 25, pp. 17395-17398.

Strack, B., Lessi, M., Calendar, R., & Lanka, E. 1992, "A common sequence motif, -E-G-Y-A-T-A-, identified within the primase domains of plasmid-encoded I- and P-type DNA primases and the alpha protein of the *Escherichia coli* satellite phage P4", *J.Biol.Chem.*, vol. 267, no. 18, pp. 13062-13072.

Surolia, N. & Padmanaban, G. 1991, "Chloroquine inhibits heme-dependent protein synthesis in *Plasmodium falciparum*", *Proc.Natl.Acad Sci.U.S.A*, vol. 88, no. 11, pp. 4786-4790.

Surolia, N. & Padmanaban, G. 1992, "de novo biosynthesis of home offers a new chemotherapeutic target in the human malarial parasite", *Biochem.Biophys.Res.Commun.*, vol. 187, no. 2, pp. 744-750.

Surolia, N. & Surolia, A. 2001, "Triclosan offers protection against blood stages of malaria by inhibiting enoyl-ACP reductase of *Plasmodium falciparum*", *Nat.Med.*, vol. 7, no. 2, pp. 167-173.

Szafranski, P., Smith, C. L., & Cantor, C. R. 1997, "Cloning and analysis of the dnaG gene encoding *Pseudomonas putida* DNA primase", *Biochim.Biophys.Acta*, vol. 1352, no. 3, pp. 243-248.

T

Taylor, A. F. & Smith, G. R. 1985, "Substrate specificity of the DNA unwinding activity of the RecBC enzyme of *Escherichia coli*", *J.Mol.Biol.*, vol. 185, no. 2, pp. 431-443.

Thommes, P. & Hubscher, U. 1992, "Eukaryotic DNA helicases: essential enzymes for DNA transactions", *Chromosoma*, vol. 101, no. 8, pp. 467-473.

Thompson, J. D., Higgins, D. G., & Gibson, T. J. 1994, "CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice", *Nucleic Acids Res.*, vol. 22, no. 22, pp. 4673-4680.

Tissier, A., McDonald, J. P., Frank, E. G., & Woodgate, R. 2000, "Polymerase iota, a remarkably error-prone human DNA polymerase", *Genes Dev.*, vol. 14, no. 13, pp. 1642-1650.

Toh, H. 1986, "T7 and *Escherichia coli* share homology for replication-related gene products", *FEBS Lett.*, vol. 194, no. 2, pp. 245-248.

Tougu, K., Peng, H., & Marians, K. J. 1994, "Identification of a domain of *Escherichia coli* primase required for functional interaction with the DnaB helicase at the replication fork", *J.Biol.Chem.*, vol. 269, no. 6, pp. 4675-4682.

Triglia, T., & Cowman, A.F. 1994, "Primary structure and expression of the dihydropteroate synthetase gene of *Plasmodium falciparum*", *Proc.Natl.Acad/.Sci.U.S.A.*, vol. 91, no. 15, pp. 7149-7153.

Tseng, T. Y., Frick, D. N., & Richardson, C. C. 2000, "Characterization of a novel DNA primase from the *Salmonella typhimurium* bacteriophage SP6", *Biochemistry*, vol. 39, no. 7, pp. 1643-1654.

U

Umek, R. M., & Kowalski, D. 1990, "The DNA unwinding element in a yeast replication origin functions independently of easily unwound sequences present elsewhere on a plasmid", *Nucleic Acids Res.*, vol. 18, no. 22, pp. 6601-6605.

V

Vaidya, A. B., & Arasu, P. 1987, "Tandemly arranged gene clusters of malarial parasites that are highly conserved and transcribed", *Mol.Biochem.Parasitol.*, vol. 22, no. 2-3, pp. 249-257.

Vaidya, A. B., Akella, R., & Suplick, K. 1989, "Sequences similar to genes for two mitochondrial proteins and portions of ribosomal RNA in tandemly arrayed 6-kilobase-pair DNA of a malarial parasite", *Mol.Biochem.Parasitol.*, vol. 35, no. 2, pp. 97-107.

van Dooren, G. G., Waller, R. F., Joiner, K. A., Roos, D. S., & McFadden, G. I. 2000, "Traffic jams: protein transport in *Plasmodium falciparum*", *Parasitol.Today*, vol. 16, no. 10, pp. 421-427.

van Dooren, G. G., Su, V., D'Ombrain, M. C., & McFadden, G. I. 2002, "Processing of an apicoplast leader sequence in *Plasmodium falciparum* and the identification of a putative leader cleavage enzyme", *J.Biol.Chem.*, vol. 277, no. 26, pp. 23612-23619.

Velankar, S. S., Soultanas, P., Dillingham, M. S., Subramanya, H. S., & Wigley, D. B. 1999, "Crystal structures of complexes of PerA DNA helicase with a DNA substrate indicate an inchworm mechanism", *Cell*, vol. 97, no. 1, pp. 75-84.

Versalovic, J., & Lupski, J. R. 1993, "The *Haemophilus influenzae* dnaG sequence and conserved bacterial primase motifs", *Gene*, vol. 136, no. 1-2, pp. 281-286.

W

Walker, J. E., Saraste, M., Runswick, M. J., & Gay, N. J. 1982, "Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold", *EMBO J.*, vol. 1, no. 8, pp. 945-951.

Waller, R. F., Keeling, P. J., Donald, R. G., Striepen, B., Handman, E., Lang-Unnasch, N., Cowman, A. F., Besra, G. S., Roos, D. S., & McFadden, G. I. 1998, "Nuclear-encoded proteins target to the plastid in *Toxoplasma gondii* and *Plasmodium falciparum*", *Proc.Natl.Acad Sci.U.S.A*, vol. 95, no. 21, pp. 12352-12357.

Waller, R. F., Reed, M. B., Cowman, A. F., & McFadden, G. I. 2000, "Protein trafficking to the plastid of *Plasmodium falciparum* is via the secretory pathway", *EMBO J.*, vol. 19, no. 8, pp. 1794-1802.

Warhust, D. C. 2001, "New developments: chloroquine-resistance in *Plasmodium falciparum*", *Drug Resist Updat.*, vol. 4, no. 3, pp. 141-144.

Warhust, D. C. 2002, "Resistance to antifolates in Plasmodium falciparum, the causative agent of tropical malaria", *Sci. Prog.*, vol. 85, no. 1, pp. 89-111.

Warrell, D. A. 1992, "Pathophysiology of severe falciparum malaria", P.N.G.Med.J., vol. 35, no. 4, pp. 229-232.

Washington, M. T., Rosenberg, A. H., Griffin, K., Studier, F. W., & Patel, S. S. 1996, "Biochemical analysis of mutant T7 primase/helicase proteins defective in DNA binding, nucleotide hydrolysis, and the coupling of hydrolysis with DNA unwinding", *J. Biol. Chem.*, vol. 271, no. 43, pp. 26825-26834.

Weber, J. L. 1987, "Analysis of sequences from the extremely A + T-rich genome of *Plasmodium falciparum*", *Gene*, vol. 52, no. 1, pp. 103-109.

Weissig, V., Vetro-Widenhouse, T. S., & Rowe, T. C. 1997, "Topoisomerase II inhibitors induce cleavage of nuclear and 35-kb plastid DNAs in the malarial parasite *Plasmodium falciparum*", *DNA Cell Biol.*, vol. 16, no. 12, pp. 1483-1492.

West, S.C. 1996, "DNA helicases: new breeds of translocating motors and molecular pumps", *Cell*, vol. 96, pp. 177-180.

White, N, J. 1997, "Assessment of the pharmacodynamic properties of antimalarial drugs *in vivo*", *Antimicrob.Agents Chemother.* vol. 41, no. 7, pp. 1413-1422.

White, J. H., Kilbey, B. J., de Vries, E., Goman, M., Alano, P., Cheesman, S., McAleese, S., & Ridley, R.G. 1993, "The gene encoding DNA polymerase alpha from *Plasmodium falciparum*", *Nucleic Acids Res.*, vol. 21, no. 16, pp. 3643-3646.

White, J. H. & Kilbey, B. J. 1996, "DNA replication in the malaria parasite", *Parasitol.Today*, vol. 12, pp. 151-155.

WHO 1998, World Health Organization. Factsheet No. 94 (World Health Organization, Geneva, 1998).

Wiesner, J., Hintz, M., Altincicek, B., Sanderbrand, S., Weidemeyer, C., Beck, E., & Jomaa, H. 2000, "*Plasmodium falciparum*: detection of the deoxyxylulose 5-phosphate reductoisomerase activity", *Exp. Parasitol.*, vol. 96, no. 3, pp. 182-186.

Williamson, D. H., Gardner, M. J., Preiser, P., Moore, D. J., Rangachari, K., & Wilson, R. J. 1994, "The evolutionary origin of the 35 kb circular DNA of *Plasmodium falciparum*: new evidence supports a possible rhodophyte ancestry", *Mol.Gen.Genet.*, vol. 243, no. 2, pp. 249-252.

Williamson, D. H., Preiser, P. R., & Wilson, R. 1996, "Organelle DNAs: the bit players in malaria parasite DNA replication", *Parasitol.Today*, vol. 12, pp. 357-362.

Williamson, D. H., Preiser, P. R., Moore, P. W., McCready, S., Strath, M., & Wilson, R. J. 2002, "The plastid DNA of the malaria parasite *Plasmodium falciparum* is replicated by two mechanisms", *Mol.Microbiol.*, vol. 45, no. 2, pp. 533-542.

Wilson, I. 1993, "Plastids better red than dead", Nature, vol. 366, no. 6456, p. 638.

Wilson, R. J., Denny, P. W., Preiser, P. R., Rangachari, K., Roberts, K., Roy, A., Whyte, A., Strath, M., Moore, D. J., Moore, P. W., & Williamson, D. H. 1996, "Complete gene map of the plastid-like DNA of the malaria parasite *Plasmodium falciparum*", *J.Mol.Biol.*, vol. 261, no. 2, pp. 155-172.

Wilson, R. J. & Williamson, D. H. 1997, "Extrachromosomal DNA in the Apicomplexa", *Microbiol.Mol.Biol.Rev.*, vol. 61, no. 1, pp. 1-16.

Wilson, R. J. 2002, "Progress with parasite plastids", J.Mol.Biol., vol. 319, no. 2, pp. 257-274.

Wittmann, II. G. 1983, "Architecture of prokaryotic ribosomes", Annu. Rev. Biochem., vol. 52, pp. 35-65.

$\underline{\mathbf{Z}}$

Zeidler, J., Schwender, J., Mueller, C., & Lichtenthaler, H. K. 2000, "The non-mevalonate isoprenoid biosynthesis of plants as a test system for drugs against malaria and pathogenic bacteria", *Biochem.Soc.Trans.*, vol. 28, no. 6, pp. 796-798.

Zhang, Z., Green, B. R., & Cavalier-Smith, T. 2000, "Phylogeny of ultra-rapidly evolving dinoflagellate chloroplast genes: a possible common origin for sporozoan and dinoflagellate plastids", *J. Mol. Evol.*, vol. 51, no. 1, pp. 26-40.

Zhang, Z., Green, B. R., & Cavalier-Smith, T. 2000, "Phylogeny of ultra-rapidly evolving dinoflagellate chloroplast genes: a possible common origin for sporozoan and dinoflagellate plastids", *J.Mol.Evol.*, vol. 51, no. 1, pp. 26-40.

Ziegelin, G., Linderoth, N. A., Calendar, R., & Lanka, E. 1995, "Domain structure of phage P4 alpha protein deduced by mutational analysis", *J.Bacteriol.*, vol. 177, no. 15, pp. 4333-4341.

Zuegge, J., Ralph, S., Schmuker, M., McFadden, G. I., & Schneider, G. 2001, "Deciphering apicoplast targeting signals--feature extraction from nuclear-encoded precursors of *Plasmodium falciparum* apicoplast proteins", *Gene*, vol. 280, no. 1-2, pp. 19-26.

Zyskind, J. W., Cleary, J. M., Brusilow, W. S., Harding, N. E., & Smith, D. W. 1983, "Chromosomal replication origin from the marine bacterium *Vibrio harveyi* functions in *Escherichia coli*: oriC consensus sequence", *Proc.Natl.Acad Sci.U.S.A*, vol. 80, no. 5, pp. 1164-1168. Appendices

٨.

÷

ł

APPENDIX A

1. Media

Luria-Bertani Broth

10 g NaCl

10 g Tryptone

5 g Yeast Extract

Make up to 1 litre with distilled water with pH adjusted to 7.0 with 5M NaOH before autoclaving.

SOB

20 g Bacto-Tryptone 24 g Yeast Extract 0.5 g NaCl 2.5 ml 1M KCl 900 ml H₂O

Adjust pH to 7.0 with 10M NaOH and add H₂O to 990 ml. Sterilise by autoclaving and store at room temperature. Before use, add 10 ml sterile 1M MgCl₂.

SOC

Identical to SOB medium, except it contains 20 ml sterile 1M glucose.

2. Buffers

PBS 10X 80 g NaCl 2g KCl 26.8 g Na₂HPO₄-7 H₂O 2.4 g KH₂PO₄ 800 ml H₂O Adjust pH to 7.4 with HCi and sterilise by autoclaving.

ł

Denhardt's Solution

250X
50 g Ficoll 400
50 g polyvinylpyroolidone
50 g BSA
600 ml H₂O
Adjust volume to 1 litre with H₂O. Divide in aliquots and store at -20°C.

DTT

1M
3.085 g DTT
20 ml 10 mM Sodium acctate (pH 5.2)
Divide into 1 ml aliquots and sterilise by filtration. Store at -20°C.

0.5M EDTA (pH 8.0)

186.12 g Na₂EDTA-2 II₂O

Adjust pH to 8.0 with NaOH and adjust volume to 1 litre with H_2O . Sterilise by autoclaving.

MgCl₂

1M 203.31g MgCl₂ 800 ml H₂O Adjust volume to 1 litre with H₂O. Sterilise by autoclaving.

MOPS

10X
41.85 g MOPS
6.8 g sodium acctate
800 ml DEPC-treated H₂O
Add 20 ml of a DEPC-treated 0.5M l

Add 20 ml of a DEPC-treated 0.5M EDTA and adjust pH to 7.0 with 10M NaOH. Adjust volume to 1 litre with DEPC-treated H₂O. Protect from light and store at 4°C. Sterilise by autoclaving.

ģ

NaOAc

3M 408.24 g sodium acctate 800 ml H₂O

Adjust to pH 5.2 with glacial acetic acid and adjust volume to 1 litre with H_2O . Sterilise by autoclaving.

a the standing of the

いたちょう たいほうゆい 日本体験 スワイト

Constraints and a first of the second

and the second of the second s

ŝ

a diservities of the

4

A strategy of the residence of a strategy of the strategy of the

. .

NaCl

5M 292.2 g sodium chloride 800 ml H₂O Adjust volume to 1 litre and sterilise by autoclaving.

SDS

10%

100 g SDS

900 ml H₂O

Heat to 68°C to dissolve crystals and adjust pH to 7.2 with HCl. Adjust volume to 1 litre.

TCA

100% 500 g trichloroacetic acid 227 ml H₂O

TRIS

1M

121.14 g TRIS 800 ml H_2O Adjust pH to desired value by adding concentrated HCl. Adjust volume to 1 litre with H_2O and sterilise by autoclaving.

SSC 20X 175.3 g NaCl 88.2 g sodium citrate 800 ml H₂O Adjust pH to 7.0 with HCl and make up to 1 litre with H₂O. Sterilise by autoclaving.

17

i Q

A SAME AND A SAME AND A SAME

このであるというではないので、

TE

1X10 ml 1M TRIS2ml 0.5M EDTA (pH 8.0)Mix and adjust volume to 1 litrc with H₂O. Sterilise by autoclaving.

3. Buffers for electrophoresis of DNA

6X Loading buffer

0.25 % bromophenol blue 40 % (w/v) sucrose

50 X TAE buffer

242 g TRIS57.1 ml Glacial acetic acidMake up to 1 litre with H₂O. Dilute to 1X prior to use.

4. Buffers for electrophoresis of RNA

20X Northern gel buffer

96.5 g Na₂HPO₄
5.5 g NaH₂PO₄
Make up to 1 litre with DEPC-treated H₂O.

1X Northern gel buffer

50 ml 20X northern gel buffer 88 ml formaldehyde 860 ml DEPC-treated H_2O Make up to 1 litre with DEPC-treated H_2O .

RNA sample buffer

50 % glycerol 1 mM EDTA 0.4 % bromophenol blue

DEPC-treatment (per 100 ml)

0.01% DEPC or 0.1 ml DEPC to 1000 ml of a solution (e.g. H₂O) Shake and incubate overnight in a fume hood. Autoclave the solution to inactivate the remaining DEPC. å

£

5. Buffers for the electrophoresis of proteins

10% Ammonium persulphate (APS)

1g APS
8 ml H₂O
Adjust volume to 10 ml with H₂O. Store at 4°C.

10X Running Buffer

10 g SDS 30.3 g TRIS 144.1 g Glycine 800 ml II₂O Make up to 1 litre and dilute to 1x prior to use.

2X SDS PAGE sample buffer

10 ml 1.5M Tris (pH 6.8)
6 ml 20% SDS
30 ml Glycerol
15 ml β-mercaptoethanol
1.8 mg Bromophenol blue

Adjust volume to 100 ml with H₂O. Aliquot in 10 ml stock solution.

10% Resolving gcl (10 ml volume)

4 mi H₂O 3.3 ml 30% acyl-bisacrylamide mix 2.5 ml 1.5M TRIS (pH 8.8) 0.1 ml 10% SDS 0.1 ml 10% APS 10 μl TEMED

5% Stacking gel (5 ml volume)

3.4 ml H₂O
0.83 ml 30 % acryl-bisacrylamide mix
0.63 ml 0.5 M TRIS (pH 6.8)
50 μl 10% SDS
50 μl 10% APS
5 μl TEMED

Coomassie Blue staining solution

2.5 g Coomassie Brilliant Blue R-250
450 ml Methanol
100 ml Acetic acid
400 ml H₂O
Adjust volume to 1 litre with H₂O.

Coomassie Blue destaining solution

450 ml Methanol
100 ml Acetic acid
400 ml H₂O
Adjust volume to 1 litre with II₂O.

6. Buffers for Western blotting

10X Transfer buffer

24.22 g TRIS
112.5 g Glycine
Make up to 1 litre with H₂O.
Use at 1 X - 100 ml 10X stock, 200 ml methanol and 700 ml H₂O. Prechill at 4°C.
10X TBS
24.22g TRIS
80 g NaCl
Adjust pH to 7.6 and make up to 1 litre with H₂O. Store at 4°C.

- 3

Ponceau S solution

0.5 % Ponceau S 1% Acetic acid

7. Buffers for indirect immunofluorescence

Blocking buffer

0.01% Tween 20 0.05% Sodium azide 1% BSA Make up in PBS.

Immunobuffer

0.1 M TRIS, pH 7.4
2.3 M NaCl
0.01% Tween 20
1% BSA
0.05% Sodium azide

8. Buffers for In situ hybridisation

Hybridisation buffer

37.5 % formamide2X SSC1 mM TRIS, pH 7.2

0.1 mM EDTA

10% Dextran sulphate

1 μg Herring sperm DNA

Detection buffer

4X SSC 0.2% Tween 20 The second in second second

いたい ビード・ビーン とうきょうかい しんたいがく

APPENDIX B

Le Roch et al (2003) transcriptome analysis data for PfPREX from Plasmodb website



PlasmoDB gene: chr14.glm 129

Raw expression data for array element N150_19 (some time points are averaged values from multiple replicates)





P. falciparum chr14.glm_129



Downloads Tools Queries BLAST History CDs & Links Browse Data Sources

Plasmodium falciparum / CHR 14 / chr14.glm_129 / Expression 40% identity to 99% of (U23502) POM1 [Plasmodium chabaudi chabaudi] protein coding gene predicted by GlimmerM

THIS PAGE:	Microarrays- Ben Mamoun time series- Scripps/GNF Lifecycle- DeRisi HB3 study- Mass spec. data- Features(G)-
LINKS TO OTHER DATA:	Summary Gene model DNA/RNA features Protein features NRDB Motifs Sequences Expression GO assignments Pathways Orthologs/Paralogs 3-D structures MR4 reagents Links
USER COMMENTS:	View/Add comments

licroarray elements I	inked to this gene	back to top-<>
element_id	array description	array size
N150_19	DeRisi P. falciparum 70-mer oligo array, version 1	7392
N150_19	DeRisi P. falciparum 70-mer oligo array, version 2	7744
N150_19	MRA-452, 23K (MR4 P. falciparum long oligo)	23232

Ben Mamoun Plasmodium falciparum developmental time series study back to top-

2003 Lifecycle Study using the Scripps/GNF Malaria Array
*** None ***

back to top-<>

DeRisi Plasmodium falciparum HB3 intraerythrocytic cycle study

back to top <>>

PlasmoDB gene chr14.glm_129 / Array element(s): N150_19 DeRisi Plasmodium falciparum HB3 lifecycle Study



x-axis

Time in hours after adding synchronized culture of HB3 parasites to fresh blood

blue plot:

averaged smoothed normalized log base(2) of Cy5/Cy3 for chr14.glm_129

grey plot:

averaged normalized log base(2) of Cy5/Cy3 for chr14.glm_129

A Warred darrent and marred ato



		L CONTRACTOR			a delete a series a s
redicted protein features					back to top
PF 3_5_exonuclease PF DNA_pol_A PF Ribosomal_S4e TMAP TOPPRED2 signal peptide				0	
predicted epitopes		1 1	1. 1.	1 1 1	1 11
				REAL REAL R	
hydropathy plot	howwww	manymym	www.analytime	www.manhorenahilipation	ananyaaman haanna
hydropathy plot AA sequence	howwww	www.thanky.th	-	n as in the second of the second	www.www.white.anth
hydropathy plot AA sequence Pf chr14.glm_129 Ok	aa	Ok aa	ok aa	1k aa	nnn hann hann hann hann hann hann hann
hydropathy plot AA sequence Pf chr14.glm_129 Ok Description:	aa Place the mous	Ok aa	Ok aa	1k aa	1k aa
hydropathy plot AA sequence Pf chr14.glm_129 Ok Description:	aa Place the mous	ok aa ID:	Ok aa	1k aa iils.	1k aa

PlasmoDB Version: 4.1 PlasmoDB Date: 2003/08/01

Plasmodium Sequence Release Date: 2002/10/03

To contact PlasmoDB, send email to help@plasmodb.org

APPENDIX C

Figure C.1. Nucleotide sequence of PfPREX. (Accession no:- AAN36724.1)

9

3

. 1911-21

j.

TANK D.F.

No. of Street, Street,

Ċ,

A COLUMN

1888 N. C. C. C.

Ì

ATGAAGTAAGAAAAATAATATATATATTATATATATTTGTTATACAATGATTTCTATGATAT ACCACCATGCTTTTGTATAAGTTTTTATATTTTTTTTTGCTCGTACATTTGTCGTTA TGCATACSTIATAGAAATCAAAATAAAACTGATTCTTATTTAAAAACAAATTATAAATTA TTAAAAAAGAGGAAAAAAAAAAACCGTAGATATAAATTTCCAAAAAACAGAAAAGGA GACACATATAAAAACCGAATATTTTTTAGCAATGAGAAAAATGATGAGAAATTTTATCAAG ACCANTAATATTCTTATTATGAAAAAAGAAATAAGATCAAATTAAAATAGTGCTACTAGC AAAAAATATGAATTTATAGAAACAGATATTAAGATAACGTTAAAGTATTGTCCATTTGT CCTCCACATAAATATAAAATATGATAA'ATGTATAAACATGAAATATTTAAAAACACAGGA AATAGTTATTGICATAGATGEGGATATAAAGGAAGTTECTATGATTTCAAATTGAAAATG GGGGATCTAATAACTAGTAATTTTGAAAGTACIGTAGTACATAATAATAATTTTTATGAA ACTTTAAAAAAGTTTTTAAEAGGTTTTTCTGTTATGGAATTTCAATCTTTTGAAAGTTCE GGTAAATTTGAAAAACACGAATGTATTATATTTCCATTTAAAAAAAGTAGATGAAAAA AATATGATAGAAACGAATGGAATCAACACTAACATGGAACAAAATGAACAACAACAACAAC GAACTGGTACGAATAAAAATTAGAAGUTTAAAGGATAAAGGTTATATGAGATTATATCCT GAAGAAATAGTTTTGACAGAAGGAGAAATTGATGCCATGACAATAAGTCAAGAAACAAAA TATCCASCTATATCTTTACCAAATGGALCAAAATCTTTACCTATATATTTATTACCATAT TTAGAAAGGTTTAAAAAAAAAAATACATTTATGGCTAGATTTTGATAAAGCTGGAAAATCAAGC GTCTTCAATTTCGTTAATAAAATCGGATTAGGAAGAACCAACGTAATTACGGATGCCAAT GTTCACTATTTAAATCCAGATGTTTTTAAGAGAGAGACAAAAGTCCAGGTTAACCAAAAG AGTCTTCTCCTTACATCAATCCCTTCTAATGCTATGGAGATTCTTCAAAAAGAA GAAAATATGCATAATATATATGATACTACTAATAATGACTATATGGATAATAAAAATACTG <u>ACTAATAATTTAAAAAGTATAACTAGTGACAAAATAAAGAAAAAAGAAGAAATTGATCTA</u> TTTAATGGACNGAAAATTTCTAGCAATAATATNAATGTAAATATACTAAAAAACAAAAAA AACGAAACAGATAATAAACAAAATAAGGAAAATAAAGTGATAATAATTTGAAAGAAGGA ATGGAAAAAAAAGAGATACAAAAATGAAATAAGTGDAATTGAAGATAATAATAATAATAAA ATTGAAGATAATATTTCTTATTTTGTTGATAATAATAATATGTATATAACCAAATAATAATATA ATTATAAAAGATGCGAATGATTGTTTGAAACATAATATTGATATAAGATTCTTTATAGAA ACTAGTGAAAACGTAAAACATAG TCAAATATTGAATTTTAATGATTTAAGGCAACGTATA **TTAGAAGAATTAAAATATCCTGATCGAATTAATGGTGTAAAAAGTAAAACCATTCCATCA** TTAAATAAA'IATTTATATGGATTAAGAATGGGACAATTATCTATAIGGACTGCTTCTACA **GGTGTAGGAAAAACTACCTTATTGTCTCAACTCTCTTTAGATTATTCTATCCAAGGAGTA** TCAACATTATGGGGATCATTTGAAATAAATGTAAAATTAGGAAAAGTCATGTTAAAAT TTTGAACTGTTACCTCTTAAATTTTTTAAAGTTCCATGGTAGTACAAATATTGATCAGGTT ATAGATGCTATGGATTATGCTGTGTGTATGCATATGATGTAAAACATATAATTATTGATAAT TTACAATTTATGTTAAATATAAAATAAATTTTCTGATATATGAATTACAAAATATTGCT ATAGATAAATTTAGATCCTTTACTACAAAACAAAAATGTTCACACAAACTTTAGTTGTTCAT CCAAGGAAAGAGGATAACAATTTATTATCTATTTCATCCGTTTTTGGTAGTGTAAAATCA ACTCAAGAAGCAGATAATGTGTTTATTATTATACAAAGACATGTATCAAAAACAAATGAAACT **GTTTTTTATAGATATTAAAAAAAATAGATTTAAAGGAAGTTTAGGGAAAATACCTTAT** GTAGT^CTCAAATACTTATTTACCATCAAATAAATTTTCT^CCTTCATCATTACAAAATAAT ACTACTAATAATAATTTTATACAAAATGATAATTTAAATTTCACTTTATGTGATGAATAT GATTATGANACAGCTAGCTGATGAATATGAATCAAAACATGCTTTTAAAAAAAGGTTAT APTGACAATACCGATATGAATAACAATAAAAAATGTTACTCTTTATGTGGATAGTTTACAG ANTATAAAAACTATATCCACTGATGATAAAACTAATGATGATAAAACACATGTAAAACAA **ATAAAAAGCATTAAAAAACAATGAACGTAAAAATACATTAAAGATTGATGGAAACAAATCA TTAGGTTCTACACAAAATTTAAATTACCAAAATGAAAATAAAAACAATAATAATAATAAT** ANTAATAATAAGAGTAATCAAGAAAATGGAAAAGAATAACATAGATGATAAATCATCAACA GGAAATAATAAAAATATAAATGGCAACTCAAAAGGTAATAATAATAATAATAATAATAAT ATTAAAAATATTAAACACATCGGCACAAAATAACATTCCTTTTAAAGATACAATATGG TCATATACACTAACTAATGAAGGATTAATAAAATTGTGTGAAGAACTAAAAGATGAAGAA GCAGAAAAACTTAAAAATCGAGTTGTCGTTCTTTCGATGAGAAATTGTATAATTGATAAT CCACCAAAAAATTTAAAAAAAATGGATATATTTATTAGTATTTTACAAAAATATACCAAAG GAATATATTACAATTAANAGTGGACAGAAATATGGACCAAATAAAGGACTACCAAATGAC AAATCAAAAACAAAAAAAAGGAACCACATAATAATAATAATGTTGGAATATAAAAGTT **GGATCTAACATCGGTGATGGACAAAATACATCATCATCATGTATGAATATTAATAAAA**ATA CCCCAAACATTATTACCAAATGATAGAAATGATTCAAACTCTCATAGTTATAATAATATT AATTATACAATGGTTAAAAAATGGTAACGAAGGAAATATAAATGATTATCTAAATGAAA TATGAAAATAATGTAGATAACCATCATAATGATGAAAATTACAAAGAAATACATAAAGGAT GATGATATAGAATCTAGATTCTTTATAAATGATAATAATTATAAATGAAAAATGAAAAT TTAATATAAAAGATATAAAAATATTGTGGATTAGATATTGAAAACAACAGGTTTAGAAGTG GATA: GTTTAATATAAACAAAAAAAAAAATATTAGATGGATTAAGAAAAGTATTAGAAAAT AAAAATATAATAAAAATAATAAAAATGGAAAATTTGATGCAAAATTTTTTTGCTTCATAAT AATTTCAAAAATTGAAAATATTTTCGATACATATAGCTAGTAAAATTATTAGATAAAAAT AAGAATATGTATGGATTTAAATTAAATAATATAGTTGAGAAAATATTTAAATGTAATATTG AAAAAAAAAYTTACATATAGTAAAYGAYATTGAAAATAAAYGTAYCCYACCYATAYGY GATATGGAATTAAAATGGTATTAAAGTAGATTTAGAAAAATTTACAAAAAAGTACAAATGAA ATTTCTAATAAACTAATAGAAAANTACATCTGATTCCAATTTAAAAAAATTTTCTAAATCAT GAAGAAATTATAAGTTTACGTAATTATAGAAGATTATATAAATTGTATTCAGCTTTTTTAT TTAAAAACATTCTCAGGAAGATTTAGTAGTGAAAAACCAAATTCACAAAATACCTAGA CAAAAAAATATAAGAGAAATATTTATACCAAACGATAATAATATCTTCATTATAGCTGAC TTTAAACAAATCGAGCTTAAAATTGCAGCAGAAATTACAAATGATGAAATTATGCTTAAA ATACCAGATATTAATAAAGAAGATAGACATATAGCAAAAGCTATCAATTTGGTCTTATA TATGGTATGAATTATGTCAACTTAAAAAATTATGCAAATACGTATTATGGTTTAAATATG TCTTTAGATCAATGTCTATATTTTTATAATTCTTTTTTGAACACTACAAAGGAATATAT AAATGGCATAATCAAGTAAAAAAAAAAAAAGAGCTTTACAGTATTCTACACTTTCAAATAGA AAAGTTATATTCCCATACTTCTCTTTTACAAAAGCTCTTAATTATCCTGTACAAGGTACA TGTGCTGATATTTITAAAATTAGCTCTAGTAGATTTATATGACAACTTAAAGGATATTAAT GAACAAGCTCTTAAAATTTTAGTACAATCTATGGAAAATTCGGCTTCTTACTTTTGAAA AAAGTTAAATGTGAGGTTTCTGTAAAAATAGCTGAAAATTGGGGGATCAAAGGATTAACAT AAATTTATAGATAIT

Figure C.2. Amino acid sequence of PfPREX.

MLLYKFYFLYFLLVILSLCIRYRNQNKTDSYLKTNYKLLKKRKKYENRRY KFPKNRKGNNNNIYDAINNNNTNNIYNNDTYKNRIFFRNEKNDENFIKTN KYSYYEKRNKIKLNSATSTFVSKYYKININDVYNYLHRKKYEFIETDIKI TLKYCPFCPPHKYKYDNMYKHEIFKNTGNSYCHRCGYKGSFYDFKLKMGD LITSNFESTVVHNNNFYEEEEEKITLNDVKVYNMNLLYSKEAENARNYLM NVRKLSIDTLKKFLIGFSVMEFQSFESSGKFEKHECIIFFFIKKVDEINM IETNGINSNMNKMNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN KGYMRLYFKNVRNEMKLFFFGDHLIKNSEEIVLTEGEIDAMTISQETKYP AISLPNGSKSLFIYLLPYLERFKKIHLWLDFDKAGKSSVFNFVNKIGLGR TNVITDANVHYLNPDVFKRRQKSRLTKKSLLLTSMASNAMEILQKDKFEN MHNIYDTTNNDYMDNKILSNNLXSISSDKIKKKEEIDLFNGQKTSSNNIN VNILKNKKNETDNITNKENKSDNNLKFGMEKKEIQNEISVIEDNNNNKNN IIENNNDDMSEKIKVEKSIEDNISYFVDNNIMYIPNNITIKDANDCLKHN IDIRFFIETSEKVKHSQILNFNDLRQRILEELKYPDRINGVKSKTIPSLN KLGKVMLNOFCGKNLEKNIELFDIYADKFELLPLKFLKFHGSTNIDQVID AMDYAVYAYDVKEIIIDNLQEMLNINKFSDIYELQNIAIDKFRSFSTNKN VHITLVVHPRKEDNNLLSISSVFGSVKSTQEADNVFIIQRHVSKTNETVF FIDIKKNRFKGSLCKIPYLYNKENMTIKEISINNFNEHVVSNTYLPSNKF SSSSLQNNTTNNNFIQNDNLNFTLCDEYDYMKQLADEYESKHAFKKGYRS KLDPNLRVNNIDMIKTEIIDNSDMNNNKNVTLYVDSLENIKTISTDDKTN DDKRDVNEIKSIKNNERKNTLKIDGNKSLGSTQNLNYENENKNNNNNNN NKSNQEMEKNNIDDKSSTGNNKNINGNSKGNNNINNNNKNKNSSSSSSS SNYNNNEGIKNILNTSAQNNIPFKDTIWSYTLTNEGLIKLCEELKDEEAE KLKNRVVVLSMRNCIIDNNSSIKDIRTFIKINKLNIKTAGKNLKKMDIFI SILONIPKEYITIKSGOKYGPNKGLPNDKSKNKIKEPHNNNMLEYNKVGS NIGDGQNTSSSCMNINKIYSEEENNIYINNNNNNNKEPQTLLPNDRNDS NSHSYNNINYTMVKNGNEGNINDYJNRNYENNVDNHHNDEI'TKKYIKDNI INVDDNIIKKKDIFKLKNENNEITECAFEYFESKKKFDDDIESRFFIIND NNYNENINLIYKDIKYCGLDIETTGLEVFDENIRLIQIAVENYPVIIYDM FNINKKDILDGLRKVLENKNIIKIIQNGKFDAKFLLHNNFKIENIFDTYI ASKLLDKNKNMYGFKLNNIVEKYLNVILDKQQQNSVWNNSLLNNNQLFYA ARDSSCLLKLYKKLKEEIKKENLHIVNDIENKCILPICOMELNGIKVDLE NLQKSTNEILNELNIEKDNLKKKLKDENINVNSQQQVUKALQKNNVRDIS NKLIENTSDSNLKNFLNHEEIISLRNYRRLYKLYSAFYLKI PLHINTKTN KIHTTFNQLKTFSGRESSEKPNLQQIPRQKNIREIFIPNDNNIFIIADFK OIELKIAAEITNDEIMUKAYNNNIDLHTLTASIITKKNIPDINKEDRHIA KAINFGLIYGMNYVNLKNYANTYYGLNMSLDQCLYFYNSFFEHYKGIYKW HNQVKQKRALQYSTLSNRKVIFPYFSFTKALNYPVQGTCADLLKLALVDL YDNLKDINGKIILCVHDEIIIEVNKKFQEEALKILVQSMENSASYFLKKV KCEVSVKIAENWGSKD

Figure C.3. Amino acid sequence of P. chabaudi PomI sequence (Accession no:- AAA84746.1).

NSFKIKRERKRKNKVINLLNPNSASLSSSEESNADASSNSNALASQQENLDQSDTEKEKENDVYFIDNGI MYIWNKIFVKDANDCLKNNIDVQFFIKNSEKVKHSQILNFNDLRQHILEELKYPDRINGVRSKTIPSLNK FLYGLRMGELSIWTCSTCVGKTTLLSQLSLDYCIQGVSTLWGSFEINNIKLAKVMLNQFCGKNLEKNIDL FDLYADKFELLPLKFLKFHGSTNIDQVLDAMDYAVYAYDVKHIIIDNIQFMININKFSDIYELQNIAIDK FRSFSTNKNVHITLVVHPRKEDNNLLSISSVFGSVKSTQEADNVFIIQRQISKNNEAVFFIDIKKNRFKG SLGRIPYLYNKENMTIKEMPINYLNDFLSNNGSIGGSSSNNYLNKNGSYSNVLQNISSASNNNMDFTLCD EYDYMKQLSDEYESKHAFRRYNNKTCKCVADSGGLSLLKSSTLGSGTRRGDIFSEIKGNGKNAENEDIEK EKEKCNSKISELLKQNDEKVAHTKTTNNSVTAQSLGSNEIRSKEKNTKKKSVTENVGLQNNDSTNLKGGN KNSEINTDTMNNKNVTKGKYSSYLLSNEGLEKLCKEF

Figure C.4. Amino acid sequence of PyPREX.

-**..** .

MPWYLFYFFYIFFINLALGIRRRKNRLKNEYYIKINYKLLKKKKNDLKKKFGIINGFVSKPNEKLLKKQF RAYYKRNNDIEWGSNEKNCYNNLNKIKNKLNAMSTFVSKYYKININDVYNYLNRKKYEYIETDVKITLKY CPFCPPHKYKYDNMYKHEIFKNTGNSYCHRCGYKGSFYDFKLKMGDLVTSNFESPIVNDTYEEEKITFND VKVYNMNILYSKEAENARNYLINERKINFETLKXYYIGFSVMEFQSLENSGKFEKHECLVFPFIRKISNT NENENNCKSIYKGIGKTEDEGDIYEVVRIKVRSLKDKGYMRL/YPKNVRSDMKLFFFGDHLITNSDEVVLT EGEIDAMTIFQETNYFALSLPNCSKSLPIYLLPHLEKFKKIHLWLDFDKAGKSSVFNFVNKTGLGRTNVI TDANVHYLNENAFKIKRERKKKNKIINLLNPNSASLLPNEDSNVDVSTDSNCKSDYQEELGQSNSEKEKE KNSENDVYFIDNGIMYIWNKISIKDANDCLKNDIDIQFFIKNSEKVKHSQILNFNDLRQHILFELKYPDR INGVKSKTIPSLNKFLYGLRMGELSIWTGSTGVGKTTLLSQLSLDYCIQGVSTLWGSFEINNIKLGKVML NQFCGKNLEKNIDLFDLYADKFELLPLKFLKFHGSTNIDQVLDAMDYAVYAYDVKHIIIDNLQFMLNINK FSDIYELQNIAIDKFRSFSTNKNVHITLVVKPRKEDNNLLSISSVFGSVKSTQEADNVFIIQRQISKNNE AVFTIDIKKNRFKGSLGRIPYLYNKENMTIKEMPINYLNDFLSNNSSIGSNNSNNYLNKNGSYSNVLQNV SSVSNNNMDFTLCDEYDYMKQLSDEYESKHAVRKYNNKIGKCVADSGGLSLLKSSTDRSETRRGDIFSEI KEYEKNTENEDIEKEKEKYNSKISELLKKSDEKLTNTKTINDNTTAQSLSSNEIRSKEKNTKKKSVTENV GFQNNDSINLKSGNKNSDIDADKMDDKNLITRKFSSYLLSNEGLEKLCKELKDDEKGKFQNVVISISMRK CVINNNSPIKDIRNFIKVNKLNIKTSGNNLKKKDVFINVLOSIPXEYISIITEGDREKYENKNGGDTSGN

 $\label{eq:start} \begin{tabular}{lltrown} \b$

Figure C.5. Amino acid sequence of PbPREX.

MPKFKTAIRORKNRLKKAYYIKINYKLLKKKKNDLKKKFGIINGFVSKPNEKLLKKOFRA YYKRNNDIEWGSNEKNSCNNLNKIKNKLNAMSTFVSKYYKININDVYNYLNRKKYEYIET DVKITLKYCPFCPPHKYKYDNMYKHEIFKNTGNSYCHRCGYKGSFYDFKLKMGDLITSNF ESPIVNDTYEEEKITFNDVKVYNMNLLYSKEAENARNYLINERKINFETLKKYYIGESVM EFOSLENSGKFEKHECLVFPFIRKISNTNENENNCKSIYKGIGNTNDEEDIYEVVRIKVR SLKDKGYMRLYPKNVRSDMKLFFFGDHLITNSDEVVLTEGEIDAMTIFQETNYPAISLPN GSKSLPIYLLFQLEKFKKIHLWLDFDKAGKSSVFNFVNKIGLGRTNVTTDANVHYLNENT FKIKRERKKKNKIINIJNSNNASFLPSEDSNVEVSTNSNCKGGHQEELDQSNTEKEKEKI SENDVYFIDNGIMYIWNKISIKDANDCLKNDIDIOFFIKNSEKVKHSQILNFNDLRQHIL EFLKYPDRINGVKSKTIPSLNKFLYGLRMGELSIWTGSTGVGKTTLLSQLSLDYCIQGVS TLWGSFEINNIKLGKVMLNQFCGKNLEKNIDLFDLYADKFELLPLKFLKFHGSTNIDOVL DAMDYAVYAYDVKHIIIDNLQFMLNINKFSDIYELONIAIDKFRSFSTNKNVHITLVVHP RKEDNNLLSTSSVFGSVKSTOEADNVFIIOROISKNNEMTIKEMPINYLNDFLSNNSSIG NNNSNNYLNKNGSYSNVLONVSSVSNNNMDFTLCDEYDYMKQLSDEYESKHAVRKYNNNI GKCVADSGGLSLLKSSTLGSGTRRGNJ FNENKEYGKNTENEDIEKDKERYNSKI IELLKQ NNEKVTNTKTTNDSTTVOCLSSNEIQSKEKNTKKKSVTENVGLHNNDSTNFKSGNKNSDI DVDKMDNTNLITGNFSSYLLSNEGLEKLCKELKDDEKGKFQNIVISISMRKCVINNNSPI **KDIRNFIKVNKLNIKTSGNNLKKKDVFINVLQSIPKEYISIITEGGREKNGGNTSDINDI** IKATCKEIFFESOROGNINTNNNKGDEYNNVRNMDEGGENNAIKNNTNNEMYNNQVDPRL DDKIIKKYIDDNIINVDGNIVKRCGLFKLINDKDKVKIEENLYYYEPEKNFNDNIETRFF TINDKNYNEKINYIYNGIKYCGLDMETTGLEVFGEKIRLIOISVENYPVIIYDMFNITNN NILDGLRKILNDENIVKIIQNGKFDTKFLLYNNFNITNIFDTYIASKLLDKNKNMYGFKI. NNIVEKYLSVYLDKQQQNSVWNNSLLNNNQLFYAARDSSCLLKLYKKLSEQIKAENMQIV NDTENKCILPICDMELNGIKVDLENINKSTDOILNELNVERDKLKKELKNEDINVNSCOO ILKALQDNNVRDLSNKLIDNTSDANLKNFINHNAVILLRNYRKLYKLYSAFYLKLPLHIN KKTNKIHTTFNQLKTFSGRFSSEKPNLQQIPRQKNIREIFIPTENNIFIIADFKQIELKI AAEITNDEIMLKAYNNNIDLHTLTASIITKKCINEINKEDRHVAKAINFGLLYGMNYVNL RTYANTYYNVKMNLDQCLYFYNSFFEHYKGLSRWHNIVKQTKALEYSTLSNRKVVFPYFS FTKALNYPVOGTCADILKLALVELYKNLRHINGKIILCVHDEIIIEVDKKHQEEALRILV **ESMENSASFFLKKVKCEVSIKIAONWGTKD**

Figure C.6. Partial amino acid sequence of PcPREX.

MGLIKLGIGREKNELKKANYIKINYKLLKKKKDLKKNIGIINELVSKPNEELLKKOFRAY YKRNNNIEWSKNENNSYNNLNKTKNKLNAMSTFVSKYYKININDVYNYLNRKKYEYIETD VKITLKYCFFCPPHKYKYDNMYKHEIFKNTGNSYCHRCGYKGSFYDFKLKMGDLVTSNFE SPIVNDTYEEEKITYNDVKVYNMNLLYSKEAENARNYLMNERKINFETLKKYYIGFSVME FOSLENSGKFEKHECLVFFFIRKITTTDENENNCKSIYKGAGKTEDEGDTYEVVRIKVRS LKDKAYMRLYPKNVRSDMKLFFFGDHLITDSDEVVLTEGEIDAMTIFQETNYPAISLPNG SKSLPIYLTPHIEKYKKIHLWLDFDKAGKSSVFNFVNKIGLGRTNVITDANVHYLNENSF KTKRERKRKNKVINLLNPNSASLSSSEESNADASSNSNAFASQQENLDQSDTEKEKENDV YFIDNGIMYIWNKIFVKDANDCLKNNIDVQFFIKNSEKVKHSQILNFNDLRQHILEELKY PDRINGVRSKTIPSLNKFLYGLRMGELSIWTGSTGVGKTTLLSQLSLDYC1QGVSTLWGS FEINNIKLAKVMLNQFCGKNLEKNIDLFDLYADKFELLPLKFHGSTNIDQVLDAMDY AVYAYDVKHIIIDNLOFMLNINKFSDIYELONIAIDKFRSFSFNKNVHITLVVHPRKEDI IYFLFHQFLVVMTIKEMPINYLND#LSNNGSTGGSSSNNYLNKNGSYSNVLQNISSASNN NMDFTLCDEYDYMKQLSDEYESKHAFRRYNNKTGKCVADSGGLSLLKSSTLGSGTRRGDI FSEIKGNGKNAENEDIEKEKEKCNSKISELLKONDEKVAETKTTNNSVTAQSLGSNEIRS KEKNTKKKSVTENVGLONNDSTNLKGGNKNSEINTDTMNNKNVTKGKYSSYLLSNEGLEK LCKELKEDDQEKFKNVIISMSMRNCVINNNSPIKDIRNFIKVNKLNIKTSGNNLKKKDVF INVLQSIPKEYISITEGDTSGSNDSINCTQKETFSGSQRQGNMSTNKNTGGEYSNMGSMD EEGDNNG1KNDMNNEMYDNQIEARLDDEITKKYMDDNIINVDGNIVKKCARFKLINDKDN TEIEENLYYYERKKTIMIILRLDFLMETTGLEVFGEKIRLIQIAVENYFVIIYDMFNIYN NNILDGLRXILNDENIVKIIQNGKFDTKFLLYNNFNITNIFDTYIASKLLDKNKNMYGFK LNNIVEKYLSVYLDKQQQNSVWNNSLLNNNQLFYAARDSSCLLKLYKKISEQIVAENMQI VNDIENKCILPICDMELNGITVDLE 1. Street

e,

Ś

Ń

S.

. 1

14.5

Section 1

Marrie -

Figure C.7. Amino acid sequence of PkPREX.

MVRCCLF1FY11L1HYVLAVRNKSRFKADFYLKTNCELVKRRKNGSKAYFKRKVRDEDLGLRSFYDQLRG GRSNCSSAGNHDNHVDRRNSGALRATSTFVSKYYKININDVYSYLNRKKYEYIETDVKITLKYCPFCPPH KYKYDNMYKHEIFKNTGNSYCHRCGYKGSFYDFKLKMGDLVTSNFESTVVNNTTYEEEKT."PNDVKVYNM NLLYSKEAEGARKYLVEERKLNLETLKKYYIGFSIMEFOSLENSGKFEKHECLTFPFIKKANDMNSIGLN GNRNNTNEKDSYEIVRIKVRSLRDKGYMRLYPKNVKDEMKLFFFGDHLVSNSEEIVLTEGEIDAMTVNQE TNYAAISLPNGSKSLPIYLLPYLERFKKIHLWLDFDKAGKSSVFNFVNKIGLGRTNVINDTNVQYLDEQV FERKRKNMLTKGGLLLPLMIGDNAIGVAEQKQDAIKENTQSGEKNGNTECIAGGNEPNTAVDPISGTKKS $\texttt{DLQITEKAKSEGDGKNNSICGDTGNKQEEAQKKVNEESKVRAFHFVQNNIMYIPNNIVVKDANDCLRHNI$ DIRFFIENSEKVKHSQILNFNDLRQNILEELKYPDRINGIKSKTIPSINKFLYGLRMGELSIWTGPTGVG KTTLLSQLSLDYCIQGVSTLWGSFEINNIKLGKVMLNQFCGKNLEKNIELFDLYADKFELLPLKFLKFUG STNIDQVLDAMDYAVYAYDVKHIIIDNLQFMLNINKFSDIYELQNIAIDKFRSFSTNKNVHITLVVHPRK EDNNLLSTASVFGSVKSTQEADNVFIIQRQVSKINEIVFFIDIKKNRFKGCLGRIPYLYNKENMTIKEMS IGYLNDAIPSSGYGTKVTTPPSNFVPNVGPLRGGLDFTLCDEYDYMKQLSEEYESKHAMRKYRVGADGRV SGVGSASTNNFNASSSVHRAPNECRNDSGATDSLRNNQNNDNKSVNQVCDAEDDPTSNNRVVNTNDKSEE CRIKKNMLKGDDQGRTTTPVGKTTKSDKPSGD11HVQNSPNTKREKQEAAPDV2KQNVSSYRLSSEGITK LCEEIKDNKNEMLKDREITISMRNCVINKDSTIKDIRNFIKTNKLNIKTAGKNLKKVDVFIAILQSIPKE YITIKYGGGDVEKDNFDKNSVGNRTGENVKRTKVHSTLAENHNGVITGCNPSGALPVGGFTKNNFNIVSS VKESSASGHNIGHSLRGYVSHEGKEDALSGEKQYSEEIKSIYGEEVTKRYIQDNIINVDDNIVKRSGMEK LEGDNKMVSNEKLEYYEPVKKFDDDIESRFFLINDNNYNEKINNIYKNVTHCGLDIETTGLEVFDEKIRL IQIAVENYPVIIYDMFNITKESILTGLREILKNEKVVKIIQNGKFDAKFLMHNNFQVDNIFDTYIASKLL DKNKNMYGFKLNNIVEKYLNVTLDKQQQNSVWNNSLLNNNQLFYAARDSSCLLKLYKXLKSEICKENMET VNDIENKÇILPICDMELNGITVDLESIKKSTNEILSELNIETSKLKAELKDEEININSQQQVLKALQNNN VRDISNKLIENTSDSNLKNFLNHKEVVLLRNYRRLYKLYSAFYLKLPQHINKKTNKIRTTFNQLKTFSGR FSSEKPNLQQIPRQKNIREIFIPQKDNIFIIADFKQIELKIAAEITNDDIMLKAYNNNIDLHTLTASIIT KKPIADVNKEDRHIAKAINFGLIYGMNYVNLKNYANTYYNLNMNLDQCLYFYNSFFEHYKGIYRWNNQIX QIRGLEYSTI.SNRKVIFPYFSFTKALNYFVQGTCADILKLSLVELYKNI.KPIHGKTILCVHDEIIIEVDK KYQEDALKTLVESMENSASFFLKKVKCEVSVKIAQNWGSKE

Figure C.8. Amino acid sequence of PvPREX.

MVRCRFFLFYLILIHYALAVRNKSRPKADFYLKTNCELVKRRKNGSKAYFKRKARGEDIG SRARYDDSQLWGRRNRCSGGNRVMRSNRERGALRATSTFVSKYYKININDVYSYLNRKKY EFIETDVKITLKYCPFCPPHKYKYDNMYKHEIFKNTGNSYCHRCGYKGSFYDFKLKMGDL ITSNFENTVVDETYEEEKITENDVKVYSMNLLYSKEAEAARKYLTEERKLNLETLKKYYV GFSIMEFQSLENSGKFEKHECLIFPFIRKANGASSLGLMGSKHSTNEIDAQENASYEVVR IKVRSLRDKGYMRLYPKNVKDEVKLFFFGDHLVGSSEFVVLTEGFIDAMTVSQETNYAAI SLPNGSKSLPIYLLPYLERFKKIHLWLDFDKACKSSVFNFVNKIGLGRTNVITDANVHYL DEQLFERKRKNLLTKGGLLVPLTVGDSPIGVAEQKQGGIKENTQRGEENGHSENTLDGRS PNTAADPTSGTNKADLKTAEKEGDRRKNFICGETGSTQEGEQKKASENCKAKTLHFVQNN $\texttt{IMYLPNSIVVKDANDCLKHNIDVRFFIQNSEKVKHSQILNFNDLRQNILEELKYPDRING$ VKSKTIPSLNKFLYGLRMGELSIWTGPTGVGKTTLLSQLSLDYCIQGVSTLWGSEEINNI KLGKVMLNQFCGKNLEKNIDLFDLYADKFELLPLKFLKFHGSTNIDQVLDAMDYAVYAYD VKHILLDNLQFMLNINKFSDIYELQNJAIDKERSFSTNKNVHITLVVHPRKEDNNLLSIA SVFGSVKSTQEADNVFITQRHVSKTNETVFFIDIKKNRFKGSLGRIPYLYNKENMTIKEM SIGYLNDSISSSGYGANGTPPSSNFLFSVNPPRDGLDFTLCDEYDYMKQLADEYESKHAV RRYRLGADGKANGVGSGSTNRPNASSLKCSQNVNSTDGGLVDSSRNNQNRDNKPVNQVGG VEDDPTRSNHVGSANGRGEKGGIKNILKGTDRCDTAPPVGGTPTSDKPSGGSSHVCTPAK PTGEKQELTADMSKQSVPSYRLSSEGIIKLCEEIKDNKGETLKDREITISMRSCVINKDS TIKDLRNFIKTNKLNIKTAGKNLKKVDVFIAILQSIPKEYITVKYVGGDVGKGNPTNNNV RMGGGENVKGGRLNGALAGNHNSVNGGGSPSRAIPIGGPTKNDLNIVTTARESPVNSHNL GHISSGHGSAPL/TGDGQVGEKHYSEEIKSLYGEEVTKRYIODNIINVDDNIVKRSGTFK1 EGDDKMVGSEKLEYYEFVKKFDDDIESRFFLIHDSNYNERINHIYKNVTHCGLDIETTGL EVFDERIRLIQIAVEDYPVIIYDMFNITKESILTGLREVLKNEKVVKIIQNGKFDAKFLM HNKFEVANIFDTYLASKLLDKNKNMYGFKLNNIVEKYLNVTLDKQQQNSVWNNSLLNNNQ LFYAARDSSCLLKLYKKLKSEICRENMGTVNDIENKCILPICDMELNGITVDLESLSKST NEILSELNEETSKUKAELKDEEINVNSQQQVLKALQNKNVRDVSNKLIENTSDSNLKNFL NHREVVLLRNYRRLYKLYSAFYLKLPQHINKKTNKIHTTENQLKTESGRESSEKPNLQQI PRQKNIREIFIPQKDNIFIIADFRQIELKIAAEITNDDIMLKAYNNNIDLHTLTASIITK KAIADVNKEDRHIAKAINFGLIYGMNYVNLKNYANTYYNLNMNLDQCLYFYNSFFEHYKG IYRWHNQIKQMRALFYSTLSNRKVIFPYFSFTKALNYPVQGTCADILKLSLVELYKNLRP INGKIILCVHDEIIIEVDKKYQEDALKILVESMENSASFFLKKVKCEVSVKIAQNWGSXE

Figure C.9. Amino acid sequence of TaPREX.

MEPTYYTNQNINETTLRIPKKRLIGRSKVAFLFFLTRLILISSLLQFSESVKEDLSQGSF SISKRGRKTPRFEIPCNIFEKLNTIHKLNAVDKFSMISMKSLKQLSNLRNLIRFKYNKGS RNYKNRNFYKCLFVKNLNEPOIDGGNIIYGSQFYAYSPSSYENVPSYNGSYGLDESNTFV SNHYRNMCEDIVEYLNRKKIEYNESPVKYTLKFCPFCPPHRFKSDNLYKHEIFKNSGNSY CHRCGYKGSLFDFKAHMGDLPGAFDIASDPITNPFGAKKVQKDPDVTITDVNKFEDNLYN NPEYACVLEYLTKTRGLKTHVLKKYHVGAGTFKFKSVTGKLEPEKCVVFPWLSTFHKPYA DMGDDVEIRDESEGFDYNREYTTDKLNVNRIKIRSIYDKSKIKILPRGGTWGMFGGELIQ EALEKDSIVLSECEIDAMSIFQETGRITTSLPNGANSLPLALLPKLEKFNEIYLWMDFDA PCQSSISHFANKLGIQRVKIVHPLNTPSTTCTTTGTTTTGAGTTTKGTKSTTTATPTSK SKMMKDANEVLLSGGVDMNNYFKNATVMTHSQILTFNDIKQLVYNELSDPVTTCGIKSIT MPGLSNLLKGHRRGELTVWTGSTGSGKTTLLSQLSLDYCLQGVSTLWGSFEINNVRLAKT MLRQFSGRNLENSLNEFDYYANKFNELPLRFLKFHGSTNIDIVLDAMDYAVYVYDVQHII IDNIQFMLSNYSCPSQNSSYGGYSSTKDVYELQNRTIEKFRRFVTNKNIHLSLVVHPRKE ADGIQLGLSSVFGSVKSTQEADNVIILQNILNESRCIDVKKNRFAGLLGRVYFKFDPVSL TAQUUKVTEYLNEYTTNNFTKAKVKKDQTESFTVPSGAFPINGGSEVKSKVSLSNFKINL PQVNSTNLSTEPMNIPITTSTMMTEDYQSRAVPASTTTNYRERERNNLGLTGIDLLNLKP VSQYVNKIIQQNYSTTSNGNQESGDSSTKQDSTESSGVDLTKSNMTGTGDNSNSSGKCRK KSSKSDPVETEYLSMRGLKTSKSSTIKEMRDFVKKNGLGELVKTAGKGIKKEAIYENIRS TIPQSAVVNTKQTEKLDASGESETVETIKVEDEGSPKSEVHSFGVSTSKSLLKLERINVV PNNDIDEFKLKYLSKLFKSLYKGESQPPSQPEKVASNDQPPIPVDVYYSVNYDPLLIDGI IYVDSYEKLESLKPLFKDPKAVGVDIETTGLDHNTNQIRLVQLSVPNQPSVIIDLFKLNT NNPESDNVV9GRNELIKCEWLKNLFKSKDTVKVFHNGKFDINFLRVYGFEFEGPIFDTMV ASKLLVASRYISCKLTHVSERYLNIVLDKTQQYSDWSTLQLFEEQLLYSARDSFVLLPLY VILEHLLKINNLADIASVENKCILATSDMELNGIKVDEDKLRILQDELKQEHQEISNDLY SQLNHSEINLNSQKQVLEKLQELKIMDR3KKKI1SDTSESTLIRNMSNPIISSLREYRKA NKALTAFTQKLPNHINFITSRIYFNYNQLGAESGRFSCDGPNLQQVPRDKKFRECFVAPK GSKFVIADFSQIELRIAAEIANDPKMIQAYQQNVDLHSLTASILKNKSINEVNKEERQLA KAVNFGLIFGMSINGLRMYAFTGYNLKLTQTEAKEIYTSFFNNFKGILNWHNSVKNSRPTMVRTLCNRLSVFESFSFTRSLNYPVQGTSADITKETMARLVDSVKPLNAKIIICVHDEII LEVPEDNAEKALKMLIDTMVKSGEKYLKKVPVEAMGSIGDSWADKS

Figure C.10. Amino acid sequence of TpPREX.

MEPTYYTNOALNEPTLRISKKRYYKVAFLFFLTRLILISCLFPFSECVKTDLSOGSF SISKRGRQFPPPRTLCNIFKTFNTIHKFSPVDKLNMISVKSLKDISNLRNLIRFRYN KRDRNYRNRSFYKCLFVSNLNKPQMESGKITDGGSFYSYPPTSYPNMPSYNGSYGLD ESNTEVSNHYRNMGEDIVEYLNRKKIEYNESPIKYTLKFCPFCPPHRFKSDNLFKHE 1FKNSGNSYCHRCGYKGSLFDFKAHMGDLPGAFDIASDPITSPFGAKKVQKDPDVTI EDVNKFEDNLYNNPEYACVLDYLTKTRGLKPHVLKKYHVGAGSFKFKSVTGKLEQEX CVVFPWLSTPNKLYPNDEVETQTDDSEGFDYNREFTTDKLNVNRIKIRSIYDKSKIK ILPRGGAWGMEGGHLIQEAQEKDSIVLSEGELDAMSIYQETGRLTISLPNGANSLPL ALLPKLEKFNEIYLWMDFDAPGQSSISHFANKLGIQRVKVVHPLNTFTTTATTTKGT **TTSTKGAGTTTKGSKSTTATTQKTTSKMLKDANEVLLSGGVDMNNYFKNATVMTHSQ** ILTFSDIRQLVYNELSDPVTTCGIKSITMPGLSNLLKGHRRGELTVWTGSTGSGKTT LLSQLSLDYCLQGVSTLKGSFEINNVRLAKTMLRQFSGRNLENNLNEFDYYANKFNE LPLRFLKFHGSTNIDIVLDAMDYAVYVYDVQHIJIDNLQFMLSNYSGPSQNSSFGGY SSSKDIYELQNRTIEKFRRFVTNKNIHLSLVVHPRKEADGIQLGLSSVFGSVKSTQE ADNVIILQNILNESRCIDVKKNRFAGILGRVYFKFDPVSLTAQEFKVTEYLNEYTTN FTKPKVKKDQPDTPTVPSGGFPINGTPETKPKVSLSNFKLNLPQVNGSNVSNEPIGI PNVGTMTSEDLQSRPTPATSTNYKDKERNNVGLIGIDLINLKPITQQINKIIQQNYT TSSNSNSDSSDSSCKPEGEDT3GDDNTNSNLVGTNDNSNS3GKGGKLKTKSELVEEE YLSMRGLKITKNSSTKEMRDFVKKNGLGELVKTAGKGIKKDAIYENIRSIIPQSAII TTKPTEEEVESVKEDSSVDTTKVEDESSARSDVHSFGVSTSKSLLKLERVNVTPNND ITEFKLKYLSKLPKSLYKGEPPS\$\$\$KPTPVTTGDQVSIPVDVYYSVNFD9LLIDGIT YVDSYDKLQSLEPLFKDPKAVGVDIETTGLDHNTNQIRLVQLSVPNQPSJIDLFKL STNNPDAESVVPGRNELIKCEWLKKLFKSKETVKVFHNGKFDINFLRVYGFEFEGPI

Ĵ.

4

A156214

4

्रम्

4

FDTMVASKLLVASRYISCKLTHVSERYLNIVLDKTQQYSDWSTLQLFEEQLLYSARD SFVLLPLYVILEHLLKINNLSDIASVENKCILATSDMELNGIQVDEDKLKFLQDELK QEHQEISNTLYSQLNQSEINLNSQKQVLEKLQELKIMDKSKKKIISDTSESTLIRNM SNFIISSLREYRKANKALTAFTQKLPNHINPITSRIYPNYNQLGAESGRFSCDGPNL QQVPRDKKFRECFVAPKGSKFVIADFSQIELRIAAEIANDPKMIQAYQQKVDLHSLT ASILKNKNINDVNKEERQLAKAVNFGLIFGMSINGLRMYAETGYNLKLTQNEAKEIY TSFFNNFKGILNWHNSVKNSRPTMVRTLGNRLSVFESFSFTRSLNYPVQGTSADITK ETMARLVDLVKPLNAKIIICVHDEIILEVPEDNAEKALKMLIDTMVKSGEKYLKKVP VEAMGSIGDSWADKS

PbPREX	1	
PYPREX	1	MPWYLFYFFWIFFMNDALCIPREENRLFNRYFFENYRELKEDLEERREGIINGFVSE
PCPREX	1	MGLIKUGIORENPIKKANTIEINYELLEEKE-ODEINIELNEL
pfprex	1	LLYKFEFLEFLEVHUSSCITY-ROOM TESTLETUK LLEVERKYENRREKFPKNRKG-
PKPREX	1	WYRCOLELFYLIIITHYVHAVIN-ISPHADEMENTICE VIRKING-SHAWFKRKVRDED
PHOPEY	49	
PUPPEN	61	
PYPREA	10	
PCPREA	48	
piprex	59	- INNN YDAINNNTN IYNNDYKNRIFR HADDENFKINKISIYEKNN IN SY
PRPREX	59	LGDASFYDDD GRSNCUSAGHDDHWDRUNSGARR-DT
PhPREX	92	STORYSIZYYN I MHNIWYN MUZHAY CYMETIOWN UMRYOPEO PPHRY RYONWYCHEN SRAT
PUPPEX	104	STEVSKYVETHINDVYNYLMREFYEY DETDVKUTLEYCPECEPHEYEYDMMYEHELEKNT
POPREX	91	STEVSKYVETNINDVYNYLNPEFYEYIETDVETTLEVCPECPPHEYEVDNMYEHETEENT
nforey	118	STEVERYVETNINGVYNVI HELEVERIETDERITLEVCPECEDBEVEYDMMYRHEIFKNT
PEPFEX	07	STEVERYVETNINDVYSVINDERVEYIETDVETTI.RVCDEC PDHEVEYDMMYKHEIFKNT
PRPREA	91	
Pbprex	152	SNSYCHROGYKGSFYDFELEMGDLITSNFESPIANDTYEEFELTFNDVKVYNMNLE
PYPREX	164	GNSYCHROGYFGSFYDEFLEMGDLVTSNFESPIVNLTYEEFEITFNDVEVYNMNLL
PCPREX	151	GNSYCHROGYLGSFYDFELEMODLVTSHFESPIVNETYEEERITFNDVEVYNMNLL
pfprex	178	GNSYCHROGYRGSEYDFELEMGULITSNFEUTVHMNNFYEEEEERITLUDVKVYNMNLL
PKPREX	157	GNSYCHROGYRGSFYDFFLEMGDLVTSNFESTVYNN-TTYEEDELTFNDVEVYNMNLL
		Standbline and Local Courts Make of Manual and and the
Pbprex	208	YSKEAENARNYLINERKINFETLERYYIGFSVMEFQSLENSGEFEKHECLVFPFIRKIS <mark>N</mark>
PYPREX	220	YSKEAENARNYLINERKINFETLEKYYIGFSVMEFQSLENSGEFEEHECLVFPFIRKISN
PCPREX	207	YSKEAENARNYLMNERKINFETLELYYIGETVMEFQSLENSGKFEEHECLVFPFIRKITT
pfprex	238	YSKEAENARNYLMI <mark>V</mark> RELSIDTLEEFILGEGVMEFQD F ESSGEFECHECIIFPFIKEVDE
PkPREX	214	YSKEAFGAFKYLVEERFINLETIKKYYIGFSIMEFOSLENSGKFEKHECLIFPFIKKAND
PHOPEY	268	
PUPREA	200	TATENENAL CONTRACTOR CONTRAC
PYPREA	200	
PCPREX	207	
piprex	290	
PRPREX	2/4	
PbPREX	312	PKNVRSDMKLFFEGDHLFTNSDEVVLTEGEIDAMPIFOETNYPATSLPNGSESLPTYLLF
PYPREX	324	PKNVRSDMKLFFFGDHLITNSDEVVLTEGEIDAMTIFOETNYPAISLPNGSKSLPIYLLP
PCPREX	311	PKNVRSDMELFFFGDHLITDSDEVVLTEGEIDAMTIFØETNYPAISLPNGSESLPIYELF
pfprex	358	PKNVPNEMELEFEGDHLIKUSEETVLTEGEIDAMTISOETKYPAISLPNGSESLPIYLLP
PKPREX	312	PENVEDEMKLEFFGDHLVSNAEFIVLTEGEIDAMTVNOETNYAAISLENGSESLEIYLLE
Pbprex	372	QDEKFEKTHLWLDFDKAGESSVENFVNEIGLGETNVITDANVHYLNENTFETEREREK
Pyprex	384	HLEKFKKIHLWLDFDEAGESSVENEVNEIGLGETNVITDANVHYLNENAFEIKREBKE
PCPREX	371	HLEEXKKIHLWEDFDEAGESSVENFVNKIGEGETNVITDANVHYENEN <mark>S</mark> EEIKREEER
pfprex	418	YLERFECKIHLWLDFDEAGESSVENFVNKIGLGETNVITDANVHYLNPEVERRQKSRLT
Pkprex	372	Y BERFFFTHLWLDFDFAGESSVFNFVNKTGLGETNVIN TNVOVIDEOVFER FKNMLT
DEDDEV	120	
PUPPEN	430	
POPREX	442	
PEPREX	429	
DIDLEX	410	

Figure C.11. Multiple alignment of amino acid sequences of Plasmodium PREX homologues

PKPREX	432	GGLI PINIGDAIGNAEQ-KAIAIKENNSGEKNGUTENIGG-NEPNTAN
		Linds, S.J. and State States States
Pbprex	469	DOSNTEKELEIISEMD
Pyprex	481	GOSHSDINDE
PCPREX	468	DOSDERKER
pfprex	537	LFUGO ISSNNINVNILKNUKNUTONITNKENKSDNNLKEGMEKKEIQNEISVIEDNNN
PKPREX	482	PISGT KSDLQITEKAK BEGGGKNNSICGDTGN
Pbprex	485	
PYPREX	497	Wight Dirght Mirwinkers Frankhulter Bird Anders Bird Anders Bird Anders Bird Anders Bird Anders Bird Anders
PCPREX	480	WYELDING IMPLIMINE THEY PARTY IN THE WILL BE THE THEY AND THE WILL BE THE THEY AND THE
pfprex	597	NKNNIIENNNDDMSEKIKVEKSIEDNIST VDANIST PUNII HJANUTUHHIDIRGE
PKPREX	516	-KQEEAQKKVNEESKVRAFHEWQ
PHOPEY	518	TENDESTAVENDATE IN MODERATINE A STATE OF DESTINATION OF A DESCRIPTION OF A STATE OF A DESCRIPTION OF A DESCRIPANTA DESCRIPTIO
PUPPEY	530	TENSPEVENSOTINFNDLPOHILFELEYPDRINGVESETIPSLNKFLYGLEMGELSIWTG
POPREX	513	TENSEEVENSOTUNENDLROHTLEEDE YPDRINGARDETIPSLNEELYGLEMGELSIWTG
oforex	657	TET SERVENSOILMENDLPOR LEELEYPDRINGVESETIPSLNEYLYGLRMGELSIWTG
PEPREX	566	TEMSEKVEHSOTLNENDLRONILEELEYPDEINGERSETIPSLNEELYGLEMGELSIWTG
		and the second of the second
Pbprex	578	STGVGETTILEQUESTBYCTOGVSTEWGSFEINNI KLGEVMENOFCGENLEINI DEFDLYA
Pyprex	590	STGVGETTELSQLSLDYCIQGV3TLWGSFEINNIKLGEVMLNQFCGENLEENIDLFDLYA
PCPREX	573	STGVGETTIL SOLSLOYCIOGVSTLWGSFEINNIFLAEVMENOFCGENLEENIDLFDLYA
pfprex	717	STGVGETTLESQL3LDYCIQGVSTLWGSFEINNVELGEVMLNQFCGENLEENIELFDEYA
PKPREX	626	PTGVGKTTLLSQLSLDYCTQGVSTLWGSEEINNIKLGKVMLNOFCGKNLEKNIELFDLYA
Phopery	638	
PUPPEY	650	DEFELLPLEFLEFHESTHIDOVLDAMDYAVYAYDVEHLIDDLOFMENINETSDIYFLON
PCPREX	633	DEFELTPLEFLEFHESTNIDOVLDAMDYAVYAYDVEHILIDNLOFMLNINEFSDIYELON
pfprex	777	DKFELLPLEFLEFHGSTNIDGVIDAMDYAVYAYDVEHIIIDNLOFMLNINEFSDIYELON
PKPREX	686	DRFELLPLEFLEFLEFLGSTNIDOVLDAMDYAVYAYDVEHIIIDNLOFMLNINEFSDIYELON
		A supervised of the second
PbPREX	698	IAIDEFRSFSTNENVHITLVVHPREEDNNLLSISSVFGSVESTQEADNVFIIQRQISENN
Pyprex	710	TAIDKFRSFSTNENVHITLVVHPEKEDNNLLSISSVFGSVKSTQEADNVFIIQRQISENN
PCPREX	693	TAIDKERSESTNKNVHITEVVHPREEITYFEFELITYFEFEL
pfprex	837	IAIDKFRSFSTNENVHITLVVHFREEDNNLLSISSVFGSVKSTQEADNVFIIQPHVSHTN
PKPREX	746	IAIDEFRSESTNENVHITLVVHERFEDNNLLSIASVEGSVESTOEADNVEITOEOVSETN
DEDDEV	750	
POPREX	758	EPTH REPRINT IN DRIVEN AND AND AND AND AND AND AND AND AND AN
PYPREX	770	AVERTED KNOPPERSSTER FOR THE END THE END THE DATE STORES TO SUME AND THE
PEPREA	897	
PLPTEX	806	ETWEETDIEKINDEKCOLCELOVI.VNKENMTIKENSICVENTATDSSCVCTKVTTD
EKENDA	000	
PbPREX	791	NGSYSNVLONVSSVSNNNMDFTL&DEYDYMROL&DEYESEHAVEL
Pyprex	830	NGSYSNVLONVSSVSHNIMDFTLCDEYDYMEQLSDEYESEHAVRF-YNNFEGECVADSGG
PCPREX	764	NGSYSNVLODIS ASNNNMDETLCDEYDYMEQLSDEYESEHAFER-YNNETSKCVADSGG
pfprex	954	S-LONTTHINFIQ-DDNMLTLCDEYDYMKOLADEYESNHAF& GORSEDDNLRVNNI
PKPREX	861	PENFNPNVGPD-RGGIOFTLCDEYDYMEQLSEEYESKHAMRE-NRVGADGRVSGV
DEDDEV	050	
POPREX	850	
POPPEN	009	
nfnrow	1010	DMUKTETTDNSDMNNNKNUTLYUDSLENTKE ISUDDUTMEDUDIUMETUSTUMEDUDI
PLPTEX	914	CASTNN-PN
L AF NEA	514	
Pbprex	887	NEEYNSKIIICHAR CHNH VTNIKUTNUSTIVICUSINDUQ
Pyprex	926	FERYNSKI <mark>SELLIKSDEFL</mark> TNTKTTNENTUA, SISSNE IRSKEKN-RKKKSV
PCPREX	860	KERCNSKISELLEONDEKWAHUKTINNSVIA,SIGSNEIRSKEKN-IKKKSV
pfprex	1072	IDGUKSUGSTQNLEYENENENNNNNNNNNKSKSNQ-EVEKKNUDDKSSTGNNKUINGNSKG
PkPREX	951	
DL DE DU	0.00	
POPREX	938	
POPPEY	911	
pfprex	1131	NNU NNU KUKNSSSS SSSSNYN NEGIKUILNTSAON MEPEDTIWSYT THEGI I

PKPREX	997	TPVEKTTKSEKPSGDIIHVENSPETKEEKQEAAPENP-EQNVERESKEETKE
PhPREX	989	FREERODEKCES ONEVISIONEK
PVPREX	1028	CKELKDDEKGREONVVISISMEKCVIMNNSPIEDIRNEIKVNKLNIETSGNNLEKKDVEI
PCPREX	962	CKELKEDDCEKFENVIISMEMENCVIDNNSPIKDIPNFIKVNKLNITTSGNNLKKKDVFI
pfprex	1191	GEELKDEEAEHLKHRVVUSMENCIIDHNSSIKDIPTFIITHKLNIKTACKULKIMDIFI
PKPREX	1052	EETKENKN MLTERE TISMRUGVLOKET I NELENTUN IN INTARKNUM VOVEL
PhPREX	1049	
PVPREX	1088	NVLOSI PNEY ISTITE GORENYENENGGETSCNNGIIKST
PCPREX	1022	NVLQSIPKEWISTTE ODTSCSNUS NGU
pfprex	1251	STICN PKEWITIKSGQKYGPNKGLPNDKSENKIKEPHNUNMLEYUKV SHIGDGQNTSS
PkPREX	1112	ALLOSIPREYITIKYGGHEVIXDNPDHESVHRTEEVKRTKHE
Pbprex	1085	
PYPREX	1128	OKETHIE SOF COMINTAL IKED
PCPREX	1050	OKETESGEORDOMMSTMKNTEGEYSUMGSMDEEGDN
pfprex	1311	SCMNINKIYSEENNYINN-NNNNKEPQTLLPNDRNDSNSHSTNNINYTMVKNG
PKPREX	1158	LANNHNGVITGGNPS AMPVGGPTKMNFNIVSSVK-ES-SASGHMGHSLRGYVS
Pbprex	1121	NATENNTNNEMYNNOVEPRIDIK <mark>I</mark> FRYIDDNIINVDGNIVFRCGEFREINDEDEM <mark>E</mark> I
PYPREX	1164	NVIKNNMINIEMMNNOVI PRLDIKIIKKYMDDNIINVDGNIVKECOREKLINDKDNVKI
PCPREX	1086	NGIKNIMIH MYDNOIBARINDEITTEEYMUDRIENVUDRI VKEGARDENIAUDMINTE
pfprex	1368	EENIMYL RNEENWINHAUDT KITTEKITEKITEKUT DUT TEKKUT KUTTEKITEKUT
PKPREX	1211	HPG BRALSGERONSEERKSIIGEONIR ROOM ENNED TO VERSEEGE - WILL VS
PHPREY	1179	RENTAVERBUNDNI FUREETINERUNEETIN TANCI RACOLIMEATOLEVECERTE
PVPREX	1222	EENLYYYEDEENENDNIETRFELINDKNYNEKINY LYNGEKYCGLDMETTGLEVFGEKIR
PCPREX	1144	EENLYYYERK IMILERIDFLMEUTGLEWEGENTE
pfprex	1425	FCAFEYFESKEKEDEDIESEFFIINENNYNENINLIYKDIKYCGLDIETTGLEVEDENIE
PkPREX	1270	NEK EYYEYVEKEDED DESPEELINEN NYNEKINNEYKNETH GLIFETTGLEVEDEKIP
PbPREX	1239	LIGISVENYPVILYPMENITNNNILDGLERFILNDENNVREIONGREUTFELLYNNENITN
PYPREX	1282	LIQIAVENYPVIIYDMFNITHMNILDGLPKILNDENIVKIIQNGFFDTEFLLYNNFNITN
PCPREX	1180	LIQIAVENYPVIIYDMENITNNUILDGLEEILNDENIVEIIQNGEEDTEELLYNNENITN
pfprex	1485	LIQIAVENYPVIIYDMFNI <mark>NKK</mark> DILDGLREMI <mark>E</mark> NKNIIKIIQNGKFL <mark>A</mark> KFLLHNNFKIEN
PkPREX	1330	TO LAVENY PVILYDMENT KESTAT TO BE TAKNE KOVELLONG KELAKI BARDAN HIN EOMDA
PBPREX	1299	IFDTYIASKLEDENKNMYGERENNIVERYLSVYLDROQQNSVWNNSLENNNOLFYAARDS
Pyprex	1342	IFDTYTASELLDENENMYGFFLDNIVEFYLSVYLDROOONSVWNNSLLNNNOLFYAARDS
PCPREX	1240	TFDTYLASELLDEDEDEDMYGFELDNIVEEYLSVYLDEOOONSVWNNSLLDNDOLFYAARDS
pfprex	1545	TEDTYLASELLDENENNYGFELINIVEKYLN <mark>UI</mark> DEQQQNSVWNNSLINNNDLEYAARDS
PRPREA	1390	TEDLETASETHIDENENVELGEEDUNEVENTEN TEDUCOONSVENUSUUUUUUEETAAMUS
Pbprex	1359	SCHENLYKKLSEQTKAENMOTVNDTENKCTIPTCDMELINGTKVDHEN <mark>HAKST</mark> DQILLINEIAN
Pyprex	1402	SCILLELYEELSEOT TAENNO I VIIDI ENKCI LPI COMELINGI KVDLESINKSTOCI I AUGIAI
PCPREX	1300	SCLLKDTKKLREQTVAERIQTVNDTERIKCTLPTCOMELINGTVDLE
PKPREX	1450	SCLIKTYKERSETCKEINETVIDIENKCILPICDMELIGITVDLESIKKSINETLEEN
Dhonou	1410	
PUPPEX	1419	VERDET FOR NED INVISOOFT FALL DINVE DATE FOR STATE OF STATE
POPPEN	1402	
pfprex	1665	TERENTER KINDON NVNSKOV AVALOKUNVRID ISNIHL ENDST STAAND ISNIEEFT ST
PKPREX	1510	TETSELFRELEDEEININSOO, VLKALONNNVEDISNKLIENTSESNLENEINEKEVVLL
Phoper	1479	
PVPREY	1522	RNYPKLYELYSAFYLELPLHINEETNEIHTTENOLKTESCHESSELPHLOOIDPOUNTPH
PCPREX		
pfprex	1725	RNYRRLYKLYSAFYLKLPLHIN <mark>T</mark> RTNKIHTTFNQLKTFSGRFSSEKPNLQQIPROKNIRE
PKPREX	1570	RNYPRLYKLYSAFYLKLF <mark>O</mark> HINEKTNKIHTTFNOLKTESGRESSEKPNLOOIPROKNIRE
Phppry	1530	
PVPREX	1582	IFIFTENNIFLIADFKOIELKIAAFITNDEIMLKAYNNNIDLHTLTASIITKKGINEINK
PCPREX		
pfprex	1785	DETEN <mark>DNNIFITADFFOIETETAAEITNDETMEEAYNNNIDLHTEPASITTEEH</mark> NI <mark>PDINE</mark>



Figure C.11. Multiple alignment of amino acid sequences of *Plasmodium* PREX homologues. Sequences were aligned using ClustalX 1.81. Black shading indicates identical amino acids; grey shading indicates conserved amino acid substitutions. The numberings on the left hand side shows the relative position of the amino acid in the respective sequences. Abbreviations: *P. falciparum* (Pf), *P. berghei* (Pb), *P. chabaudi* (Pc), *P. knowlesi* (Pk), *P. vivax* (Pv) and *P. yoelii* (Py). Introns boundaries were determined using ORF finder at http://www.ncbi.nlm.nih.gov/gorf/gorf.html.

APPENDIX D

Figure D.1. Full alignment of deduced amino acid sequences of DNA polymerase domains of DNA polymerases from family A

r7	1 DVVWTKALEELSDKH FPPELDFTDVGYTTFWS
73	1 DVVWTKALLELLSDKHWFPPELDFTDWGYTTFWS
biye03-12	1 DVVWTVKLUE FLTDLHWFPAGDDFTKYDADLFWR
Tvanonhage	1 DVEWSVALAKTFVPKIP
Synechocystis	GETHGNATYPEVLONAK DRRUNAHHETE GGVVL
F aquaticus	1 GILAKDLAWLAWREGLDLFPEDEDED
I.aquacicus	1 AAW ADDAKP ALHEAKAAVHD AGR WT EGVTS
Myccu E coli	1 KPL EDEKAL GONLKEDRG ANY HE RGIAF
E.COII	
S.typhimurium	
Rhizobium	1 KALEDESHLARANIKADI JARANIKATEDUPPYF
Thermotoga	
A.thallana	
Rice	
Dictyostelium	1 KEYFEDESULT WHNIADRAFTING COACHER COACHER
A.aeolicus	
A.pyrophilus	
polN	1 FQT0KCKCPVICFNAKD4VKIWIQFFNDGSWKHV
MmpolN	1 FWTIRCKCPVVCFNARDEVRIV OLITELDGWRRV
Mouse	1 QMRVTPKLMALTWDGFPLHYSDSHGWYLWPGRRDNLTEPPVSPIVESAAVICPIRAIES
Human	1 QMRWTPKLWALTWDGFPLHYSERHGWYLWPGRRDNLARLPTGTTLESAGVVCPTRATES
Fly	1 GMQHAPKLMSLCWEGYPLHYEREQGW FLWPFRSDSEGVDRLPMEQLLARCPVPEFAR
Miplp	1 KSRIPILFKLSWENSPVIWSKESGWCFNVP
PfPREXpol	1 RKVIENKNII IIII IQNGKIDAKFILHNNFKIENIFD
eta	1 RKESDKECSVVIYDFIQSYKIIIIILSCOISIEQSYE
SP6	1 CEELDQKIDETWAAFRPHMPMRIKSKPFKPQEKQE
M2	1 YDSUKKLPFPVKKIAKDUQLPLUKGDIDYHTERPVG
т7	36ESTEAVDIEHRAAWLLAKQERNGFP
Т3	36BSLEAVDVEHRAAWLLAKQERNGFP
phiYeO3-12	36PAGESVDLEHRAAWLLAKQERNGFP
Cvanophage	18 YOWSVDTEHEIARIMSWQEQMGFP
Synechocystis	36
T aquaticus	23 PWLLAY LDPSNTT-PEGVARRYGG
Myctu	36
Fcoli	36
S turbimurium	36
S. Cyphinairian Bhigchium	35NULLSYNLDAGTGAHGMDPLSEKFLGHTPIPYKDVAGSGK
Rhizobium	26 AVI EPNEKKENLDDLALKELGYKMTSYOELMSESEP
Thermotoga	30 TIATIBUT MARK WAS ADDIVCGYSLEAT TSDPKULGGTOTKEEAEFLGKIS
A. thalland	30
Rice	30
Dictyostellum	
A.aeolicus	
A.pyrophilus	36VERELEPRER
polN	36ADFIGLDPRIAAWLIDPSDAIPSFEDLVBRICERSITVING
MmpolN	36ADFVGLDPRVAAWLIDPSDIAPSFEDLVAARLEASIIVAFSS
Mouse	61 LYRKHCLEQGKQQLEPQEVDIAEEFLLTDSSAMwQTVEELGCLDVEAEAAMENSGLS
Human	61 LYRKHCLEQGKQQLMPQEAGHAEEFLLTDNSAIWQTVEELDYLEVEAEAKMENLKAAVPG
Fly	59 LSABKAESDBAFDMLPGQVEQHLGKREHYKKLSQKQQ
Miplp	32HEQVETYKAKNYVLADSVSQEEEEIRTHNLGLQCTG
PfPREXpol	36WIASKULDKNKNMYGFKLNNIVEKYLNVILDKQQQNSVWN
eta	36PPKVACWLDPDSQEPTLHSIVTSFFPHELPLLEGMETSQG
SP6	36QVDAANSFSDQNHTGVTLGADAFIHAER
M2	37
Т7	61FDTKAIEELYVELAARRSEELRKETETFGSWY
Т3	61FDTKAIELYVELAERRSELLRN TETFGSWY
phiVeO3-12	61FNTOAIELYVELAMKRAELIRKITETFGSWY
Cyanophage	43FDVRAAOALEGKLRLELDTUSDDWRETF
Sunachocystis	77OTIADLPL TAGOYC LDCYATYL ASK OKE
Taquatique	48EWTEDA ERAL AER FOTIKER
Muctu	77
Reali	77 NOLTENOTAL FACEYAMEDADATIONAL KAWPD
E.COII	
S. typnimurium	
Rhizobium	
Thermotoga	/8LFGFSFADVPVVKAANYSCEDADITYK YKTISLK
A.thaliana	83 MKTIFGKRKLKKDGSEGKIVVIPPVEELQREDREAWISYSELDAISTLKEYESWTKK
Rice	81 MKTIFGRKKVRKDGSEGKTISIEPVEKLQREDRELWICYSSLDSMSTLKEYESEKNK

Dictyostelium	71	IKDLFGTRKIKADGEEGKSIVIPPLERIQRDSKFILNWIEYSSLDSETTWLERENHIK
A.aeolicus	47	BMLNELDEERGEELLKTRTAKIFD
A pyrophilus	47	MINELE ERGEELLKTRTARIFG
nolN	78	
POIN	70	TERRACIANT CONTRACTOR AND CON
MUDOIN	10	TEREASKNILSONVERNIKA LIDATRIDESK
Mouse	118	QPLVLPAACAPKSSQPTYHHGNGPYNDVNIPGCWFFKLPHKDENNYNVGSPFAKDFLPKC
Human	121	QPLALTARGGPKDTQPSYHHGNGPYNDVDIPGCWFFKLPHKDCNSCNCGSPFAKDFLPK
Fly	96	RLETQYQGSGVWCNKVLDDCCFFLKLPHKN PSFR GNP SKDFLNKF
Mip1p	68	VLFKVPHPN PTFNCTNL TKSYNHFF
PfPREXpol	77	NSLLNNNOLFYARDSSCLLKWYKKKKEE
eta	77	IOSLGLNAGSEHSGRYRASVESI
SP6	64	
510	27	
MZ	21	
т7	93	QPKGGTEMFCHPRTGKPEPKYPRIKTPKVGGEFKKPKNKAQREGREPCEL
T3	93	OPKGGTEMFCHPRTGKPPKYPRIKIPKVGGFFKKPKNKAOREGREPCEL
phiYeO3-12	93	OPKGGKELFKHPRTGK PLPSYPRVWYPKOGGWYKKPRNKAOBEGLEPCDL
Cyanophage	71	
Cyanophage	110	
Synechocystis	110	DQIPELKEILKE DOI ERI AANDDROIRHCDYIQTLSQQLAEN LT
T.aquaticus	12	KGEERLLWLYEEVEKE SRVIARCHAT SYRIFYAY QALSLEVEAEVRQ
Myctu	110	ARIDSTALLGEORIE VQRVLAKORSASIAVELPMETELQSQFGDQURD
E.coli	112	QKHKGPLNVFEN DEED VPW SR PRNCVK PKV HNHSEELTLR AE
S.typhimurium	112	QQHKGPLNVFENISMU VPV SRVPRN VKI PAVIHKHSEEITLR AE
Rhizobium	111	AAAG-LTSVYER BRELEW ARE DARE TWEROTE SRLSGELAOGAAR
Thermotoga	114	HEAD-LENVFYKNEMD UNWARRANT NAWYN TEF YKKLEFFYCYN TEF
A thaliana	141	OT MOWIL DOK DUL OPTIMI DEVIDE DUDING THE DEVIDENCE TO A DEVIDE THE DEVIDENCE TO A DEVIDE THE DEVIDENCE TO A DEVIDE TO A DEVIDENCE TO A DEVID
Rico Rico	130	CHIDWIDDORPVDORTHLDFINETWKFGENVKWARDULTKEINAEIEKVAKAEQQV
Rice	139	EAKEWIFDDCPRGTMYDFYEEYWR FGALLVKMFTERL
Dictyostelium	131	RDMEWNQGTTMWDFYYLWWR FGHINTEN QROKKALIDY KSLEEKAYTDEE
A.aeolicus	72	KSPTAIVEMAFVREAKLEINEFPVUVEETTNKLKAVERETQK
A.pyrophilus	72	KSPPAINDAFVKEWAKING PULIET ESTLKDIERKTOK
DOIN	110	KDYGIWOLFRT AND TPH AVIONSHA LOW KEEKTSALLCAPHKE
MmpolN	110	
Manag	170	RAIG - LWOLF CITATION IF HAVE AVER AND RELEVANT ALL GARAGE
Mouse	1/8	EDGTLQAGPGGASGPRATEINKETSFWRNAHKRISSQMMVWLPRSATPR
Human	181	EDGTLQAGPGGASGPRALEHNKONSFWRNAHKRISSQMWVWLPRSALPR
Fly	144	AENVLSSGDPSCQAAARVIDHAROMSYWRNNRDRIMGQLWVWLDSQQHPN
Miplp	95	EKGVLKSESELAHQALQHNSSGSYWMSARERHQSQFWVPSCKFPNEFQSL
and the set of the set	107	
PfPREXpol	107	KKENLHIVND NKCILPHCDOMLNCHKMLENKOKSTNEILNEMNI
PfPREXpol	113	KKENLHIVNDHENKCILPECDUBLNCHKELENEQKSTNEILNEENI
PIPREXpol eta SP6	113	KKENLHIVNDUKNKCILPICDULINGIKZULEN QKSTNEILNEINI QKENLQDVFRKVEN SQYCHALLULIN IGFSTAECESQKHIMQAKUDA NDTPSIKHIGPYTPUTFEDILGNEDTIKOVI VDEGNPUTFUTFOTSIN DEHVUI BKDW
PfPREXpol eta SP6 M2	107 113 96	KKENCHIVNDUCNKCILPICDGULNGIKZILEN OKSTNEILNEINI OKENQDVFRKYEMESQYCDALGULNGIKZIAECESOKHIMQAKUDA NDTPSIKHIGPYTPVTFEDIPLGNRDTWKOVIYDFGWREVEFNDTEQSYLDEHGVLPKPW
PfPREXpol eta SP6 M2	107 113 96 63	KKENLHIVNDUTNKCILPICDULNGIKZILEN OKSTNEILNEINI OKENLQDVFRKZEMESQYCHALUELNGIGFSTAECESOKHIMOAKUDA NDTPSIKHIGPYTPVTFEDIPLGNRDTVKQVIYDFGWREVEFNDTEQSYLDEHGVLPKPW KQGLDRWTAGSDSLKGFKILSTKKFNKVFPKLSLPMDKEIRK
PIPREXpol eta SP6 M2	107 113 96 63	KKENLHIVNDUTNKCILPICDGELNGIKZELEN OKSTNEILNEINI OKENLQDVFRKZERSOYCHALLEIN IGFSTAECESOKHIMQAKUDA NDTPSIKHIGPYTPVTFEDIPLGNRDTVKOVIYDFGWRVEFNDTEQSYLDEHGVLPKPW KQGLDRWTAGSDSLKGFK ILSTKKFNKVFPKLSLPMDKERK
PfPREXpol eta SP6 M2	107 113 96 63	KKENLHIVNDIC NKCILPICOGU LNCIKVILEN QXSTNEILNENI QKENLQDVFRKYDM SQYCLALLI LNUIGFSTAECESQKHIMQAKUDA NDTPSIKHIGPYTPVTFEDIPLGNRDTVKQVIYDFGWRVEFNDTEQSYLDEHGVLPKPW KQGLDRWTAGSDSLKGFK ILSTKKFNKVFPKLSLPMDKENRK
PfPREXpol eta SP6 M2 T7	107 113 96 63 143	KKENLHIVNDIG NKCILPICDGULNGIKZILEN QKSTNEILNEINI QKENLQDVFRKYDMESQYCLALLULN IGFSTAECESQKHIMQAKUDA NDTPSIKHIGPYTPVTFEDIPLGNRDTVKQVIYDFGRRVEFNDTEQSYLDEHGVLPKPW KQGLDRWTAGSDSLKGFKULSTKKFNKVFPKLSLPMDKEURK
PfPREXpol eta SP6 M2 T7 T3	107 113 96 63 143 143	KKENLHIVNDIGNKCILPICDELNGIKYILEN OKSTNEILNEINI QKENLQDVFRKVEMESQYCLALLU IN IGFSTAECESQKHIMQAKUDA NDTPSIKHIGPYTPVTFEDIPLGNRDTKKOVIYDFGWROVEFNDTEQSYLDEHGVLPKPW KQGLDRWTAGSDSLKGFKILSTKKFNKVFPKLSLPMDKEIRK DTREYVAGNPYTPVEHVVENPSERDHQKKTOPAG-WVPTKYTDKGAPVVDDEVLEGVRV DTREYVAGNPYTPVEHVVENPSERDHQKKTOPAG-WVPTKYTDKGAPVVDDEVLEGVRV
PfPREXpol eta SP6 M2 T7 T3 phiYeO3-12	107 113 96 63 143 143 143	KKENLHIVNDIGNKCILPICDELNGIKKILEN OKSTNEILNEINI OKENLQDVFRKVEM SQYCHALELN IGFSTAECESOKHIMQAKUDA NDTPSIKHIGPYTPVTFEDIPLGNRDTWKOVIYDFGWREVEFNDTEQSYLDEHGVLPKPW KQGLDRWTAGSDSLKGFKILSTKKFNKVFPKLSLPMDKEIRK DTREYVAGAPYTPVEHVVENPSRDHFOKKIOFAG-WVPTKYTDKGAPVVDDEVLEGVRV DTREYVAGAPYTPVEHVVENPSRDHFOKKIOFAG-WVPTKYTDKGAPVVDDEVLEGVRV DTREYVAGAPYTPVEHVVENPSRDHFOKKIOFAG-WVPTKFTDKGAPVVDDEVLEGVRV
PfPREXpol eta SP6 M2 T7 T3 phiYeO3-12 Cyanophage	107 113 96 63 143 143 143 143 86	KKENLHIVNDIENKCILPICOGILNEIKKILEN QKSTNEILNEINI QKENLQDVFRKYDM SQYCLALHILN IGFSTAECESQKHIMQAKHDA NDTPSIKHIGPYTPVTFEDIPLGNRDTVKQV YDFGWR VEFNDTEQSYLDEHGVLPKPW KQGLDRWTAGSDSLKGFK ILSTKKFNKVFPKLSLPMDKEHRK DTREYVAGAPYTPVEHVVENPSRDHFQKKTOFAG-WVPTKYTDKGAPVVDDEVFEGVRV DTREYVAGAPYTPVEHVVENPSRDHFQKKTOFAG-WVPTKFTDKGAPVVDDEVFEGVRV DTRDYMEGAPYTPVEFVTFKPSRDHFQKKTOFAG-WVPTKFTDKGAPVVDDEVFEGVRV DTRDYMEGAPYTPVEFVTFKPSRDHFQKKTOFAG-WVPTKFTDKGAPVVDDEVFEGVRV DTRDYMEGAPYTPVEFVTFKPSRDHFQKKTOFAG-WVPTKFTEKGAPVVDDEVFEHVRV -VRHYFENAPFCKLR-EFNPTFRHHAWAFEHHRGWEPKETAGGQFKIDDEITREINT
PfPREXpol eta SP6 M2 T7 T3 phiYe03-12 Cyanophage Synechocystis	107 113 96 63 143 143 143 143 86 159	KKENLHIVNDIGNKCILPICDGLINGIKZILEN QKSTNEILNEINI QKEN
PfPREXpol eta SP6 M2 T7 T3 phiYeO3-12 Cyanophage Synechocystis T_aquaticus	107 113 96 63 143 143 143 143 86 159 121	KKENLHIVNDIGNKCILPICDGLINGIKYILEN QKSTNEILINEINI QKEN
PfPREXpol eta SP6 M2 T7 T3 phiYeO3-12 Cyanophage Synechocystis T.aquaticus Myctu	107 113 96 63 143 143 143 143 86 159 121	KKENLHIVNDIGNKCILPICDELINGIKKILEN OKSTNEILNEINI QKEN
PfPREXpol eta SP6 M2 T7 T3 phiYeO3-12 Cyanophage Synechocystis T.aquaticus Myctu E.coli	107 113 96 63 143 143 143 143 143 86 159 121 158	KKENLHIVNDIENKCILPICOGLINGIKKILEN QKSTNEILNEN QKENLQDVFRKYDM SQYCLALHILN IGFSTAECESQKHIMQAKHDA NDTPSIKHIGPYTPVTFEDIPLGNRDTVKQVIYDFGWR VEFNDTEQSYLDEHGVLPKPW KQGLDRWTAGSDSIKGFKIISTKKFNKVFPKLSLPMDKEHRK DTREYVAGAPYTPVEHVVFNPSTRDHQKKLOBAG-WVPTKFTDKGAPVVDDEVLEGVRV DTREYVAGAPYTPVEHVVFNPSTRDHQKKLOBAG-WVPTKFTDKGAPVVDDEVLEGVRV DTREYVAGAPYTPVEHVVFNPSTRDHQKKLOBAG-WVPTKFTDKGAPVVDDEVLEGVRV -VRHYFENDFCKLR-EFNPTRHHAWAFEHHRGWEPKETTAGQPKIDDEITREINT IETAAYEAGESINGSPKLIGTILFFKLGIDRKKSRKTKTGYSTDHATLEKLQ LEEEVFRLAGKQINIGPKCIQPKCIQVUFTELGCPKTKTKTGYTTDADALQSLFDK
PfPREXpol eta SP6 M2 T7 T3 phiYe03-12 Cyanophage Synechocystis T.aquaticus Myctu E.coli C.tubicut	107 113 96 63 143 143 143 143 143 143 143 143 143 14	KKENLHIVNDIGNKCILPICOGLINGIKKILEN QKSTNEILNEINI QKENLQDVFRKYDM SQYCLALMI LNIIGFSTAECESQKHIMQAKHDA NDTPSIKHIGPYTPVTFEDIPLGNRDTVKQVIYDFGWR VEFHDTEQSYLDEHGVLPKPW KQGLDRVTAGSDSIKGFK ILSTKKFNKVFPKLSLPMDKEHRK DTREYVAGAPYTPVEHVVFNPSRDHLQKKLQFAG-WVPTKYTDKGAPVVDDEVLEGVRV DTREYVAGAPYTPVEHVVFNPSRDHLQKKLQFAG-WVPTKFTDKGAPVVDDEVLEGVRV DTREYVAGAPYTPVEFVTKPSRDHLQKKLQFAG-WVPTKFTDKGAPVVDDEVLEGVRV DTREYVAGAPYTPVEFVTKPSRDHLQKKLQFAG-WVPTKFTDKGAPVVDDEVLEGVRV LTREYVAGAPYTPVEFVTKPSRDHLQKKLQFAG-WVPTKFTDKGAPVVDDEVLEGVRV LTREYVAGAPYTPVEFVTKPSRDHLQKKLQFAG-WVPTKFTDKGAPVVDDEVLEGVRV LTREYVAGAPYTPVEFVTKPSRDHLQKKLQFAG-WVPTKFTDKGAPVVDDEVLEGVRV LTREYVAGAPYTPVEFVTKPSRDHLQKKLQFAG-WVPTKFTDKGAPVVDDEVLEHVRV -VRHYFENAPFCKLR-ENPTSRHHAWAFEHHRGWEPKERTAGQPKIDDEILREINT IETAAYEAGESTUSSPKIGTLFFKLGDRKKSRKTKTGYSTDHATLEKLQ LEEEVFRLGGKQINIGSPKIQVUFFELGPAIGKTEKTGKRSTSAAVLEALR AAEAAYGVIGKQINIGSPKIQVUFFELGPKTKRTKTGYTTDADALQSLFDK
PfPREXpol eta SP6 M2 T7 T3 phiYeO3-12 Cyanophage Synechocystis T.aquaticus Myctu E.coli S.typhimurium	107 113 96 63 143 143 143 143 143 143 143 143 159 121 158 161 161	KKENLHIVNDIGNKCILPICDGELNGIKZILEN QKSTNEILNEINI QKENLQDVFRKVEM SQYCLALMELN IGFSTAECESQKHIMQAKHDA NDTPSIKHIGPYTPVTFEDIPLGNRDTVKQVIYDFGRRVEFHDTEQSYLDEHGVLPKPW KQGLDRWTAGSDSIKGFKILSTKKFNKVFPKLSLPMDKEHRK DTREYVAGAPYTPVEHVVINPSRDHQKKLQFAG-WVPTKFTDKGAPVVDDEVLEGVRV DTREYVAGAPYTPVEHVVINPSRDHQKKLQFAG-WVPTKFTDKGAPVVDDEVLEGVRV DTROYMEGAPYTPVEHVVINPSRDHQKKLQFAG-WVPTKFTDKGAPVVDDEVLEGVRV UTRYVAGAPYTPVEHVVINPSRDHQKKLQFAG-WVPTKFTDKGAPVVDDEVLEGVRV UTRYVAGAPYTPVEHVVINPSRDHQKKLQFAG-WVPTKFTDKGAPVVDDEVLEGVRV LTRYVAGAPYTPVEHVVINPSRDHQKKLQFAG-WVPEKFTEKGAPVVDDEVLEHVRV -VRHYFENAPFCKLR-EFNPTRHHAWAFEHHRGWEPKERTAGGQPKIDDEILREINT LETAAYEAGSINSPKIGTLFEKLGDRKKSRKTKTGYSTDHATLEKLQ LEEEVFRLAGKQINIGSPKOLQVVIFFELGUPAIGKTEKTGKRSTSAAVLEALR AAEAAYGVIGEENISTKLQTILFEKQGIKP-LKKTPGGAPSTSEEVLEELA LEKKAHDINGEAFNISTKLQTILFEKQGIKP-LKKTPGGAPSTSEEVLEELA
PfPREXpol eta SP6 M2 T7 T3 phiYeO3-12 Cyanophage Synechocystis T.aquaticus Myctu E.coli S.typhimurium Rhizobium	107 113 96 63 143 143 143 143 143 143 143 143 143 14	KKENLHIVNDIGNKCILPICDGULNGIKYILEN OKSTNEILNEINI OKEN
PfPREXpol eta SP6 M2 T7 T3 phiYeO3-12 Cyanophage Synechocystis T.aquaticus Myctu E.coli S.typhimurium Rhizobium Thermotoga	107 113 96 63 143 143 143 143 143 143 143 143 143 14	KKENLHIVNDIENKCILPICOGILNEIKKILEN QKSTNEILNEN QKENLQDVFRKYDM SQYCLALMILNI IGFSTAECESQKHIMQAKHDA NDTPSIKHIGPYTPVTFEDIPLGNRDTVKQVIYDFGWR VEFNDTEQSYLDEHGVLPKPW KQGLDRWTAGSDSIKGFKIISTKKFNKVFPKLSLPMDKEHRK DTREYVAGAPYTPVEHVVFNPSTRDHIQKKLOBAG-WVPTKFTDKGAPVVDDEVLEGVRV DTREYVAGAPYTPVEHVVFNPSTRDHIQKKLOBAG-WVPTKFTDKGAPVVDDEVLEGVRV DTREYVAGAPYTPVEHVVFNPSTRDHIQKKLOBAG-WVPTKFTDKGAPVVDDEVLEGVRV -VRHYFENAPFCKLR-EFNPTERHHIAWAFEHHRGWEPKETTAGQPKIDDEITREINT IETAAYEAGGESINTSPKLIGTILFFKLGIDRKKSRKTKTGYSTDHATLEKLQ LEEEVFRLIGKQINIG PKCIQVIFTELGC PKIKKTKGKRSTSAAVLEALR AAEAAYGVIGKQINIG PKCIQVIFTELGC PKIKKTKGGAPSTSEEVLEELA LEKKAHDIG
PfPREXpol eta SP6 M2 T7 T3 phiYeO3-12 Cyanophage Synechocystis T.aquaticus Myctu E.coli S.typhimurium Rhizobium Thermotoga A.thaliana	107 113 96 63 143 143 143 143 143 143 143 143 143 14	KKENLHIVNDIGNKCILPICDGULNGIKKILEN QKSTNEILNEINI QKENLQDVFRKYDM SQYCLALMULNI IGFSTAECESQKHIMQAKHDA NDTPSIKHIGPYTPVTFEDIPLGNRDTVKQVIYDFGWR VEFHDTEQSYLDEHGVLPKPW KQGLDRWTAGSDSLKGFK ILSTKKFNKVFPKLSLPMDKEHRK DTREYVAGAPYTPVEHVVFNPSRDHIQKKLQFAG-WVPTKYTDKGAPVVDDEVLEGVRV DTREYVAGAPYTPVEHVVFNPSRDHIQKKLQFAG-WVPTKFTDKGAPVVDDEVLEGVRV DTREYVAGAPYTPVEHVVFNPSRDHIQKKLQFAG-WVPTKFTDKGAPVVDDEVLEGVRV DTREYVAGAPYTPVEFVTFKPSRDHIQKKLQFAG-WVPTKFTDKGAPVVDDEVLEGVRV DTREYVAGAPYTPVEFVTFKPSRDHIQKKLQFAG-WVPTKFTDKGAPVVDDEVLEGVRV DTREYVAGAPYTPVEFVTFKPSRDHIQKKLQFAG-WVPTKFTDKGAPVVDDEVLEGVRV DTREYVAGAPYTPVEFVTFKPSRDHIQKKLQFAG-WVPTKFTDKGAPVVDDEVLEGVRV DTREYVAGAPYTPVEFVTFKPSRDHIQKLQFAG-WVPTKFTDKGAPVVDDEVLEGVRV DTREYVAGAPYTPVEFVTFKPSRDHIQKLQFAG-WVPTKFTDKGAPVVDDEVLEGVRV LEKANPFCKLR-EFNFSRDHIQKLQFAG-WVPTKFTDKGAPVDDEVLEHRVV LEKKSAHGAGEFNFSRDHIQKLQFAG-WVPTKFTTAGAPVLDELFLKLGL LEEVFRLGGEFNFSTKIGTLFFKGGKFFKTGGRSTSAAVLEALR AAEAAYGVIGKQINIG PKCIQVLFTELGVPKTKTTGGAPSTSEEVLEELA LEKKAHDIGGEFNISTKIQTTLFFKQGKFFLKKTFGGAPSTSEEVLEELA LEKKAHDIGGEFNISTKIQTLFFKQGKFFLKKTGGVSTSAQVLEDLA LAEEIYRIGG
PfPREXpol eta SP6 M2 T7 T3 phiYeO3-12 Cyanophage Synechocystis T.aquaticus Myctu E.coli S.typhimurium Rhizobium Thermotoga A.thaliana Rice	107 113 96 63 143 143 143 143 143 143 143 159 121 158 161 159 162 201 197	KKENLHIVNDIGNKCILPICOGULNGIKKILEN QKSTNEILNEINI QKENLQDVFRKVEN SQYCLALGULNGIKKILEN QKSTNEILNEINI NDTPSIKHIGPYTPVTFEDIPLGNRDTVKQVIYDFGRKVEFHDTEQSYLDEHGVLPKPW KQGLDRVTAGSDSIKGFKILSTKKFNKVFPKLSLPMDKEIRK DTREYVAGAPYTPVEHVVFNPSRDHQKKLQFAG-WVPTKYTDKGAPVVDDEVLEGVRV DTREYVAGAPYTPVEHVVFNPSRDHQKKLQFAG-WVPTKFTDKGAPVVDDEVLEGVRV DTRDYMEGAPYTPVEFVTFKPSRDHQKKLQFAG-WVPTKFTDKGAPVVDDEVLEGVRV DTRYNAFGAPYTPVEFVTFKPSRDHQKKLQFAG-WVPTKFTDKGAPVVDDEVLEGVRV DTRYNAFGAPYTPVEFVTFKPSRDHQKKLQFAG-WVPTKFTDKGAPVVDDEVLEHVRV -VRHYFENAPFCKLR-EFNPTRHHAWAFEHHRGWEPKERTAGGQPKIDDEILREINT IETAAYEAGESNIS PKCIGTLFFKLG DRKKSRKTKTGYSTDHATLEKLQ LEEEVFRLAGKQINIGSPKCIQVULFELGUPAIGKTEKTGKRSTSAAVLEALR AAEAAYGVIGKQINIGSPKCIQVULFELGUPAIGKTEKTGGAPSTSEEVLEELA LEKKAHEINGESNIS TKCIQTIFFKQGIKP-LKKTPGGAPSTSEEVLEELA LEKKAHDINGESNIS TKCIQTIFFKQGIKP-LKKTPGGAPSTSEEVLEELA LEKKAHDING
PfPREXpol eta SP6 M2 T7 T3 phiYeO3-12 Cyanophage Synechocystis T.aquaticus Myctu E.coli S.typhimurium Rhizobium Thermotoga A.thaliana Rice Dictyostelium	107 113 96 63 143 143 143 143 143 143 159 121 158 161 159 162 201 197 185	KKENLHIVNDIGNKCILPICDGULNGIKKILEN QKSTNEILNEINI QKENLQDVFRKVEM SQYCLALGULNGIKKILEN QKSTNEILNEINI NDTPSIKHIGPYTPVTFEDIPLGNRDTVKQVIYDFGWRVEFHDTEQSYLDEHGVLPKPW KQGLDRVTAGSDSIKGFKILSTKKFNKVFPKLSLPMDKETRK DTREYVAGAPYTPVEHVVFNPSSRDHQKKLQAG-WVPTKFTDKGAPVVDDEVLEGVRV DTRDYMEGAPYTPVEHVVFNPSSRDHQKKLQAG-WVPTKFTDKGAPVVDDEVLEGVRV DTRDYMEGAPYTPVEHVVFNPSSRDHQKKLQAG-WVPTKFTDKGAPVVDDEVLEGVRV UTRYVAGAPYTPVEHVVFNPSSRDHQKKLQAG-WVPTKFTDKGAPVVDDEVLEGVRV UTRYVAGAPYTPVEHVVFNPSSRDHQKKLQAG-WVPTKFTDKGAPVVDDEVLEGVRV UTRYFNAPFCKLR-EFNPTSRHHAWAFEHHRGWEPKERTAGGQPKIDDEILREINT IETAAYEAGSSNLSPKLIGTLFFKLGDRKKSRKTKTGYSTDHATLEKLQ LEEEVFRLAGKQINIG PKOLQVLFFELGPAIGKTEKTGKSTSAAVLEALR AAEAAYGVIGEFNISTKOLQTILFFKQGIKP-LKKTPGGAPSTSEEVLEELA LEKKAHDIGGEFNISTKOLQTILFFKQGIKP-LKKTPGGAPSTSEEVLEELA LEKKAHDIGG
PfPREXpol eta SP6 M2 T7 T3 phiYeO3-12 Cyanophage Synechocystis T.aquaticus Myctu E.coli S.typhimurium Rhizobium Thermotoga A.thaliana Rice Dictyostelium A.aeolicus	107 113 96 63 143 143 143 143 143 143 143 14	KKENLHIVNDIENKCILPICDELINEIKKILEN QKSTNEILNENI QKENLQDVFRKYDM SQYCLALM LNIIGFSTAECESQKHIMQAKHDA NDTPSIKHIGPYTPVTFEDIPLGNRDTVKQVIYDFGWR VEFNDTEQSYLDEHGVLPKPW KQGLDRWTAGSDSLKGFK ILSTKKFNKVFPKLSLPMDKEHRK DTREYVAGAPYTPVEHVVFNPSTRDH QKKLOBAG-WVPTKFTDKGAPVVDDEVLEGVRV DTREYVAGAPYTPVEHVVFNPSTRDH QKKLOBAG-WVPTKFTDKGAPVVDDEVLEGVRV DTREYVAGAPYTPVEHVVFNPSTRDH QKKLOBAG-WVPTKFTDKGAPVVDDEVLEGVRV -VRHYFENPFCKLR-EFNPT RHH AWAFEHHRGWEPKETTAGQPKIDDEITREINT IETAAYEAGESINTSPKLIGTTLFFKLGIDRKKSRKTKTGYSTDHATLEKLQ LEEEVFRLAGESINTSPKLIGTTLFFKLGIDRKKSRKTKTGYSTDHATLEKLQ LEEEVFRLAGESINTSPKLIGTTLFFKLGIPKIKTGKSTSAAVLEALR AAEAAYGVIGKQINIG PKOLQVILFELGC PKIKKTKTGYSTDHATLEKLQ LEKKAHDINGEFNISTKKIQTILFFKQGKP-LKKTPGGAPSTSEEVLEELA LEKKAHDINGEFNISTKKIQTILFFKQGKP-LKKTPGGAPSTSEEVLEELA LEAEIYRIG
PfPREXpol eta SP6 M2 T7 T3 phiYeO3-12 Cyanophage Synechocystis T.aquaticus Myctu E.coli S.typhimurium Rhizobium Thermotoga A.thaliana Rice Dictyostelium A.aeolicus A.pyrophilus	107 113 96 63 143 143 143 143 143 143 143 14	KKENLHIVNDIGNKCILPICDELINGIKYILEN QKSTNEILNENI QKENLQDVFRKYDM SQYCLALMI LNIIGFSTAECESQKHIMQAKHDA NDTPSIKHIGPYTPVTFEDIPLGNRDTVKQVIYDFGWR VEFNDTEQSYLDEHGVLPKPW KQGLDRWTAGSDSLKGFK ILSTKKFNKVFPKLSLPMDKENK DTREYVAGAPYTPVEHVVFNPSRDHLQKKLOFAG-WVPTKYTDKGAPVVDDEVLEGVRV DTREYVAGAPYTPVEHVVFNPSRDHLQKKLOFAG-WVPTKFTDKGAPVVDDEVLEGVRV DTREYVAGAPYTPVEHVVFNPSRDHLQKKLOFAG-WVPTKFTDKGAPVVDDEVLEGVRV DTREYVAGAPYTPVEFVTKPSRDHLQKKLOFAG-WVPTKFTDKGAPVVDDEVLEGVRV DTREYVAGAPYTPVEFVTKPSRDHLQKKLOFAG-WVPTKFTDKGAPVVDDEVLEGVRV DTREYVAGAPYTPVEFVTKPSRDHLQKKLOFAG-WVPTKFTDKGAPVVDDEVLEGVRV DTREYVAGAPYTPVEFVTKPSRDHLQKKLOFAG-WVPTKFTDKGAPVVDDEVLEGVRV LETAAYEAGEFNTSRDHLQKKLOFAG-WVPTKFTDKGAPVDDEVLEHRV -VRHYFENAPFCKLREFNTSRDHLQKLOFAG-WVPTKFTTAGAPSTDHATLEKLQ LEEVFRLGGEFNTSPKLGTTLFFKQGVFTFELGPAIGKTEKTGKRSTSAAVLEALR AAEAAYGVIGKQINIG PKLQVLFTELGPAIGKTEKTGKRSTSAAVLEALR AAEAAYGVIG
PfPREXpol eta SP6 M2 T7 T3 phiYeO3-12 Cyanophage Synechocystis T.aquaticus Myctu E.coli S.typhimurium Rhizobium Thermotoga A.thaliana Rice Dictyostelium A.aeolicus A.pyrophilus polN	107 113 96 63 143 143 143 143 143 143 159 121 158 161 161 161 159 162 201 197 185 115 115 155 155 155 155 155	KKENLHIVNDIGNKCILPICDGULNKIKILEN QKSTNEILNEINI QKENLQDVFRKYDH SQYCLALGULNGIKYJLEN QKSTNEILNEINI NDTPSIKHIGPYTPVTFEDIPLGNRDTVKQVIYDFGWRVEFHDTEQSYLDEHGVLPKPW KQGLDRVTAGSDSLKGFK ILSTKKFNKVFPKLSLPMDKEHRK DTREYVAGAPYTPVEHVVFNPSRDHUCKKLQFAG-WVPTKYTDKGAPVVDDEVLEGVRV DTRDYMEGAPYTPVEHVVFNPSRDHUCKKLQFAG-WVPTKFTDKGAPVVDDEVLEGVRV DTRDYMEGAPYTPVEFVTKPSRDHUCKKLQFAG-WVPTKFTDKGAPVVDDEVLEGVRV DTRDYMEGAPYTPVEFVTKPSRDHUCKKLQFAG-WVPTKFTDKGAPVVDDEVLEGVRV DTRDYMEGAPYTPVEFVTKPSRDHUCKKLQFAG-WVPTKFTDKGAPVVDDEVLEGVRV LEXAYEAGESTNTSPKIGTLFFKLGDAG-WVPKKTGKGAPSTDEVLEHRV -VRHYFENAPFCKLRESTNTSPKIGTLFFKLGD DRKKSRKTKTGVSTDHATLEKLQ LEEEVFRLGGKQINIGPKCIQVULFELGUPAIGKTEKTGKRSTSAAVLEALR AAEAAYGVIG
PfPREXpol eta SP6 M2 T7 T3 phiYeO3-12 Cyanophage Synechocystis T.aquaticus Myctu E.coli S.typhimurium Rhizobium Thermotoga A.thaliana Rice Dictyostelium A.aeolicus A.pyrophilus polN	107 113 96 63 143 143 143 143 143 143 143 14	KKENLHIVNDIGNKCILPICDGULNGIKKILEN QKSTNEILNEINI QKENLQDVFRKVEN SQYCLALGULNGIKKILEN QKSTNEILNEINI NDTPSIKHIGPYTPVTFEDIPLGNRDTVKQVIYDFGWRVEFHDTEQSYLDEHGVLPKPW KQGLDRVTAGSDSIKGFKILSTKKFNKVFPKLSLPMDKETRK DTREYVAGAPYTPVEHVVINPSRDHQKKLQEAG-WVPTKFTDKGAPVVDDEVLEGVRV DTREYVAGAPYTPVEHVVINPSRDHQKKLQEAG-WVPTKFTDKGAPVVDDEVLEGVRV DTRYMAGAPYTPVEHVVINPSRDHQKKLQEAG-WVPTKFTDKGAPVVDDEVLEGVRV DTRYMAGAPYTPVEHVVINPSRDHQKKLQEAG-WVPTKFTDKGAPVVDDEVLEGVRV DTRYMAGAPYTPVEHVVINPSRDHQKKLQEAG-WVPTKFTDKGAPVVDDEVLEGVRV UTRYYKGAPYTPVEHVVINPSRDHQKKLQEAG-WVPEKFTEKGAPVVDDEVLEHVRV -VRHYFENAPFCKLR-ETNPTRHHAWAFEHRGWEPKERTAGGQPKIDDEILREINT IETAAYEAGSIJSPKIGTIFFKLGDRKKSRKTKTGYSTDHATLEKLQ LEEEVFRLAGKQINIGSPKIGTIFFKLGDPKKKRKTKGYTTDADALQSLFDK LEKKAHEIGGEENISTKIQVVIDFELGQPKTKRTKGGAPSTSEEVLEELA AAEAAYGVIGEENISTKIQVVIDFELGQPKTKRTKGGYSTSAQVLEAL AAEAAYGVIG
PfPREXpol eta SP6 M2 T7 T3 phiYeO3-12 Cyanophage Synechocystis T.aquaticus Myctu E.coli S.typhimurium Rhizobium Thermotoga A.thaliana Rice Dictyostelium A.aeolicus A.pyrophilus polN MmpolN	107 113 96 63 143 143 143 143 143 143 143 14	KKENLHIVNDIENKCILPICOGILNEIKKILEN QKSTNEILNENI QKENLQDVFRKYDM SQYCIALHILNIGFSTAECESQKHIMQAKHDA NDTPSIKHIGPYTPVTFEDIPLGNRDTVKQVIYDFGWR VEFNDTEQSYLDEHGVLPKPW KQGLDRWTAGSDSIKGFKIISTKKFNKVFPKLSLPMDKEHRK DTREYVAGAPYTPVEHVVFNPSTRDHQKKLOBAG-WVPTKFTDKGAPVVDDEVLEGVRV DTREYVAGAPYTPVEHVVFNPSTRDHQKKLOBAG-WVPTKFTDKGAPVVDDEVLEGVRV DTREYVAGAPYTPVEHVVFNPSTRDHQKKLOBAG-WVPTKFTDKGAPVVDDEVLEGVRV -VRHYFENNPFCKLR-EFNPTERHHAWAFEHHRGWEPKETAGGQPKIDDEITREINT IETAAYEAGESINTSPKLIGTTLFFKLGIDRKKSRKTKTGYSTDHATLEKLQ LEEEVFRLAGESINTSPKLIGTTLFFKLGIDRKKSRKTKTGYSTDHATLEKLQ LEEEVFRLAGESINTSPKLIGTTLFFKLGIDRKKSRKTKTGYSTDHATLEKLQ LEEEVFRLAGESINTSPKLIGTTLFFKLGIDRKKSRKTKTGYSTDHATLEKLQ LEEEVFRLAG
PfPREXpol eta SP6 M2 T7 T3 phiYeO3-12 Cyanophage Synechocystis T.aquaticus Myctu E.coli S.typhimurium Rhizobium Thermotoga A.thaliana Rice Dictyostelium A.aeolicus A.pyrophilus polN MmpolN Mouse	107 113 96 63 143 143 143 143 143 143 143 14	KKENLHIVNDIGNKCILPICDGULNGIKYILEN QKSTNEILNENNI QKENLQDVFRKYMM SQYCLALMI LNIIGFSTAECESQKHIMQAKMDA NDTPSIKHIGPYTPVTFEDIPLGNRDTWKQY YDFGWR VEFNDTEQSYLDEHGVLPKPW KQGLDRWTAGSDSLKGFK ILSTKKFNKVFPKLSLPMDKENK DTREYVAGAPYTPVEHVVFNPSRDHIQKKOGAG-WVPTKYTDKGAPVVDDEVLEGVRV DTREYVAGAPYTPVEHVVFNPSRDHIQKKOGAG-WVPTKFTDKGAPVVDDEVLEGVRV DTREYVAGAPYTPVEHVVFNPSRDHIQKKOGAG-WVPTKFTDKGAPVVDDEVLEGVRV DTRDYMEGAPYTPVEFVTKPSRDHIQKKOGAG-WVPTKFTDKGAPVVDDEVLEGVRV DTREYVAGAPYTPVEFVTKPSRDHIQKKOGAG-WVPTKFTDKGAPVVDDEVLEGVRV DTRDYMEGAPYTPVEFVTKPSRDHIQKKOGAG-WVPTKFTDKGAPVVDDEVLEGVRV LETAAYEAGEFNTSRDHIQKKOGAG-WVPTKFTDKGAPVVDDEVLEHVRV -VRHYFENAPFCLREFNTSRDHIQKKOGAG-WVPTKFTDKGAPVVDDEVLEHVRV LEEVFRLIGEFNTSRDHIQKKOGAG-WVPTKFTTAGAPLALALLAL AAEAAYGVIGKQINIG PKOLGTILFFKQG PKTKKTKTGYTTDADALQSLFDK LEKKAHEINGGEFNTSTKOLQTILFFKQGKFV-LKKTPGGAPSTSEEVLEELA LEEKAHDIGG
PfPREXpol eta SP6 M2 T7 T3 phiYeO3-12 Cyanophage Synechocystis T.aquaticus Myctu E.coli S.typhimurium Rhizobium Thermotoga A.thaliana Rice Dictyostelium A.aeolicus A.pyrophilus polN MmpolN Mouse Human	107 113 96 63 143 143 143 143 143 169 121 158 161 159 162 201 197 185 115 155 155 158 158 158 158	KKENLHIVNDIGNKCILPICOGULNEIKYILEN QKSTNEILNEINI QKENLQDVFRKYMM SQYCLALMULNI IGFSTAECESQKHIMQAKHDA NDTPSIKHIGPYTPVTFEDIPLGNRDTWKQVI YDFGWR VEFHDTEQSYLDEHGVLPKPW KQGLDRWTAGSDSIKGFK ILSTKKFNKVFPKLSLPMDKEHRK DTREYVAGAPYTPVEHVVFNPSRDHHQKKLQFAG-WVPTKYTDKGAPVVDDEVLEGVRV DTRDYMEGAPYTPVEHVVFNPSRDHHQKKLQFAG-WVPTKFTDKGAPVVDDEVLEGVRV DTRDYMEGAPYTPVEFVTKPSRDHQKKLQFAG-WVPTKFTDKGAPVVDDEVLEGVRV DTRAYAGAPYTPVEHVVFNPSRDHQKKLQFAG-WVPTKFTDKGAPVVDDEVLEGVRV DTRAYAGAPYTPVEHVVFNPSRDHQKKLQFAG-WVPTKFTDKGAPVVDDEVLEGVRV DTRAYAGAPYTPVEFVTKPSRDHQKKLQFAG-WVPTKFTDKGAPVVDDEVLEHVRV -VRHYFENAPFCKLR-EFNPTSRHHAWAFEHHRGWEPKERTAGQQPKIDDEITREINT IETAAYEAGSTNISPKIGTIFFKLGDRKKSRKTKTGYSTDHATLEKLQ LEEEVFRLAG
PfPREXpol eta SP6 M2 T7 T3 phiYeO3-12 Cyanophage Synechocystis T.aquaticus Myctu E.coli S.typhimurium Rhizobium Thermotoga A.thaliana Rice Dictyostelium A.aeolicus A.pyrophilus polN MmpolN Mouse Human Fly	107 113 96 63 143 143 143 143 143 143 143 14	KKENLHIVNDIGNKCILPICOGULNGIKKILEN QKSTNEILNEINI QKENLQDVFRKUM SQYCLALMULNI IGFSTAECESQKHIMQAKHDA NDTPSIKHIGPYTPVTFEDIPLGNRDTWKQVI YDFGWR VEFHDTEQSYLDEHGVLPKPW KQGLDRWTAGSDSIKGFK ILSTKKFNKVFPKLSLPMDKENRK DTREYVAGAPYTPVEHVVFNPS RDHQKKLQFAG-WVPTKFTDKGAPVVDDEVLEGVRV DTREYVAGAPYTPVEHVVFNPS RDHQKKLQFAG-WVPTKFTDKGAPVVDDEVLEGVRV DTRYMAGAPYTPVEHVVFNPS RDHQKKLQFAG-WVPTKFTDKGAPVVDDEVLEGVRV UTRYMAGAPYTPVEHVVFNPS RDHQKKLQFAG-WVPTKFTDKGAPVVDDEVLEGVRV UTRYMAGAPYTPVEHVVFNPS RDHQKLQFAG-WVPTKFTDKGAPVVDDEVLEGVRV UTRYMAGAPYTPVEHVVFNPS RDHQKLQFAG-WVPEKFTEKGAPVVDDEVLEHVRV -VRHYFENAPFCKLR-ESNPT RHHAWAFEHHRGWEPKERTAGQQFKIDDEILREINT IETAAYEAGSSTIS PKIGT LFFKLG DRKKSRKTKTGYSTDHATLEKLQ LEEEVFRLAG
PfPREXpol eta SP6 M2 T7 T3 phiYeO3-12 Cyanophage Synechocystis T.aquaticus Myctu E.coli S.typhimurium Rhizobium Thermotoga A.thaliana Rice Dictyostelium A.aeolicus A.pyrophilus polN MmpolN Mouse Human Fly Miplp	107 113 96 63 143 143 143 143 143 143 143 14	KKENLHIVNDIENKCILPICOGLINGIKKILEN QKSTNEILINEN QKENLQDVFRKYDM SQYCIALMI LNIIGFSTAECESQKHIMQAKHDA NDTPSIKHIGPYTPVTFEDIPLGNRDTVKQVIYDFGWR VEFNDTEQSYLDEHGVLPKPW KQGLDRWTAGSDSIKGFKIISTKKFNKVFPKLSLPMDKEHRK DTREYVAGAPYTPVEHVVFNPSTRDHIQKKOGAG-WVPTKFTDKGAPVVDDEVLEGVRV DTREYVAGAPYTPVEHVVFNPSTRDHIQKKOGAG-WVPTKFTDKGAPVVDDEVLEGVRV DTREYVAGAPYTPVEHVVFNPSTRDHIQKKOGAG-WVPTKFTDKGAPVVDDEVLEGVRV -VRHYFENDFCKLR-EFNPT RHHIAWAFEHHRGWEPKETTAGQPKIDDEITREINT IETAAYEAGESINTSPKCIGTTLFFKLGIDRKKSRKTKTGYSTDHATLEKLQ LEEEVFRLGGESINTSPKCIGTTLFFKLGIDRKKSRKTKTGYSTDHATLEKLQ LEEEVFRLGGESINTSPKCIGTTLFFKLGIDRKKSRKTKTGYSTDHATLEKLQ LEEEVFRLGGEFNITSFKCIGTTLFFKLGIDRKKSRKTKTGYSTDHATLEKLQ LEEEVFRLGGEFNITSFKCIGTTLFFKQGKP-LKKTPGAPSTSEEVLEELA LEKKAHEINGEFNITSFKCIGTTLFFKQGKP-LKKTPGAPSTSEEVLEELA LEKKAHDIG
PfPREXpol eta SP6 M2 T7 T3 phiYeO3-12 Cyanophage Synechocystis T.aquaticus Myctu E.coli S.typhimurium Rhizobium Thermotoga A.thaliana Rice Dictyostelium A.aeolicus A.pyrophilus polN MmpolN Mouse Human Fly Miplp PfPREXpol	107 113 96 63 143 143 143 143 143 143 143 14	KKENLHIVNDIGNKCILPICOGULNKIKCILEN QKSTNEILNENNI QKENLQDVFRKYMM SQYCLALMULNI GFSTAECESQKHIMQAKMDA NDTPSIKHIGPYTPVTFEDIPLGNRDTWKQY YDFGWR VEFNDTEQSYLDEHGVLPKPW KQGLDRWTAGSDSLKGFK ILSTKKFNKVFPKLSLPMDKENK DTREYVAGAPYTPVEHVVFNPSRDHIQKKOGAG-WVPTKYTDKGAPVVDDEVLEGVRV DTREYVAGAPYTPVEHVVFNPSRDHIQKKOGAG-WVPTKFTDKGAPVVDDEVLEGVRV DTREYVAGAPYTPVEHVVFNPSRDHIQKKOGAG-WVPTKFTDKGAPVVDDEVLEGVRV DTREYNAGAPYTPVEFVTKPSRDHIQKKOGAG-WVPTKFTDKGAPVVDDEVLEGVRV DTREYNAGAPYTPVEFVTKPSRDHIQKKOGAG-WVPTKFTDKGAPVVDDEVLEGVRV DTREYNAGAPYTPVEFVTKPSRDHIQKKOGAG-WVPTKFTDKGAPVVDDEVLEGVRV DTREYNAGAPYTPVEFVTKPSRDHIQKKOGAG-WVPTKFTDKGAPVVDDEVLEGVRV DTREYNAGAPYTPVEFVTKPSRDHIQKKOGAG-WVPTKFTDKGAPVVDDEVLEGVRV DTREYNAGAPYTPVEFVTKPSRDHIQKKOGAG-WVPTKFTDKGAPVVDDEVLEGVRV DTREYNAGAPYTPVEFVTKPSRDHIQKKOGAG-WVPTKFTDKGAPVVDDEVLEGVRV DTREYNAGAPYTPVEFVTKPSRDHIQKKOGAG-WVPTKFTDKGAPVVDDEVLEGVRV DTREYNAGAPYTPVEFVTKPSRDHIQKKOGAG-WVPTKFTDKGAPVVDEVLEGVRV DTREYNAGAPYTPVEFVTKPSRDHIGKT IETAAYEAGEFNTSPKLIGTLEFKGGTPVEKTRTGGVSTSAAVLEALR AAEAAYGVIGKQINIG PKIGVVLFTELGVFKKGTKTFGGNSTSAAVLEALR AAEAAYGVIG
PfPREXpol eta SP6 M2 T7 T3 phiYeO3-12 Cyanophage Synechocystis T.aquaticus Myctu E.coli S.typhimurium Rhizobium Thermotoga A.thaliana Rice Dictyostelium A.aeolicus A.pyrophilus polN MmpolN Mouse Human Fly Mip1p PfPREXpol eta	107 113 96 63 143 143 143 143 143 143 143 14	KKENLHIVNDIGNKCILPICOGULNEIKYILEN QXSTNEILNEINI QKENLQDVFRKYMM SQYCLALMU LNEIGFSTAECESQKHIMQAKHDA NDTPSIKHIGPYTPVTFEDIPLGNRDTWKQVI YDFGWR VEFHDTEQSYLDEHGVLPKPW KQGLDRWTAGSDSIKGFK ILSTKKFNKVFPKLSLPMDKEHRK DTREYVAGAPYTPVEHVVFNPSRDHIQKKLQFAG-WVPTKYTDKGAPVVDDEVLEGVRV DTREYVAGAPYTPVEHVVFNPSRDHIQKKLQFAG-WVPTKYTDKGAPVVDDEVLEGVRV DTREYVAGAPYTPVEHVVFNPSRDHIQKKLQFAG-WVPTKFTDKGAPVVDDEVLEGVRV DTRYMEGAPYTPVEFVTKPSRDHIQKKLQFAG-WVPTKFTDKGAPVVDDEVLEGVRV DTRYMEGAPYTPVEFVTKPSRDHIQKKLQFAG-WVPTKFTDKGAPVVDDEVLEHVRV -VRHYFENAPFCKLR-EFNPT RHHAWAFEHHRGWEPKERTAGGQPKIDDEITREINT IETAAYEAGESTNIS PKOIGTIFFKLGIDRKKSRKTKTGYSTDHATLEKLQ LEEEVFRLGGHPFNIN RDDIERVLFDELGPAIGKTEKTGKRSTSAAVLEALR AAEAAYGVIGKQINIG PKOIQVUFFELGG PKTKRTKTGYTTDADALQSLFDK LEKKAHEINGESTNIS TKOIQTIFFKQGKF-LKKTPGGAPSTSEEVLEELA LEDEIYVLGG
PFPREXpol eta SP6 M2 T7 T3 phiYeO3-12 Cyanophage Synechocystis T.aquaticus Myctu E.coli S.typhimurium Rhizobium Thermotoga A.thaliana Rice Dictyostelium A.aeolicus A.pyrophilus polN MmpolN Mouse Human Fly Miplp PfPREXpol eta cpc	107 113 96 63 143 143 143 143 143 143 143 14	KKENLHIVNDIGNKCILPICDGULNEHKULEN QKSTNEILNEN QKENLQDVFRKYDM SQYCIALD IN UGFSTAECESQKHIMQAKDA NDTPSIKHIGPYTPVTFEDIPLGNRDT KQVI YDFGWR VEFNDTEQSYLDEHGVLPKPW KQGLDRWTAGSDSLKGFK ILSTKKFNKVFPKLSLPMDKERK DTREYVAGAPYTPVEHVVFNPSRDH QKKLOAG-WVPTKFTDKGAPVVDDEVLEGVRV DTREYVAGAPYTPVEHVVFNPSRDH QKKLOAG-WVPTKFTDKGAPVVDDEVLEGVRV -VRHYFENAPFCKLREFNPT RHH AWAFEHHRGWEPKETAGGQPKIDDEILREINT IETAAYEAGESNLSPKLGTLFFKLG DRKKSRKTKTGYSTDHATLEKLQ LEEEVFRLGGESNLSPKLGTLFFKLG DRKKSRKTKTGYSTDHATLEKLQ LEEEVFRLGG
PfPREXpol eta SP6 M2 T7 T3 phiYeO3-12 Cyanophage Synechocystis T.aquaticus Myctu E.coli S.typhimurium Rhizobium Thermotoga A.thaliana Rice Dictyostelium A.aeolicus A.pyrophilus polN MmpolN Mouse Human Fly Miplp PfPREXpol eta SP6	107 113 96 63 143 143 143 143 143 143 143 14	KKENLQIVFRKYMM SQYCIAL BLNYKY LENGOKSTKEI LNEINI QKENLQDVFRKYMM SQYCIAL BLNYKY LENGOKSTKEI LNEINI NDTPSIKHIGPYTPVTFEDI PLGNRDT KQVI YDFGWR WEFNDTEQSYLDEHGVLPKPW KQGLDRWTAGSDSIKGFK ILSTKKFNKVFPKLSLPMDKEIRK DTREYVAG PYTPVEHVVN PSSRDH QKKIOBAG-WVPTKYTDKGAPVVDDEVLEGVRV DTREYVAG PYTPVEHVVN PSSRDH QKKIOBAG-WVPTKYTDKGAPVVDDEVLEGVRV UTREYVAG PYTPVEHVVN PSSRDH QKKIOBAG-WVPTKYTDKGAPVVDDEVLEGVRV -VRHYFEM PFCKLRENPTSRHH AWAFEHHRGWEPKERTAGGQPKIDDEI REINT IETAAYEA GESINTSPKIGTILFKLGIDRKKSRKTKTGYSTDHATLEKLQ LEEEVFRL GESINTSPKIGTILFKLGIDRKKSRKTKTGYSTDHATLEKLQ LEEEVFRL GESINTSPKIGTILFKLGIDRKKSRKTKTGYSTDHALEKLQ LEEEVFRL GESINTSPKIGTILFKLGIDRKKSRKTKTGYSTDHALEKLQ LEEEVFRL GESINTSPKIGTILFKLGIDRKKSRKTKTGYSTDHALEKLQ LEEVFRL GESINTSPKIGTILFKLGIDRKKSRKTKTGYSTDHALEKLQ LEEVFRL G
PFPREXpol eta SP6 M2 T7 T3 phiYeO3-12 Cyanophage Synechocystis T.aquaticus Myctu E.coli S.typhimurium Rhizobium Thermotoga A.thaliana Rice Dictyostelium A.aeolicus A.pyrophilus polN MmpolN Mouse Human Fly Miplp PfPREXpol eta SP6 M2	107 113 96 63 143 143 143 143 143 143 143 14	KKENLQUVFRKVMM SQYCIAL HE LNIK MENDOKSTKEI LNEINI QKENLQDVFRKVMM SQYCIAL HE LNIK GFSTAECESQKHIMQAKUDA NDTPSIKHIGPYTPVTFEDI PLGNRDTVKQM YDFGWR VEFDDTEQSYLDEHGVLPKPW KQGLDRWTAGSDSIKGFKIILSTKKFNKVFPKLSLPMDKEHRK DTREYVAG PYTPVEHVVM PSSRDH QKKLOBAG-WVPTKFTDKGAPVVDDEVLEGVRV DTREYVAG PYTPVEHVVM PSSRDH QKKLOBAG-WVPTKFTDKGAPVVDDEVLEGVRV UTRDYMEG PYTPVEHVVM PSSRDH QKKLOBAG-WVPEKFTEKGAPVVDDEVLEGVRV -VRHYFEN PFCKLRENPTSRHH AWAFEHHRGWEPKERTAGGQPKIDDEI FREINT IETAAYEA GESTUSSPK IGTIF KLGIDRKKSRKTKTGYSTDHATLEKLQ LEEEVFRLGHPENINSRD LERVIFFELGIPAIGKTEKKGKSTSAAMLEALR AAEAAYGVIGKQIIGSPKJIQVIFFELGIPAIGKTEKKTGYTTDADALQSLFDK LEKKAHEII GERENISSTKJIQTIHEKGGIKP-LKKTPGGAPSTSEEVLEELA LEKKAHDI GERENISSTKJIQTIHEKGGIKP-LKKTPGGAPSTSEEVLEELA AGSRFRNW SKYCPDAKYMUGSDTIRQIFGGISN-SSHDEVLPVEKLFKVPNIDKVI AADKFRKW SKHCPDAKYMUNDDN IRQIFFGGISN-SSHDEVLPVEKLFKVPNIDEGIA NRKVFLNW KTHSPDAENM PVDDA IQIFFGGISN-SSHDEVLPVEKLFKVPNIDEGIA NRKVFLNW KTHSPDAENM PVDDA IQIFFGGISN-SSHDEVLPVEKLFKVPNIDEGIA NRKVFLNW KTHSPDAENM PVDDA IQIFFGGISN-SSHDEVLPVEKLFKVPNDEGIA NRKVFLNW KTHSPDAENM PVDDA IQIFFARVPON-KKTKESLPLEKDFCDNLGT IGQENJGGIJTSSA VVTRHPAPDEEGYGAILPOVTAGTT TRRAVS PTW TASNARPDRVGSELKAMVQAPG SKSLNNEKTNDLAIIIFKVPMGTT TRRAVS PTW TASNARPDRVGSELKAMVQAPG SKKSLNNEKTNDLAIIIFKVPMGTT TRRAVS PTW TASNARPDRVGSELKAMVQAPG SKKSLNNEKTNDLAIIFKVPMGTT TRRAVS PTW TASNARAPDRVGSELKAMVQAPG SKDNLKKLKLDENINN QQ KLKALQKNNVRDISNKLIENTSDSNLKNFLN IETQAYQU
PfPREXpol eta SP6 M2 T7 T3 phiYeO3-12 Cyanophage Synechocystis T.aquaticus Myctu E.coli S.typhimurium Rhizobium Thermotoga A.thaliana Rice Dictyostelium A.aeolicus A.pyrophilus polN MmpolN Mouse Human Fly Mip1p PfPREXpol eta SP6 M2	107 113 96 63 143 143 143 143 143 143 143 14	KKENLQUVFRKVMI SQYCIAL SUNG KKLEN QKITNEILNEINI QKENLQUVFRKVMI SQYCIAL SUNG KKLEN QKINAKAN NDTPSIKHIGPYTPVTFEDIPLGNRDYKQY YDFGWR WEFDTESSYLEHGVLFKPW KQGLDRWTAGSDSLKGFK ILSTKKFNKVFPKLSLPMDKERK DTREYVAG PYTPVEHVVFIPSSRDH OKKICAGG-WVPTKFTDKGAPVVDDEVLEGVRV DTREYVAG PYTPVEHVVFIPSSRDH OKKICAGG-WVPTKFTDKGAPVVDDEVLEGVRV UTRDYMEG PYTPVEHVVFIPSSRDH OKKICAGG-WVPTKFTDKGAPVVDDEVLEGVRV VRHYFEN PFCKLR-EFIPTSRHH AWAFEHHRGWEPKERTAGGORKIDDEILREINT IETAAYEAGGBSTUSSRK JGTLDFKLGIPAKKSRKTKTGYSTDHATLEKLO LEEVFRIGGBSTUSSRK JGTLDFKLGIPAKKSRKTKTGYSTDAALLEALR AAEAAYGVIG
PFPREXpol eta SP6 M2 T7 T3 phiYeO3-12 Cyanophage Synechocystis T.aquaticus Myctu E.coli S.typhimurium Rhizobium Thermotoga A.thaliana Rice Dictyostelium A.aeolicus A.pyrophilus polN MmpolN Mouse Human Fly Miplp PfPREXpol eta SP6 M2	107 113 96 63 143 143 143 143 143 143 143 14	KKENLQUVFRKVHM SQYCHALMSIKELEN QKSTNEILNEINI QKENLQUVFRKVHM SQYCHALMSIKELEN QKSTNEILNEINI QKEN
PfPREXpol eta SP6 M2 T7 T3 phiYeO3-12 Cyanophage Synechocystis T.aquaticus Myctu E.coli S.typhimurium Rhizobium Thermotoga A.thaliana Rice Dictyostelium A.aeolicus A.pyrophilus polN MmpolN Mouse Human Fly Miplp PfPREXpol eta SP6 M2 T7	107 113 96 63 143 143 143 143 143 143 143 14	KKENLQUVFRKYMM SQYCHALLNUKTLEN QKSTNEILNEINI QKENLQUVFRKYMM SQYCHALLNUKTLEN QKSTNEINEN NDTPSIKHIGPYTPVTFEDIPLGNRDYKQY YDFGWR WEFDTESSYLEHGYLPKPW KQGLDRYTAGSDSIKGFK ILSTKKFNKVFPKLSLPMDKERK DTREYVAGD PYTPVEHVYT PSSRDH QKKICGAG-WVPTKYTDKGAPVVDDEVIEGVRY DTREYVAGD PYTPVEHVYT PSSRDH QKKICGAG-WVPTKFTDKGAPVVDDEVIEGVRY DTRYMAGD PYTPVEHVYT PSSRDH QKKICGAG-WVPTKFTDKGAPVVDDEVIEGVRY DTRYMAGD PYTPVEFVT KPSSRDH QKKICGAG-WVPTKFTDKGAPVVDDEVIEGVRY DTRYMAGD PYTPVEFVT KPSSRDH QKKICGAG-WVPTKFTDKGAPVVDDEVIEGVRY VRHYFEND PFCKLR-ETIPT SRHH AWAFEHHRGWEPKERTAGGQFKIDDEI FREINT IETAAYEAGGESTURS PKJIGT IF KLGIDRKKSKKTKGYSTDHATLEKLQ LEEVFRLGG
PFPREXpol eta SP6 M2 T7 T3 phiYeO3-12 Cyanophage Synechocystis T.aquaticus Myctu E.coli S.typhimurium Rhizobium Thermotoga A.thaliana Rice Dictyostelium A.aeolicus A.pyrophilus polN MmpolN Mouse Human Fly Miplp PfPREXpol eta SP6 M2 T7 T3	107 113 96 63 143 143 143 143 143 143 143 14	KKENLQUVFRKEN SQYC ALLE IN I GFTAECESQKHIMQAKIDA QKENLQUVFRKEN SQYC ALLE IN I GFTAECESQKHIMQAKIDA NDTPSIKHIGPYTPVTEDIPLGNRDI KQV YDFGWR WEF OTEQSYLDEHGVLPKFW KQGLDR TAGSDSI KGFK I LSTKKFNKVFPKLSLPMDKERK DTREYVAG PYTPVEHVVEN PS RDH QKKIO AG-WVPTKFTDKGAPVVDDEV EGVRV DTREYVAG PYTPVEHVVEN PS RDH QKKIO AG-WVPTKFTDKGAPVVDDEV EGVRV DTREYVAG PYTPVEHVVEN PS RDH QKKIO AG-WVPTKFTDKGAPVVDDEV EGVRV DTREYVAG PYTPVEHVVEN PS RDH QKKIO AG-WVPTKFTDKGAPVVDDEV EGVRV VRHYFENAPFCKLRESPIT RHH AWAFEHHRGWEPKETAGGQPKIDDEI REINT IETAAVEAGESPITS PK I TI TI KLG DRKKSRTKTGYSTDHATLEKLQ LEEEVFRLG
PFPREXpol eta SP6 M2 T7 T3 phiYeO3-12 Cyanophage Synechocystis T.aquaticus Myctu E.coli S.typhimurium Rhizobium Thermotoga A.thaliana Rice Dictyostelium A.aeolicus A.pyrophilus polN Mouse Human Fly Miplp PfPREXpol eta SP6 M2 T7 T3 phiYeO3-12	107 113 96 63 143 143 143 143 143 143 159 121 158 161 159 162 201 197 185 115 155 155 158 227 230 194 165 166 202 202 202 202 202 202 202 2	KKENLQUVFRKUM SQYC ALLELN I GFTAECESQKHIMQAKLDA QKENLQUVFRKUM SQYC ALLELN I GFTAECESQKHIMQAKLDA NDTPSIKHIGPYTPVTEDIPLGNRDI KQV YDFGWR WEFGDEQSYLDEHGVLPKPW KQGLDR TAGSDSI KGFK I LSTKKFNKVFPKLSLPMDKERK DTREYVAG PYTPVEHVVN PS RDH QKKLOBAG-WVPTKFTDKGAPVVDDEV EGVRV DTRDYMEG PYTPVEHVVN PS RDH QKKLOBAG-WVPTKFTDKGAPVVDDEV EGVRV DTRDYMEG PYTPVEHVVN PS RDH QKKLOBAG-WVPTKFTDKGAPVVDDEV EGVRV VRHYFENAPFCKLREDIPT RHH AWAFEHRGWEPKETAGGOPVIDDEV EGVRV UTRDYMEG PYTPVEFVT KPS RDH QKKLOBAG-WVPTKFTDKGAPVVDDEV EHVRV -VRHYFENAPFCKLREDIPT RHH AWAFEHRGWEPKETAGGOPSIDDEI ELEINT IETAAVEAGGESNIS PKOLGT LEG VALGEVERTAGGOPSTAELELA LEEEVFRLGROUTING PKOLOVULTIELG PAIGKTEKTGKSTSAALEALR AAEAAYGVIGKQINIG PKOLOVULTIELG PKTKRTKTGYTDADALOSLFDK LEKKAHEI AGEESNIS TKOLOTILERKQGI KP-LKKTPGGAPSTSEE LEELA LEELYNLGG
PFPREXpol eta SP6 M2 T7 T3 phiYeO3-12 Cyanophage Synechocystis T.aquaticus Myctu E.coli S.typhimurium Rhizobium Thermotoga A.thaliana Rice Dictyostelium A.aeolicus A.pyrophilus polN MmpolN Mouse Human Fly Miplp PfPREXpol eta SP6 M2 T7 T3 phiYeO3-12 Cyanophage	107 113 96 63 143 143 143 143 143 143 143 14	KKENLQUVFRKEN SQYC ALLE NI I GFTAECESQKHIMQAKIDA NDTPSIKHIGPYTPVTEDIPLGNRDTUKQV YDFGWR WEFGDEQSYLDEHGVLPKPW KQGLQR TAGSDSIKGFK ILSTKKFNKVFPKLSLPMDKERK DTREYVAG PYTPVEHVVEN PS RDH QKKIOFAG-WVPTKFTDKGAPVVDDEV EGVRV DTREYVAG PYTPVEHVVEN PS RDH QKKIOFAG-WVPTKFTDKGAPVVDDEV EGVRV DTREYNAG PYTPVEHVVEN PS RDH QKKIOFAG-WVPTKFTGKAFTGGQPKIDDEI EGVRV UTREYNAG PYTPVEHVNEN PS RDH QKKIOFAG-WVPTKFTGKATTGGAPVDDEV EGVRV LEKKAHD DE SUBSPKICTT LITEKGG PAIGKTKTGGYTDADALOSLFDK LEKKAHD GKQINIG PKO QVVUT ELG PAIGKTEKTGGYSTSEE LEELA LEDEIVVLG
PfPREXpol eta SP6 M2 T7 T3 phiYeO3-12 Cyanophage Synechocystis T.aquaticus Myctu E.coli S.typhimurium Rhizobium Thermotoga A.thaliana Rice Dictyostelium A.aeolicus A.pyrophilus polN MmpolN Mouse Human Fly Miplp PfPREXpol eta SP6 M2 T7 T3 phiYeO3-12 Cyanophage Sumachage Sumachage Sumachage Sumachage Sumachage Sumachage Sumachage	107 113 96 63 143 143 143 143 143 143 143 14	KKENLQDVFRKUEM SQYC LALLE IN I GFSTAECESOKHIMOAKIDA NDTPSIKHIGPYTPVTFEDIPLGNRDTUKQV YDFGWR VEFBOTEQSYLDEHGVLPKKUA KQGLQRVTAGSDSIKGFK ILSTKKFNKVFPKLSLPMDKERK DTREYVAG PYTPVEHVVEN PSSRDH QKKIOBAG-WVPTKYTDKGAPVVDDEV EGVRV DTREYVAG PYTPVEHVVEN PSSRDH QKKIOBAG-WVPTKYTDKGAPVVDDEV EGVRV DTREYVAG PYTPVEHVVEN PSSRDH QKKIOBAG-WVPTKFTDKGAPVVDDEV EGVRV DTRUMEG PYTPVEHVVEN PSSRDH QKKIOBAG-WVPTKFTDKGAPVVDDEV EGVRV DTRUMEG PYTPVEHVVEN PSSRDH QKKIOBAG-WVPTKFTDKGAPVVDDEV EGVRV DTRUMEG PYTPVEHVNEN SSRDH QKKIOBAG-WVPTKFTDKGAPVVDDEV EHVRV -VRHYEEN PTCKLRENPTRHH AWAFEHHRGWEPKERTAGQOPKIDDEI REINT IETAAYEAGGESNHSSPKIGTILFEKLGIDRKKSRKTKTGYSTDHATLEKLQ LEEEVFRLAGESNHSSPKIGTILFEKLGIDRKKSRKTKTGYSTDHATLEKLQ LEEKAHDIGGESNHSSPKIGTILFEKLGIDRKKSRKTKTGYTDADALOSLFDK LEKKAHEIGGESNHSSPKIGTILFEKLGIDRKKSRKTKTGYTDADALOSLFDK LEKKAHDIAGESNHSSPKIGTILFEKLGIDRKKSRKTKTGYSTDADALOSLFDK LEKKAHDIAGESNHSSPKIGTILFEKLGIDRKKRKTKTGYSTDADALOSLFDK LASEIYRIAGESNHSSPKIGTILFEKLGIKG SG-GSKTKTGGWSTSAOLEDLA AGSRFRNWSKYCPDAKYMUNG DTOIROLF GGISN-SSHDEVLPVEKLFKVPNIDKVI AADKFKKWSKHCPDAKYMUNG DTOIROLF GGISN-SSHDEVLPVEKLFKVPNIDKVI AADKFKKWSKHCPDAKYMUNG DTOIROLF GGISN-SSHDEVLPVEKLFKVPNIDGGIA NRKVFLNWKTHSPDAENMIPV DNU ROLF GGISN-SSHDEVLPVEKLFKVPNDEGIA NRKVFLNWKTHSPDAENMIPV DNU ROLF GGISN-SSHDEVLPVEKLFKVPNDEGIA NRKVFLNWKTHSPDAENMIPV DNU ROLF GGISN-SSHDEVLPVEKLFKVPNDEGIA NRKVFLNWKTHSPDAENMIPV DAGT TRRAV PTWITASNRFDRVGSELKAM QAPFG AVIRHPDYDEGLYGAILPOVVIGT TRRAV PTWITASNRFDRVGSELKAM QAPFG SAKSSLNNEKTNDLAILI FK VPNGT TRRAV PTWITASNRAPDRVGSELKAM QAPFG SAKSSLNNEKTNDLAILI FK VPNGT TRRAV PTWITASNRAPDRVGSELKAM QAPFG SAK
Musta	210 mg	
---------------	---------	--
Myctu	212 TG-	
E.COII	214 LD-	
S.typhimurium	214 LD-	
Rhizobium	212 AAG	G
Thermotoga	216 GEH	H
A.thaliana	260 EE	GKKTPTKFRNIKLHRISDSPLSTENFTASGWPSVGGDVLKELAGKVSAEYDFMDDVSD
Rice	257 TEC	GKKTP-KSRTIKLFTIVED-LKIDMFTPTGWPSVSGDVLRSLAGKIPTDHIYKIDD
Dictyostelium	244 EV(GKTKPKKKKTFYLRGIGMESKSLTTNGWPSVDSSSLRELAGKNPIDGKYGSAY
A.aeolicus	163	
A.pyrophilus	163	
polN	212 VL	N
MmpolN	212 ML	N
Mouse	287 YVI	LVGADVD
Human	290 YT	LVGADVD
Fly	252 YRJ	LVGADVD
Mip1p	205 YCI	FVGADVD
PfPREXpol	206	
eta	215 NG	RKLRLGRQFSTSKDVLN
SP6	216 QAG	GVRKCRG
M2	158 KDI	E
Т7	209	AID IKEYL KOK
Т3	209	AID IKEYL IOK
phiYeO3-12	209	CID IKEYLNIOK
Cyanophage	146	LAFARILELOK
Synechocystis	215	HPI DAN LEHRT
T.aquaticus	177	HPT DRULOYRE
Myctu	214	HPFIOHLAHRD
E.coli	216	YPLPKVWLEYRC
S.typhimurium	216	YPLPKV LEYRG
Rhizobium	215	FELPRK VDWROW
Thermotoga	219	EIIP-LHLEYRK
A.thaliana	320 IS1	LEEVVEDDDVETSETQKSKTDDETDTSAYGTAYVAFGGGERGKEACHALASICEVCS
Rice	313	GQEFDEDGSSLELPEQDIEDTSPYGTAYEAFGGGKKGREACHAHAALCEVFSI
Dictyostelium	299	EQMGREASIAINALLELGS
A.aeolicus	163	YQDVEPVKLVLE
A.pyrophilus	163	YAHVEPWRLLE
polN	215	ALRDLHPLPKI <mark>I</mark> LEYRQV
MmpolN	215	SLQDLHPLPKL
Mouse	296	SQELWIAAVIGDA
Human	299	SQEIWIAAVIGDA
Fly	261	SQEIWIASVIGDA
Mipip	214	SEE WIASLWGDS
PIPREAPOI	200	HEE ISTRNYRR
eta	234	KLKALHPLPGL
SP0	161	LUPVAFNKELG
112	101	QYPLYMQRWRFE
т7	222 RIC	GORAEGDKAWLRYVAEDGRHIGSVNPNGAVIERAWHAF
тЗ	222 RTC	GORAEGDKAWLRYVAEDGKHIGSVNPNGAVHORAWHAR
phiYeO3-12	222 RIC	GOAAEGDKAWLRYVODDGHHGAVNPNGAV HALHSF
Cyanophage	157 HLC	GOLSEGKNAWLKLERK-GBHHHSCVLN-INN GROCHMP
Synechocystis	228 A	KSTYVDA PELVNGOTOKHADEN AVES CHESSING
T.aquaticus	190 T	KNTHIDPEPALVHPKTGE
Myctu	227 TR	KVTVDGL OAVAADGBHHTTONTIAANHE STEL
E.coli	229 AK	KSTTTTDK PLMINPKTGWHISTHEAV AVER STD
S.typhimurium	229 A	KOTWTOK PLMINPKTGBWHTSWHEAVHAUGHUSSTDEWHONT
Rhizobium	228 T	KSTTTDA PGYVHPETKEWHESTSLASET COMPASSE
Thermotoga	231 0	KST IDA PKMVNPKTGR HASEN TG ANGLISSD
A.thaliana	380 DS	ISNEILPHOGSNVSGKDGRVHCSLWINNEWGRUMARR
Rice	366 DK	ISGUIVE OGDRISCKEGRIHCSLNINGE GELWARTNICONO
Dictyostelium	338 GT	LNTHIP OKLADSNSRLHTSINVNHENGLISSKK
A.aeolicus	176 R	KKIADKLKELKEHLKNGWYPERKOIGAVIGHN SAHENIONT
A.pyrophilus	176 R.	KKLSDKFKEIKENLKGDEEYPEEKOIGAVICENSLKENYONK
polN	233 H	KSTEVDG LACMKKGSUSSTENNTGEVIGELSAKHENECUSKHET
MmpolN	233 H	KSTOIDG LAYMKKGSUSSTONOTG VILLE AKH HECUSKHPT
Mouse	309 HLA	AGMHGCTAFGWMTLQGRKSRGUDLHSKTAA VE HREHAKIF-N
Human	312 HFF	AGMHGCTAFGWMTLQGRKSRGUDLHSKTATUVGLEREHAKUF-N
Fly	274 YAC	CGEHGATP GWMTLSGSKSNGSDMHSITAKAVG RDHAKWI-N
Miplp	227 IFM	N-VHGGTALGWMCLEGTKNEGUDLHTKTAQILGC RNEAKUF-N
PfPREXpol	219 YKI	YEARYLK PLHINTKTNKHUTENELK FEGEFESEK NLOOIP
eta	252 TNA	AITKVVFPTQREKCLNPFLGMERIYPVSCSHTATGRITFTETNIONVPRDFE
SP6	237 NAC	QAYYETYGYWPTSDKDDGE-WRWPAVAISIG SFFMRHRNWVW
M2	173 FE	KEGMIPT QIKKNPFFKGNEMLKNSGVEPVELYLTN

T /	
Т3	270VRSPIGEQCRAAFGAEMED
phiYeO3-12	270VRSPIGEQCKSAFGAE
Cyanophage	202SAHEYRSLFKPSD
Synechocystis	274PIRSDFSRQIMMAF
T.aquaticus	236PVRTPLGQRINNAF
Myctu	271PIRTDAGRRIEDMF
E.coli	275PVRNEEGRRIPOFF
S.typhimurium	275PVRNEEGRRINO F
Rhizobium	274PVRTAEGRKITTF
Thermotoga	277PTKSEEGKEIREAI
A.thaliana	426PALEKDRYKIPKAF
Rice	412PVLEKDRYKIRH#F
Dictvostelium	382PALEKDRYKIRNAF
A apolicus	
A. acorreus	
A. pyrophilus	
POIN	
MIDOTN	
Mouse	354IGRIIGAGOSTABILL
Human	357YGRIYGAGQPFAEDLL
Fly	319YARIYGAGQLFAETLL
Miplp	271YGRIYGAGAKFASQLL
PfPREXpol	266RQKNIREI
eta	306 IKMPTLVGESPPSQAVGKGLLPMGRGKYKKGFSVNPRCQAQMEERAADRGMPFSISMEHA
SP6	282IPARGLYPLEDL
M2	212VDLE
Τ7	290 GITGKPWVOAGI AGURING CHANFMARFDNGEYAHDTLNG HUKNOTAAELPTRENAK
m 3	290 CITCKPWVOACT ASCHURCCALEMARFDNGEVAHRTINGALEMKNOMAAFLPTRINAK
nhive03-12	200 GITCKPWWQAGIDA GODA CAMPANAK DAGINATI NG TUKKANA AFI DTA NAK
phile03-12	250 GIIGREWUGAGI A GIANG AN MAREDVOCUEA FINICIANG THANQNAAEDER NAA
Cyanophage	213NALQAGG AGENCIC CHILLSRIDGGRAPE VIOLOUT ALAE IGIDRASGA
Synechocystis	
T.aquaticus	250 VA-EEGWVLWVLIYSGTBHIVIATINGGEENHIRVIQECRDTHIQHASWWFGVSPEGVD
Myctu	285 VVGDGYAELJTAWHON BOT BOT BOT SGREGETEARNTCEDT SFVMSRAFGVPI DEWT
E.coli	289 IAP-EDYVIVSANYHOUSINE CAMPEREKGATTALABCKDIMRAMALEWFGIPLETET
S.typhimurium	289 IAP-EDYLINSADYGDIEBEIMANISRIKGHITAFABCKDIHRATAAENFGUPLISUT
Rhizobium	288 IST-PGHKLISADYSTIDDEVEAUVAEIPOTKAPEDOVDIHAMPASEWFGUPVEGDP
Thermotoga	291 VPQDPNWWIVSADYSCIELELLAHLSGEENLIRAFEEGIDWHTLEASRIFNWKPEEVT
A.thaliana	440 VAS-PGNTLWVALYC, LEDEL AHLTGCKSWEARKACG FISRTAMNMYPHVRAME
Rice	426 VAA-PGNTLEVALVG PARTELEANDANCKSTEARKAGGFERRAMNOYOHVRAAKE
Dictvostelium	396 TCE-PGNTLUVADYG AND THAL TINCKS TABOYGG- FESTAMGY PHYKAND
A applique	229 FRAFFONTERT STATES THE PLAFFY WE PLAT DARK WE - DELEVER SUBLICK FORT
A. acorrections	
A. pyrophilus	225 FRAPEGNITVIALFOOT FOR A BIVNED FRANKSK - ULKI SLOUGREEFI
POIN	305 FVSSRGHTFLAATISOTELETITILSGEPETELEDESERDIVFSTETSOWKDPVCOVT
MmpolN	305 FVSSEGHTFLAANSKULDING MALINSGEPEN KLEOMSERDINFSTLTSOWKDHPIERUT
Mouse	370 MQFNHRLTRQEMAEKAQOQYAMTHGLRRYR SADGEWLVKQLNLPVDRTEDGWUSLQDIR
Human	373 MQFNHRLTQQEMAEKAQWAYAATHGLRWYR SDEGEWLVRELNLPVDRTEGGWISLQDMR
Fly	335 RQFNPTFSASE KAKAQKUFSHTWGKRVYRUR-EEFHDELMDRAYSSYEASRUAIQRNR
Miplp	287 KRFNPSLTDEETKKIANKYENTKGKTKRSK
PfPREXpol	274 FIPNDNNIFTIATEK, IEJK ARETTNELITIKATNNNIDITTLERSITTKKNIPDEN
eta	366 FVPFPGGSILAADYSTOLELKILAALSHORRE QVLNTA VFRSIAAEWKMEPPSVG
SP6	294 FIAGKGKMINGCHGAGUDIRVISHEMMIPEYOEIVLHEDIHMHNOLKAGLPKRMAK
M2	216 LIOEHYELYNVEYIDGFKFEEKTGFFKEFIDKWTWYKTHEEGABKOLBKLELNSLYGKFA
Τ7	350 TFT TO FOR ACDER
T3	350 TFT FT ACDER OT CACKE-
15 nhiVe02-12	
phireos-12	350 I
Cyanophage	269 G
Synechocystis	343 PAEENLETTINDEVITUMGAQREERETGISA
T.aquaticus	307 PLMBRAAPTINEOVLYCMSAHRISGEIS PY
Myctu	343 GBLERRVANSYGLANGLSAYG SQUAK ST
E.coli	346 SWORSAHAINEGIN YGMSAFGDARQUN PR
S.typhimurium	346 RBORRSARGISROIN PR
Rhizobium	345 GSVERRAANSSAVGISAFGANOISER
Thermotoga	349 E
A.thaliana	497 NGOVILEWHPEPGEDKPPVPLLKDAFGSEBRKAMMINESTAWCKTAVCUSRDWKWST
Rice	483 EKKVLLEWHPOPGODKPPVPLLKDAFGAFIPKIMANISSIAWSKTAVCKGWDWKWGV
Dictucctolium	153 DOEWLI EWDCECEDDEDI I KNUVACE DELEMANDOLA WAAUC CODWOWN
Diccyoscerium Discosli	455 ROEVILLEWDGEGEFFRELIKINVIASERKKAMITINNISTARUKTANGI SQDWGVTL
A. aeolicus	
A nurophi lus	287 KBERQLATARNIG FROM ISAKGI ZEYAKI GYGVE
A. Pyrophrido	287 KBERQLA AINGHITTISAKGAEYAKIGYGVE 287 KBERQLAAINEGHIVGISAKGAEYARTGYGVE
polN	287 KBERQLA AINEGHTIGISAKGAEYAKIGIGVE 287 KBERQLAAINEGLIYGISAKGAEYARIGIGVE 365 HADREQTKWYAVYGAGKERBAACIGVPI
polN MmpolN	287 KBERQLA AINEGHTIGISAKGAEYAKIGIGVE 287 KBERQLAHAINEGLIYGISAKGAEYARIGIGVE 365 HADREQTKWYYAVYYGAGKERLAACIGVPI 365 HMDREQTKWYYSVYGAGKERLAACIGVTV
polN MmpolN Mouse	287 KBERQLA AINNEGLI TUISAKG AEYAK GYGVE 287 KBERQLE AINNEGLI YOISAKG AEYAK GYGVE 365 HADREQT KWYYAVY CAGKER ACIGYPI 365 H
polN MmpolN Mouse Human	287 KBEPOLA AINEGLI TUISAKGAEYAKUGIGUE 287 K
polN MmpolN Mouse Human Fly	287 KBEPOLA AINFOLI TOISAKGAAEYAKIGYGVE 287 KBEPOLA AINFOLI TOISAKGAAEYAKIGYGVE 365 HADREQT KUVYAVUTGAGKERLAACIGVPI 365 HADREQT KUVYSVUTGAGKERLAACIGVTV 430 MIRREASRKSRWKKWEVASERAWIGGT SEMFINILESIAMSDIPRTPVLGCCISRALEPS 433 KVQRETARKSQWKKWEVVAERAWKGGT SEMFINILESIATSDIPRTPVLGCCISRALEPS 393 TLAEVFHRPNWQGGT SAMFINILEEIATGSQPRTPFLGGRESRALEAD

268

Miplp	322	FWYGGS SILFN LESI EQETPKTPV CGUTYSLMKK
PfPREXpol	332	KBDEHIMANNFGLIYCMNYVNEKNYANTYYGLN
eta	424	DWL QQ RQ CYGUY CMGAKS EQUG KE
SP6	351	TFIVAFILYCSGIAN AVCGUTED
M2	276	SNPDWTGKWPWLKDDGSLGFRWGDEEYKD
T 7	374	REKWLKKKWLENTPANAALRESIOOTLVESSOWVAGEOOVWKRRWIKGLDGR
Т3	374	REKELKKKELENTPAHAALRESIQOTLVESSQWVAGEQQVEWKRRWIKGLDGRE
phiYeO3-12	374	REKELKKKELENTPAHAALREAIQQSLVNSSAWIGGEQEVQWKRRWIKGLDGRK
Cvanophage	294	-ASRKGODIRGKIMOGLSGFADLNAAIQERANSGWLKGLDGRP
Synechocystis	374	VERMERTHAD FOULETMKLEAIAK Y TO VERMEYFNFWTEALROLEGK
T.aguaticus	338	EPAVAL ERMFOSNPKWRAMIEGTLEEGERREYWEITEFHEREYVEDINARVKSVE
Myctu	374	ERANDONDAWFAREGGERDYLRAVVERAEKDEYTSTULBEREYLEELDSSNROVE
Ecoli	377	KENOKWOLWFERNPGWLEWMERTRAOANEO YWEINDSERLYLEDIKSSNGAR
S.typhimurium	377	KHOKWIDLWFERWPGWLEYMERTRAQANEOCYVELLEHERLYLTDUKSSNAARK
Rhizobium	376	SPECIALKKYFERPECIRDYMESRKAMAPDKEYVETIFSVRINYELRSSNPSVE
Thermotoga	380	KENEKM VN FVLYPKWRDVIORVVSEANEK, YWRTIF HKEDI OLMARDRNTQ
A.thaliana	554	KENOLTUDL YNDROEWRKWOEMRKKEAIED YVLULL WSERFASKSRAOE
Rice	540	
Dictvostelium	507	NEWKETENREYEDRPEELVEORKTIETAEKYEWTREEM HYBOLE-DIKGNKGMB
A.aeolicus	321	ISLEW OVLRER FKN KAFKENHDRVKKEL EK EVKGHTLLG RFSANTFND
A.pvrophilus	321	ISEE TERNREFKNEKAFKLEHEKVKKELSEK VFRGRTLLG RFTATTFND
polN	396	ORAON ESILOKIKKIKDI ARAAIAOCIOTICUVSIMHERIPLIRIHAHDOOLE
MmpolN	396	LEATHERELOKYKKHKDEAOTVIGOCESAEYVTSELEERECAODOOLE
Mouse	490	VVQGPFTTSRVNWVVQSSAVDYLHLMLVAMKWLFEEFAIDGPFCISHDEVRYLVR-
Human	493	AVOE FWTSRVNWVVOSSAVDYLHLMLVAMKWLFEEFAIDG FCIS HDEVRYLVR-
Flv	441	TGPEREORE PTRINWVVOSGAVDFLHLMLVSMRWL CG WEFCLSFHDELRYLVK-
Min1n	361	NLEWNSELPSEINWATOSSGVDYLHLLCCSMEYTIKKYNLEALCIS HDEIRELVS-
PFPREXpol	366	MSLD CLYRYNS FEHNKC YKGHNOVKOBRALOYS SN KVIFTYFSFTK
eta	455	
SP6	375	MKWWAREETELPSWARLRENVIAAGNKEYYOOAPD WGRIRMSGGELKEHT
M2	305	PVYTPMGVETTAMAR FTTTTTAAOACYDRIIYCDTDSIHLTGTEVPEIIKDIVDPKKLG
112	505	I'IIIMOVETAMAKIIIIIAQACIDAIIICDIDOMIDICIDIIIADIVDIAADO
T 7	428	VHVBSP-INA WTI JOSA WI JC WWW KTDEM VEKCUKHGWDGDE
773	428	VHVRSP-HAAAWTLUGSAFALICE WINKTZEMIVEKGIKHGWDGDF
nhiVe03-12	428	VIIVRSP HANDETLIGSACHLICE WINKTERM TEKGIKHGWDGDF
Cyanophage	336	TRUCKNEW AND YLANSA MILLORAGE PTHELLOFAGEDY
Cyanophage	431	TUTEL DI UDUKMININDAOT I TUTELA CANALANTA CANALANT
Taquatique	303	
Muctu	129	FARE ALL ARTOSTADITIVANIOURANNEOU
F coli	423	
S turbimurium	432	
S. Cyphinalian Phisobium	432	
Thermotoga	431	AFREMATINATION AND THANKER AVEVOR
1 thaliana	400	
R. Charlana	505	
Dictucctolium	561	
A apolique	375	
A. acorreus	375	
A. pyrophilus	151	
ManolN	451	AQAL OT THE VIEW OF THE AVENUE AND TAMAASHTLTARLVAQ
Manao	431	EEDDVERVALALOUTIVITERATING THAT IK STAWATSPTLTAKLVAQ
Human	540	EEDEVEAUALOUTNULTROFFAUKIONDLPQSVAFFS
numan	549	BELODENA AMUERNI MEDALOTINLLTROMFAYKIGUNDLPQSVAFFS
Minle	49/	EKDEVELATION AMERICAN STORE ST
MIDID	418	EKDKIGAANALQUSNIWTKAMFCQQMGUNELPQNCAFFS
PIPREXPOI	418	Y WOCH CALD IN A WOLK YOU KDING
eta	510	AHAERON HTINGSAAD VI HTINN QKOLETFHSTFKSHGHR
SP6	429	MINVLINAMINELCONYMIKAFAVØRREGØALDNLGNP
M2	363	YWAHESTFKKAKYLKQKTYNQDHYWKEWNGKWKECSPDE
	Sec.	
T7	474	AYMAWVHD
Т3	474	AYMAWIHD
phiYeO3-12	474	AQMGWIHD
Cyanophage	377	YPLAFVHD
Synechocystis	483	RMLLQ
T.aquaticus	429	RMLLQ
Myctu	467	RMLLQ
E.coli	468	RVRMIMQ
S.typhimurium	468	RVRMIMQ
Rhizobium	469	RVRMLLQ
Thermotoga	471	RSKMIIQ
A.thaliana	645	RLLLQWSLLDWKLSNKGNLYWK
Rice	633	RLLL
Dictuostelium	500	KT.TM

A.aeolicus A.pyrophilus polN MmpolN Mouse Human Fly Miplp PfPREXpol eta SP6 M2	111 111 195 195 195 195 195 195 195 195 195 195 195 195 195 195 195 195 195 195	
T7 T3 phiYeO3-12 Cyanophage Synechocystis T.aquaticus Myctu E.coli S.typhimurium Rhizobium Thermotoga A.thaliana Rice Dictyostelium A.aeolicus A.pyrophilus polN MmpolN Mouse Human Fly Miplp PfPREXpol eta SP6 M2	182	AKAAMLDDDDSYYYTIGGKKSKGGKF
T7 T3 phiYeO3-12 Cyanophage Synechocystis T.aquaticus Myctu E.coli S.typhimurium Rhizobium Thermotoga A.thaliana Rice Dictyostelium	529 ICH 529 ECH 529 VCH 531 DCH 534 EAK 534 EAK 518 AAAH 521 QAH 522 QAH 523 EAK 524	
A.aeolicus A.pyrophilus polN MmpolN Mouse Human Fly Miplp PfPREXpol eta SP6 M2	161 KD	

Figure D.1. Full alignment of deduced amino acid sequences of DNA polymerase domains of DNA polymerases from family A. Details of sequences used are shown in Table 6.1, alongside PfPREXpol. These sequences were aligned using DIALIGN. Black shading indicates identical amino acids; grey shading indicates conserved amino acid substitutions. The numberings on the left hand side is the relative position of the amino acid in the respective sequences.

Figure D.2. Full alignment of deduced amino acid sequence of DNA primases

S.typhimurium	1 KLKKQGKNYHACCPFHNEKTPSFTWNGEKQGYHEFGGGAHENAIDLANYHKLEEVETV
E.coli	1 KLKKQGKNFHACCPFHNEKTPSFTUNGUKQUYHGFGGGAHINAIDULMNYUKLEUVETV
P.aeruginosa	1 QLKKTGKNYSACCPFHKEKTPSFTWSPKKQTYYFFGGGAGNALGTVADHQLEPQAV
phiYeO3-12	1MEQEN SIELYHIPEEN COMANNERS CHTYCHA
π3	1MERED SIGLFHAP EN SSDGNSWYSDCHE@C
T7	1MDNSHDSBSVBLYHIPCDNCCSSDGNSLFSDGHTBCWVC
ab-1	1MSE-OES SVEIGHVPEN CSEDANS FSDCHMECTA
B putida	1MDMKNRKHETVSERVHLPENNSSDANSUYTENHOMC AL
P.putiua	
RP	
Aquitex	I NLERVGSNIRINCFFFFDDIPSFINSFSKUFFFFGGGGGDIA VSDIDISFIAA
B.subtilis	1 QLKKQGKNIFGLCPFHGESTPSFSVSPKQIFFFFGGGGGGVVFSLKQFMISALSV
Synechocystis	1 VLKKRGRDHLGLCPFHDEKSPSFSWSPAKOMII FGGGGGGWFFNLDELGKKSTDVA
Arabidopsis	1HSGLICPTMEGGNSGERSLSLFMAPDESSATMN
P4	1MKMNVTATVSHALGHWPRILPANGIQVLKNRHQPPPVMCGRDRARFDDREGRGTWY
PfPREXpri	1TLKYCPFCPPHKYKYDNMYKHENFKNTGNSYNHR GYKCSFYDDKMKMGDLITSNFE
Τ4	1 MSSIPWIDNEFAYRALAHLPKFTQWNNSSTWKLRFRWPVMCDWKTDQNKARWYMGDNN
S.tvphimurium	61 ELAAMHNLEIPYEAGTGLSQIERHORON-LYQLMNCLNDFYQQST-HPAAKPARDY
E.coli	61 ELAAMHNLEVPFEAGSGPSQIERHOROT-LYOLMDELNTFYOOS O-OPVATSAROY
P.aeruginosa	61 ELAWRAGMDVPREERGGRGHTPROPTDSPLYPLLSWAAEFYKOAKKSHPARKWAVNYK
nhiVe03-12	39 NWTPEDEOKAEOLSTRERTEGSEPMSYDWNFGDSNERSSDUT
m3	39 HEVENNEREAKISTERT CSEPARATOS
13	41 KWTACH
ab 1	41 AWIADA
gn-1	40 EIVREDGESSGERVQSRNADCIEFARSQERFQDP
P.putida	44 HIVIGDGETPTTGGRTKRMEGLISCEFKPDM
RP	39 YSERGTGQSTHTSRAPMIPLEDDFHRMP
Aquifex	61 ELAMRYGKKLDLEKISKDEKVYVMLDRVCDFYRESLLKNREASEYWK
B.subtilis	61 HLADKYQIDFPDDITVHSGARPESSGEQKMAEAHELLKKFYHHLHINTKEGQEALDY
Synechocystis	61 DLARRYQIQIQTLEPAQKQELQRQLSLREQLYEIMWVAAGFYHHTLFQPQGQEALTY
Arabidopsis	36 RGKCELKGGVRADGGLASADPIEKVEEKITVEGIELEPLCDEIQDYF
P4	59 QCGAEDGLKLVEKVFGVSPSDAAAKVAAVTGSLPPEDPAVTTAANDETDAARKNAAA
PfPREXpri	59 TVVENNNFYEEEEEKITLNDVEVYNMNELYSKEAENARNYD
T4	61 GNIECYNCNYHAPIGIYLKEFEPDLYREYHFEIRKEKEKSRPHE
S typhimurium	118 KEGESABITO BARTGER PEGWENALKERGNNSENKALLLEAGMLVNNEOGSTYDER
S.typhimurium	118 KRCLSAFIIORFATGFAPPGWDNALKRFGNNSDNKALLLDAGMLVNNEQGSTYDRF
S.typhimurium E.coli	118 KECLSAFIIQEFATGFAPPGWDNALKRFGNNSDNKALLLDAGMLVNNEQGSTYDRF 118 KECLSHFVIARFAIGFAPPGWDNVLKRFGGNPENRQSLIDAGMLVTNDQGRSYDRF
S.typhimurium E.coli P.aeruginosa	118 KEGLSAFIIQRFATGFAPPGWDNALKRFGNNSDNKALLLDAGMLVNNEQGSTYDRF 118 KEGLSHFVIAKFAIGFAPPGWDNVLKRFGGNPENRQSLIDAGMLVTNDQGRSYDRF 120 GRGITARDFGLGFAPPGWDNLLKHLGGDNLQLKAMLDAGLLVENS-DTGKRYDRF
S.typhimurium E.coli P.aeruginosa phiYeO3-12	118 KRGLSAFIIQRFATGFAPPGWDNALKRFGNNSDNKALLLDAGMLVNNEQGSTYDRF 118 KRGLSHEVIAKFAIGFAPPGWDNVLKRFGGNPENRQSLIDAGMLVTNDQGRSYDRF 120 GRGLTGFIARDFGLGFAPPGWDNLLKHLGGDNLQLKAMLDAGLLVENS-DTGKRYDRF 82 APGTSKETCOFACYWLAKVDN
S.typhimurium E.coli P.aeruginosa phiYeO3-12 T3	118 KRGLSAFIIORFAIGFAPPGWDNALKRFGNNSDNKALLLDAGMLVNNEQGSTYDRF 118 KRGLSHFVIAKFAIGFAPPGWDNVLKRFGGNPENRQSLIDAGMLVTNDQGRSYDRF 120 GRGLTGEIARD GIGFAPPGWDNLLKHLGGDNLQLKAMLDAGLLVENS-DTGKRYDRF 82 ARGISKETCO ACYWLAKVDN
S.typhimurium E.coli P.aeruginosa phiYeO3-12 T3 T7	118 KECLSAFIIORTATGFAPPGWDNALKRFGNNSDNKALLLDAGMLVNNEQGSTYDRF 118 KEGLSHFVIAKTAIGFAPPGWDNULKRFGGNPENRQSLIDAGMLVTNDQGRSYDRF 120 GRGLTGEIARDGIGFAPPGWDNLLKHLGGDNLQLKAMLDAGLLVENS-DTGKRYDRF 82 ARGTSKETCO AGYWLAKVDN
S.typhimurium E.coli P.aeruginosa phi¥eO3-12 T3 T7 gh-1	118 KECLSAFIIQFFATGFAPPGWDNALKRFGNNSDNKALLLDAGMLVNNEQGSTYDRF 118 KECLSHFVIAKTAIGFAPPGWDNULKRFGGNPENRQSLIDAGMLVTNDQGRSYDRF 120 GRGITGFIARDFGLGFAPPGWDNULKRFGGNPENRQSLIDAGMLVTNDQGRSYDRF 120 GRGITGFIARDFGLGFAPPGWDNLLKHLGGDNLQLKAMLDAGLLVENS-DTGKRYDRF 120 ARGTSKETCO AGYWLAKVDN
S.typhimurium E.coli P.aeruginosa phiYeO3-12 T3 T7 gh-1 P.putida	118 KEGLSASIIQERAIGFAPPGWDNALKRFGNNSDNKALLLDAGMLVNNEQGSTYDRF 118 KEGLSHEVIARTAIGFAPPGWDNVLKRFGGNPENRQSLIDAGMLVTNDQGRSYDRF 120 GRGLTGSIARDFGLGFAPPGWDNLKHLGGDNLQLKAMLDAGLLVENS-DTGKRYDRF 82 ARGTSKETCOLACYWLAKVDN
S.typhimurium E.coli P.aeruginosa phiYeO3-12 T3 T7 gh-1 P.putida RP	118 KEGLSAFIIQERAIGFAPPGWDNALKRFGNNSDNKALLLDAGMLVNNEQGSTYDRF 118 KEGLSHFVIAKRAIGFAPPGWDNVLKRFGGNPENRQSLIDAGMLVTNDQGRSYDRF 120 GRGITGFIARDFGLGFAPPGWDNLKHLGGDNLQLKAMLDAGLLVENS-DTGKRYDRF 12 ARGISKETCOLAGYWLAKVDN
S.typhimurium E.coli P.aeruginosa phiYeO3-12 T3 T7 gh-1 P.putida RP Aquifex	118 KECLSADIIORTATGFAPPGWDNALKRFGNNSDNKALLLDAGMLVNNEQGSTYDRF 118 KEGLSHDVIAKTATGFAPPGWDNALKRFGNNSDNKALLLDAGMLVTNDQGRSYDRF 120 GEGLTGSIARDGIGFAPPGWDNLKRFGGNPENRQSLIDAGMLVTNDQGRSYDRF 120 GEGLTGSIARDGIGFAPPGWDNLKRFGGNPENRQSLIDAGMLVTNDQGRSYDRF 12 ARGISKETCO ACYWLAKVDN
S.typhimurium E.coli P.aeruginosa phiYeO3-12 T3 T7 gh-1 P.putida RP Aquifex B.subtilis	118 KECLSAFIIOFATGFAPPGWDNALKRFGNNSDNKALLLDAGMLVNNEQGSTYDRF 118 KECLSAFIIOFATGFAPPGWDNALKRFGNNSDNKALLLDAGMLVNNEQGSTYDRF 120 GRGLTGFIARD GLGFAPPGWDNULKRFGGNPENRQSLIDAGMLVTNDQGRSYDRF 120 GRGLTGFIARD GLGFAPPGWDNLLKHLGGDNLQLKAMLDAGLLVENS-DTGKRYDRF 82 ACTSKFTCO ACYWLAKVDN
S.typhimurium E.coli P.aeruginosa phiYeO3-12 T3 T7 gh-1 P.putida RP Aquifex B.subtilis Synechocystis	118 KECLSAFIIQFATGFAPPGWDNALKRFGNNSDNKALLLDAGMLVNNEQGSTYDRF 118 KECLSHFVIAKTAIGFAPPGWDNULKRFGGNPENRQSLIDAGMLVTNDQGRSYDRF 120 GRGITGFIARDFGIGFAPPGWDNULKRFGGNPENRQSLIDAGMLVTNDQGRSYDRF 120 GRGITGKETCC 121 GRGITGKETCC 122 ARGTSKETCC 123 ARGTSKETCC 124 ARGTSKETCC 125 ARGTSKETCC 126 ARGTSKETCC 127 ARGTSKETCC 128 ARGTSKETCC 129 ARGTSKETCC 120 ARGTSKETCC 121 ARGTSKETCC 122 ARGTSKETCC 123 ARGTSKETCC 124 FTQESICROTCYWERVMHP
S.typhimurium E.coli P.aeruginosa phiYeO3-12 T3 T7 gh-1 P.putida RP Aquifex B.subtilis Synechocystis Arabidopsis	118 KECLSAFIIQERALGFAPPGWDNALKRFGNNSDNKALLLDAGMLVNNEQGSTYDRF 118 KECLSHEVIARTALGFAPPGWDNULKRFGNNSDNKALLLDAGMLVTNDQGRSYDRF 120 GRGITGFIARD GIGFPPGWDNULKRFGGNPENRQSLIDAGMLVTNDQGRSYDRF 82 ARGISKETCO ACYWLAKVDN
S.typhimurium E.coli P.aeruginosa phiYeO3-12 T3 T7 gh-1 P.putida RP Aquifex B.subtilis Synechocystis Arabidopsis P4	118 K.GLSASIIQERAIGFAPPGWDNALKRFGNNSDNKALLLDAGMLVNNEQGSTYDRF 118 K.GLSHEVIAEFAIGFAPPGWDNVLKRFGGNPENRQSLIDAGMLVTNDQGRSYDRF 120 GRGITGSIARDFGIGFAPPGWDNLKHLGGDNLQLKAMLDAGLLVENS-DTGKRYDRF 82 ARGISKETCOLACYWLAKVDN
S.typhimurium E.coli P.aeruginosa phiYeO3-12 T3 T7 gh-1 P.putida RP Aquifex B.subtilis Synechocystis Arabidopsis P4 PfPREXpri	118 KECLSADIIONTAT GFAPPGWDNALKRFGNNSDNKALLLDAGMLVNNEQGSTYDRF 118 KEGLSHDVIAKTAIGFAPPGWDNULKRFGGNPENRQSLIDAGMLVTNDQGRSYDRF 120 GELTGSIARD GIGFAPPGWDNLLKHLGGDNLQLKAMLDAGLLVENS-DTGKRYDRF 82 ARGISKETCO ACYWLAKVDN
S.typhimurium E.coli P.aeruginosa phiYeO3-12 T3 T7 gh-1 P.putida RP Aquifex B.subtilis Synechocystis Arabidopsis P4 PfPREXpri T4	118 KCISAFIIOFATGFAPPGWDNALKRFGNNSDNKALLLDAGMLVNNEQGSTYDRF 118 KCGLSHVIAKTAIGFAPPGWDNULKRFGGNPENRQSLIDAGMLVTNDQGRSYDRF 120 GRITGFIARDGIGFPPGWDNULKRFGGNPENRQSLIDAGMLVTNDQGRSYDRF 120 GRITGFIARDGIGFPPGWDNLKHLGGDNLQLKAMLDAGLLVENS-DTGKRYDRF 82 ACTSKETCO ACYWLAKVDN
S.typhimurium E.coli P.aeruginosa phiYeO3-12 T3 T7 gh-1 P.putida RP Aquifex B.subtilis Synechocystis Arabidopsis P4 PfPREXpri T4	118 KCISACIIO, FATGFAPPGWDNALKRFGNNSDNKALLLDAGMLVNNEQGSTYDRF 118 KCOLSHEVIAK FAIGFAPPGWDNULKRFGGNPENRQSLIDAGMLVTNDQGRSYDRF 120 GGITGEIARD GIGFAPPGWDNLLKHLGGDNLQLKAMLDAGLLVENS-DTGKRYDRF 120 GRGTSKETCO ACYWLAKVDN
S.typhimurium E.coli P.aeruginosa phiYeO3-12 T3 T7 gh-1 P.putida RP Aquifex B.subtilis Synechocystis Arabidopsis P4 PfPREXpri T4	118 KCISACIIQTAIGFAPGWDNALKRFGNNSDNKALLLDAGMLVNNEQGSTYDRF 118 KCISACIIQTAIGFAPGWDNLKRFGNNSDNKALLLDAGMLVNNEQGSTYDRF 120 GRGITGCIARD GIGF PPGWDNLKRFGGNPENRQSLIDAGMLVTNDQGRSYDRF 120 GRGITGKETCC ACYWLAKVDN
S.typhimurium E.coli P.aeruginosa phiYeO3-12 T3 T7 gh-1 P.putida RP Aquifex B.subtilis Synechocystis Arabidopsis P4 PfPREXpri T4	118 KECLSADIIONTAT GFAPPGWDNALKRFGNNSDNKALLLDAGMLVNNEQGSTYDRF 118 KECLSADIIONTAT GFAPPGWDNALKRFGNNSDNKALLLDAGMLVNNEQGSTYDRF 120 GELTGSIARD GLGFAPPGWDNLKRFGGNPENRQSLIDAGMLVTNDQGRSYDRF 120 GELTGSIARD GLGFAPPGWDNLKHLGGDNLQLKAMLDAGLLVENS-DTGKRYDRF 12 ARGTSKETCC ACYWLAKVDN
S.typhimurium E.coli P.aeruginosa phiYeO3-12 T3 T7 gh-1 P.putida RP Aquifex B.subtilis Synechocystis Arabidopsis P4 PfPREXpri T4 S.typhimurium	118 KECLSAFIIOFATGFAPPGWDNALKRFGNNSDNKALLLDAGMLVNNEQGSTYDRF 118 KECLSAFIIOFATGFAPPGWDNLKRFGGNPENRQSLIDAGMLVTNDQGRSYDRF 120 GRITGFIARDGIGFAPPGWDNLKRFGGNPENRQSLIDAGMLVTNDQGRSYDRF 120 GRITGFIARDGIGFAPPGWDNLKHLGGDNLQLKAMLDAGLLVENS-DTGKRYDRF 120 GRITGFIARDGIGFAPPGWDNLKHLGGDNLQLKAMLDAGLLVENS-DTGKRYDRF 120 GRITGFIARDGIGFAPPGWDNLKHLGGDNLQLKAMLDAGLLVENS-DTGKRYDRF 120 ARGTSKETCO AGYWLAKVDN
S.typhimurium E.coli P.aeruginosa phiYeO3-12 T3 T7 gh-1 P.putida RP Aquifex B.subtilis Synechocystis Arabidopsis P4 PfPREXpri T4 S.typhimurium E.coli	118 KECLSAFIIOFATGFAPPGWDNALKRFGNNSDNKALLLDAGMLVNNEQGSTYDRF 118 KECLSAFIIOFATGFAPPGWDNALKRFGNNSDNKALLLDAGMLVNNEQGSTYDRF 120 GRITGFIARDFGIGFAPPGWDNULKRFGGNPENRQSLIDAGMLVTNDQGRSYDRF 120 GRITGFIARDFGIGFAPPGWDNLLKHLGGDNLQLKAMLDAGLLVENS-DTGKRYDRF 120 GRITGFIARDFGIGFAPPGWDNLLKHLGGDNLQLKAMLDAGLLVENS-DTGKRYDRF 120 GRITGFIARDFGIGFAPPGWDNLLKHLGGDNLQLKAMLDAGLLVENS-DTGKRYDRF 121 ARGTSKETCO AGYWLAKVDN
S.typhimurium E.coli P.aeruginosa phiYeO3-12 T3 T7 gh-1 P.putida RP Aquifex B.subtilis Synechocystis Arabidopsis P4 PfPREXpri T4 S.typhimurium E.coli P.aeruginosa	118 KTCLSAFIIOFTATGFAPGWDNALKRFGNNSDNKALLLDAGMLVNNEQGSTYDRF 118 KTCLSAFIIOFTATGFAPGWDNALKRFGNNSDNKALLLDAGMLVNNEQGSTYDRF 120 GRGTTGFIARDFGIGFAPGWDNULKRFGGNPENRQSLIDAGMLVTNDQGRSYDRF 120 GRGTTGKETCC 120 GRGTTGKETCC 120 GRGTTGKETCC 120 GRGTTGKETCC 120 GRGTTGKETCC 22 ARGTSKETCC 24 GTSKETCC 25 ARGTSKETCC 26 ARGTSKETCC 27 ARGTSKETCC 28 ARGTSKETCC 29 ARGTSKETCC 20 ARGTSKETCC 20 ARGTSKETCC 21 KETCC 21 KETTGCR 22 SKETCC 23 ARGTSKETCC 24 FIQESICROYCYWURVURVHP
S.typhimurium E.coli P.aeruginosa phiYeO3-12 T3 T7 gh-1 P.putida RP Aquifex B.subtilis Synechocystis Arabidopsis P4 PfPREXpri T4 S.typhimurium E.coli P.aeruginosa phiYeO3-12	118 K.C.I.S.A.S.I.I.Q.F.A.I.GF.A.P.G.W.DNALKRFGNNSDNKALLLDAGMLVNNEQGSTYDRF 118 K.G.L.S.H.V.I.A.K.A.I.GF.A.P.G.W.DNULKRFGGNPENRQSLIDAGMLVTNDQGRSYDRF 120 G.G.I.T.G.F.A.I.GF.A.P.G.W.DNULKRFGGNPENRQSLIDAGMLVTNDQGRSYDRF 120 G.G.I.T.G.F.A.I.GF.A.P.G.W.DNULKRFGGNPENRQSLIDAGMLVTNDQGRSYDRF 121 G.G.I.T.G.F.G.F.A.Y.W.LAKVDN
S.typhimurium E.coli P.aeruginosa phiYeO3-12 T3 T7 gh-1 P.putida RP Aquifex B.subtilis Synechocystis Arabidopsis P4 PfPREXpri T4 S.typhimurium E.coli P.aeruginosa phiYeO3-12 T3	 118 KECLSADIIOTAT GFAPPGWDNALKRFGNNSDNKALLLDAGMLVNNEQGSTYDRF 118 KEGLSHDVIAKTAIGFAPPGWDNULKRFGGNPENRQSLIDAGMLVTNDQGRSYDRF 120 GROLTGEIARD GIGFAPPGWDNLLKHLGGDNLQLKAMLDAGLLVENS-DTGKRYDRF 120 ACTSKETCO ACYWLAKVDN
S.typhimurium E.coli P.aeruginosa phiYeO3-12 T3 T7 gh-1 P.putida RP Aquifex B.subtilis Synechocystis Arabidopsis P4 PfPREXpri T4 S.typhimurium E.coli P.aeruginosa phiYeO3-12 T3 T7	118 KCISAFIIOFATGFAPPGWDNALKRFGNNSDNKALLLDAGMLVNNEQGSTYDRF 118 KCGLSHFVIAKFAIGFAPPGWDNALKRFGNNSDNKALLLDAGMLVNNDQGRSYDRF 120 GGLTGFIARDGIGFPPGWDNULKRFGGNPENRQSLIDAGMLVNNDQGRSYDRF 120 GGLTGFIARDGIGFFPGWDNLLKHLGGDNLQLKAMLDAGLLVENS-DTGKRYDRF 12 ACTSKETCO AGYWLAKVDN
S.typhimurium E.coli P.aeruginosa phiYeO3-12 T3 T7 gh-1 P.putida RP Aquifex B.subtilis Synechocystis Arabidopsis P4 PfPREXpri T4 S.typhimurium E.coli P.aeruginosa phiYeO3-12 T3 T7 gh-1	 118 KCISACIIOTAT GFAPPGWDNALKRFGNNSDNKALLLDAGMLVNNEQGSTYDRF 118 KCISACIIOTAT GFAPPGWDNALKRFGNNSDNKALLLDAGMLVNNDQGRSYDRF 120 GRGLTGCIAFDTGGFAPPGWDNLLKHLGGDNLQLKAMLDAGLLVENS-DTGKRYDRF 120 ACTSKFTCO ACYWLAKVDN
S.typhimurium E.coli P.aeruginosa phiYeO3-12 T3 T7 gh-1 P.putida RP Aquifex B.subtilis Synechocystis Arabidopsis P4 PfPREXpri T4 S.typhimurium E.coli P.aeruginosa phiYeO3-12 T3 T7 gh-1 P.putida	118 KICLSAFIIOFATGFAPGWDNALKRFGNNSDNKALLLDAGMLVNNEQGSTYDRF 118 KICLSAFIIOFATGFAPGWDNALKRFGNNSDNKALLLDAGMLVNNEQGSTYDRF 120 GRUTGFIARD FOLGFAPGWDNULKRFGGNPENRQSLIDAGMLVTNDQGRSYDRF 120 GRUTGFIARD FOLGFAPGWDNULKRFGGNPENRQSLIDAGMLVTNDQGRSYDRF 120 GRUTGFIARD FOLGFAPGWDNULKRFGGNPENRQSLIDAGMLVTNDQGRSYDRF 120 GRUTGFIARD FOLGFAPGWDNULKHLGGDNLQLKAMLDAGLLVENS-DTGKRYDRF 122 ARGTSKFTCO ACYWLAKVDN
S.typhimurium E.coli P.aeruginosa phiYeO3-12 T3 T7 gh-1 P.putida RP Aquifex B.subtilis Synechocystis Arabidopsis P4 PfPREXpri T4 S.typhimurium E.coli P.aeruginosa phiYeO3-12 T3 T7 gh-1 P.putida RP	 118 KECLSADIIOTAT GFAPPGWDNALKRFGNNSDNKALLLDAGMLVNNEQGSTYDRF 118 KECLSADIIOTAT GFAPPGWDNALKRFGNNSDNKALLLDAGMLVNNEQGSTYDRF 120 GGLTGSIARD GLGFAPPGWDNLKRFGGNPENRQSLIDAGMLVTNDQGRSYDRF 120 GGLTGSIARD GLGFAPPGWDNLKHLGGDNLQLKAMLDAGLLVENS-DTGKRYDRF 120 AGTSKETCO ACYWLAKVDN
S.typhimurium E.coli P.aeruginosa phiYeO3-12 T3 T7 gh-1 P.putida RP Aquifex B.subtilis Synechocystis Arabidopsis P4 PfPREXpri T4 S.typhimurium E.coli P.aeruginosa phiYeO3-12 T3 T7 gh-1 P.putida RP Aquifex	118 K.GLAA IIQEFAIGF PPGWDNALKRFGNNSDNKALLLDAGMLVNNEQGSTYDRF 118 K.GLAH VIAKFAIGF PPGWDNULKRFGGNPENRQSLIDAGMLVTNDQGRSYDRF 120 G.GLTG IARDAGIGF PPGWDNULKRFGGNPENRQSLIDAGMLVTNDQGRSYDRF 120 G.GLTG IARDAGIGF PPGWDNLLKHLGGDNLQLKAMLDAGLLVENS-DTGKRYDRF 22 A.GTKKTCCHASYMLKVDN
S.typhimurium E.coli P.aeruginosa phiYeO3-12 T3 T7 gh-1 P.putida RP Aquifex B.subtilis Synechocystis Arabidopsis P4 PfPREXpri T4 S.typhimurium E.coli P.aeruginosa phiYeO3-12 T3 T7 gh-1 P.putida RP Aquifex B.subtilis	118 KEGLSAL I I QEFALGET PPGWDNALKREGNNSDNKALLLDAGMLVNNEQGSTYDREF 118 KEGLSHEVIAKEA I GET PPGWDNULKREGGNPENRQSLIDAGMLVTNDQGRSYDREF 120 GEGTEG I ARDEGI GET PPGWDNULKHLGGDNLQLKAMLDAGLLVENS-DTGKRYDREF 121 GEGTEG I COLASYML KVDN
S.typhimurium E.coli P.aeruginosa phiYeO3-12 T3 T7 gh-1 P.putida RP Aquifex B.subtilis Synechocystis Arabidopsis P4 PfPREXpri T4 S.typhimurium E.coli P.aeruginosa phiYeO3-12 T3 T7 gh-1 P.putida RP Aquifex B.subtilis Synechocystis	 118 KRGLS ALLICKER ALGE PPGWDNALKRFGNNSDNKALLLDAGMLVNNEQGSTYDRF 118 KRGLS ALLICKER ALGE PPGWDNUKRFGGNESNEALLDAGMLVTNDQGRSYDRF 120 GRGLG FIARD GLGF PPGWDNLKHLGGDNLQLKAMLDAGLLVENS-DTGKRYDRF 120 BAGT KETCC ACYWL KVDN
S.typhimurium E.coli P.aeruginosa phiYeO3-12 T3 T7 gh-1 P.putida RP Aquifex B.subtilis Synechocystis Arabidopsis P4 PfPREXpri T4 S.typhimurium E.coli P.aeruginosa phiYeO3-12 T3 T7 gh-1 P.putida RP Aquifex B.subtilis Synechocystis Arabidopsis	118 KRGLS ALLICKPAIGE PPGWDNALKRFGNNSDNKALLLDAGMLVNNEQGSTYDRF 118 KRGLS ALLICKPAIGE PPGWDNULKRFGGNPENRQSLIDAGMLVTNDQGRSYDRF 120 GRGTG GIARD CLGF PPGWDNULKHLGGDNLQLKAMLDAGLLVENS-DTGKRYDRF 82 ARGTCKETCC ACYWL KVDN
S.typhimurium E.coli P.aeruginosa phiYeO3-12 T3 T7 gh-1 P.putida RP Aquifex B.subtilis Synechocystis Arabidopsis P4 PfPREXpri T4 S.typhimurium E.coli P.aeruginosa phiYeO3-12 T3 T7 gh-1 P.putida RP Aquifex B.subtilis Synechocystis Arabidopsis P4	118 KRGLS AS I I CERTAT GET PPGWDNALKRFGNNSDNKALLLDAGMLVNNEQGSTYDRF 118 KRGLS AS I I CERTAT GET PPGWDNUKRFGGNPENRQSLIDAGMLVTNDQGRSYDRF 120 GRGIT GE IARD GL GET PPGWDNLKHLGGDNLQLKAMLDAGLLVENS-DTGKRYDRF 120 ARGIT KETCC A SYWLAKUDN
S.typhimurium E.coli P.aeruginosa phiYeO3-12 T3 T7 gh-1 P.putida RP Aquifex B.subtilis Synechocystis Arabidopsis P4 PfPREXpri T4 S.typhimurium E.coli P.aeruginosa phiYeO3-12 T3 T7 gh-1 P.putida RP Aquifex B.subtilis Synechocystis Arabidopsis P4 PfPEFyeri	118 KEGLSAB I I QEFALGER PEGWDNALKREGNNSDNKALLLDAGMLVNNEQGSTYDRE 118 KEGLSHEVIAKEALGER PEGWDNVLKREGGNPENRGSLIDAGMLVTNDQGRSYDRE 120 GEGLTGEIARDEGLGE PEGWDNVLKREGGNPENRGSLIDAGMLVTNDQGRSYDRE 120 GEGLTGEIARDEGLGE PEGWDNVLKREGGNPENRGSLIDAGMLVTNDQGRSYDRE 120 GEGLTGEIARDEGLGE PEGWDNVLKREGGNPENRGSLIDAGMLVTNDQGRSYDRE 122 AEGLSKETCCHASYMLAKVDN
S.typhimurium E.coli P.aeruginosa phiYeO3-12 T3 T7 gh-1 P.putida RP Aquifex B.subtilis Synechocystis Arabidopsis P4 PfPREXpri T4 S.typhimurium E.coli P.aeruginosa phiYeO3-12 T3 T7 gh-1 P.putida RP Aquifex B.subtilis Synechocystis Arabidopsis P4 PfPREXpri T4	118 KEGLSAB I I QEFA GF P PGWDNALKRFGNNSDNKALLLDAGMLVNN EQGSTYDRF 120 GEGLTGDIARD GF P PGWDNVLKRFGGNPENRQSLIDAGMLVTN DQGRSYDRF 120 GEGLTGDIARD GF P PGWDNVLKRFGGNPENRQSLIDAGMLVTN DQGRSYDRF 120 GEGLTGDIARD GF P PGWDNVLKRFGGNPENRQSLIDAGMLVTN DQGRSYDRF 120 GEGLTGDIARD GF P PGWDNLKHLGGDNLQLKAMLDAGLLVENS - DTGKRYDRF 120 GEGLTGDIARD GF PGWLMKVDN

S.typhimurium E.coli P.aeruginosa	235 1 235 1 238 1	LVVECYMEVVALAQYDINYAVISICTSTIADHMHMLERATNNVICCYLGR- LVVECYMEVVALAQY INYAVISICTSTIADHIQLLERATNNVICCYLGR- MVVECYMEVIALAQY INYAVISICTSTIADHIQLLERAVPSILFCELGQ- MVVECYMEVIALAQY INYAVISICTATSEEHIKRUFRLVPSILFCELGQ-
ph1YeO3-12 T3	163	TWELOOCKMPWSLCHCASCAKKCCAANTETEDO
т7	164	TVMELQDCKNPWVSLCHCASCAKKNCAANYEYNDQFEQICIMMELMDE-
gh-1	163	TWACLQGGKYPWYSICHCSKMAKK CASNYEYDTFDE THME MOE-
P.putida	157	SWSOVQNNKAPWALPNCAPMARKAIQRNIEYLESFEEWLAMALMAB
RP	144	TI VICY FULL DE DE TRNUVARI CTAL ONOANLI SKETKKVY ILVICA
B. subtilis	238	VIF GFALVITAVSDVKESITTIGTSL DDHVKI RRNVEELLCYLSK-
Synechocystis	240	ILVECYFFVIAHHES IKQTVHAHCIALSRDQVQSHMRFSQSKQTHFNHLACK-
Arabidopsis	163	AVE BGFLNCVS PD APAKVSSKEIPSEDKDTKYK LWNCNDYLKKASR VIAT GGG-
P4	221	TWHHLTGETVMMALSSVNLLSLASLARQKHPACQIMLAADRDLSGDGQKKAAAAAAAAAA
PIPREXPII	185	VNAN KARSBOCRET, KKDAPOKYETTEAYPEATKIYGVERVKDG-
14	105	
S.typhimurium	286	GROVEFMFULDGEEPDTLVEKE
E.coli	286	ACED AWRALDTALPYMTD-GROUP FMFID DGE PDTLWKE
P.aeruginosa	289	
T3	210	GUVRYVIE-CKDANECHIMG
T 7	211	GXVRVAVIE-CKDANECHLNG
gh-1	210	GWMMWVIE-FKDANCCWGQG
P.putida	204	
RP	275	
B. subtilis	289	CIVEVAMUEDGLUPDDYLKKF
Synechocystis	293	AGINGTQRAIQETEPLVYSGQVNLRTLNLPAG&DADOFTHSS
Arabidopsis	222	PEQAMAEEIARRIGKERCWRVKWPEKSEDEHFEDANEVIMSK
P4	281	VVALPPVFGDWNDAFTQYG EATRKAIYEA RPPAESPFDT SEDEFSAMSTSEKANDIY
TA	282	
14	200	A LADRET DORT TRUCTULE TO CONDUCT TO THE REAL OF THE R
S.typhimurium	327	GKAAFEAR -EQAQPSTFUEFN
E.coli	327	GKEAFEARC-EQAMPLSAFLFN
P.aeruginosa	330	GEDAFRAR TQQAQPHAPYFFQ
phire03-12	245	EDKAILEQWWWPRGWS
T7	246	HDREIMEQWWN#GPWIIPEGWVS
gh-1	245	NAKAWTDALWNWQPFWPDGWVS
P.putida	239	REQEIVTAIWNAKLYRPIGWN
RP	226	DYDAVTQAIWNKSTYTPKSID
Aquifex B subtilis	313	GGEKFKNDITDESVTMAFKO
Synechocystis	335	AENKEIYQTIVKQAPLWVWQQQ
Arabidopsis	264	GPHLIKEAHLD@EPYPILG
P4	341	EHYGEALAVDANGQLLSRYENGVWKVLPPQDFARDVAGLFQRIRAPFSSCKVASVVDTIK
PfPREXpri T4	327 271	KEENGHNIYDTTNNDYMMNKML EPRHPDTMKRMTKLMDAGERV
S.typhimurium	348	SHLPQVDHSSPDGSTQLAALALPLINQVPGDAHRIQLRQTLGLKLGIFDDSQLDRL
E.coli	348	SLMPQVDLSTPDGRARLSTLALPLISQVPGETLRIYLRQELGNKLGILDDSQLERL
P.aeruginosa	352	OUMLEADPATLEGKAHLATLAAPLLEKIPGNNLRLLMRQRLSEITGLSGENIGQLAHHSP
рп11e03-12 тз	267	
T7	268	A STRER R
gh-1	267	AKS KAR
P.putida	261	VRDILEETRKALVMGLPWFLDPLTQLTYGRRYGEVYGLGAGTGVGKTDFLTQQIAYDIQV
RP	248	GRT FDV T
Aquifex	335	ARENLEEKTREFRYYLG
Synechocystis	358	OKLKOKNIKDPLDFEOVARGMVDILKRLTDONKRAYYLOLCGEILSOGDSRLISLOVNNI.
Arabidopsis	286	FKDFFDEIDAYYDRTHGHEYGVSTGWKNLDNLYSVVPGELTVVTGIPNSGKSEWIDAMLC
P4	401	LIPQQEAPSRRLIGFRNGVLDTQNGTFHPHSPSHWMRTLCDVDFTPPVDGETLETHAPA
PfPREXpri	349	SNNKKSISSDKIKKKEE
т4	292	MFWDKSPWK

s turbimurium	404 VPKOAESGVS
E coli	404 MPKAAESGVS
P aeruginosa	412 PPSSMDHGASGVLDGDDYFAASAYYENEPSHAPFDAAPGYVEAOPRKSWNKDKKPWDG
nhiYe03-12	276
т3	276
т7	277
ah-1	276
P. putida	321 LGERVGTIFLEOKPTETAKRVAGKIAGKRFHVPKDTAGWTDEELDAAVDALGENLVMYDA
RP	257
Amifex	352
B.subtilis	409 KONKPADNSG
Synechocystis	418 SSOLTYGDRPGKNGSR
Arabidopsis	346 NLNHSVGWKFALCSMENKVRDHARKLLEKHIKKPFFDADYGRSVORMSVEEP
P4	461 FWRWLDRAAGGRAEKRDVILAALFMVLANRYDWOLFLEVTGPGGSGKSIMAEIATLLAGE
PfPREXpri	366
т4	301
5. cypnimurium	414REAL AND A REAL AND A R
E.COII	414KEVEQUARTTMKILIGLUVQNELATUVEPLENLDENKU
P.aeruginosa	470 KKWDGKKKWDKGGRGDFKAPQKTPVSVESTTLNALKTLLHHPQLALKVDDAGTLAKEQD
phiYeO3-12	2/6
T3	2/6
Т/	2//
gh-1	
P.putida	381 FGETEWDIVKRKVRYMAVSEGIKLIYIDHLTAMADTADEKGSLEQIMKEMAGLANELGI
RP	257
Aquifex	352FISDGVRRFALASEFHTKYKVI
B.subtilis	419ETKTRRAHLTTKARQKRLRPAYENAERLLLAHMLRDRSVII
Synechocystis	434HWQAKDPTSSLLEKAEALLLKIYLHCPQERPTIDQILTENDLLFSFAHHRLLWQKII
Arabidopsis	398 DEGKKWLNDTFYPIRCEMDSLPSIDWVLERAKAAVLRYGIRGLVIDPYNELDHQRTPRQ
P4	521 DNATSATIETLESPRERAALTGFSLIRLPDQERWSGDGAGLKAITGGDAVSVDPKYRDA
PIPREXpri	366IDLFI
Τ4	301
S.typhimurium	454 GLGLFKELVKTCLAOPGLTTGOLLELYRGTNDAATLEKLSMWDEIADKAIAEKTFTDSL
E.coli	454 GLGLFRELVNTCLSOPGLTTGOLLEHYRGTNNAATLEKLSMWDDIADKNIAEOTFTDSL
P aeruginosa	530 YAOLLVSLLEALOKNPROSSMOLTARWHGTPOGRILOALGEKEWI.IVOENLEKOFFDTT
nhiYe03-12	276 AMTSEDAVGLLEDGC
Τ3	276PAMTSEDAVGLLFDGC
Τ7	277
ah-1	276
P. putida	441 ITFISHLTTPEGKPHEEGGRVTIRHFKGSRAIGFWSYFMFGLERDOOAEDPVVROTTFI
RP	257TPLHGRDAEYPF
Aquifex	374 MEILLMKIEKNSOEKEIKLSFKEKIFLKGLIELKPKIDLEVLNLSPELKELAVNALNGE
B subtilis	460 KVIDBVGFOFNIDEHBALAAYLYAFYEEGAELTPOHLMABVTDEHISOLLSDILMLOVNO
Sumechocystis	491 OVREVENI.DSDPDNOLPLI.VOLAVLEOEGDENSVESI.FOLTETSAEDI.FRANI.RIPEAT
Arabidonsis	458 ETEYVSOMLTKIKRESOHHSCHVWEVAHDKOLOHWDGGADNLY
PA	581 STHIPAVILAVNNNPMPTTDRSGGVSRBBVITHEPEOTAPOER POLKOKITELAVIV
PfPPFYpri	371 COKISSNNINVNILKNKKNETDNITNKENKSDNNLKECMEKKETONEISVIEDNNNKNI
TIEVENDIT	301
14	501SKDVNDM1KKEGA1F
S.typhimurium	514 HWFESLLQLRQESLIARDRTHGESSEERRELWTLNQELARK
E.coli	514 HOFTSLLELRQESLIARERTHGESNEERLELWTLNQELAKK
P.aeruginosa	590 KISESQRFGEREERIRSVMQKSYSE TDEEKALLREHYSVAASSPSQS
phiYeO3-12	293 GUNERTLGARGGOVI
Т3	293 GINDRTLGARG
т7	294 GINEKTLGARGGEVI
gh-1	290 ETKKMTKDCRGGWTLVT
P.putida	501 IKERYTGQATGOVLYLAYDRDTGLESLTEAPEPSSPFKDESEF
RP	270 AMNSTTGGLRLGULWTITAGS
Aquifex	434 HILPKEVLEYQVENEEKLFNNILRDEQKSGKKRKKRGLKNVNT
B.subtilis	520 ELSEAELSDYVKKVLNQRNWSMIKEKEAERAEAERQKDFLRAASLAQEIVTLNRSLK
Synechocystis	551 IMEKVPWESYQKHCFGKLQQLDPRTQAEDFRYYQEQWQKAHQEIQRLESQRLNQPLN
Arabidopsis	518 VHRNRDENAGPLDLVQIGVRKVRNKVAGQIGDAYLCYDRTTGSYSDSPVTPGMPERRSP
P4	641 HIMQKFSDPMLARSILQSQQNSDEALNIKRDADPTFDFIGYLETLPQTSGMYMGNASII
PfPREXpri	431 ITENNNDDMSEKIKWEKSIEDNISYFVDNNIMYIPNNIIIKDANDCLKHN
Т4	317 OMMEYMKNNTAOG

S.typhimurium							
E.coli							
P.aeruginosa							
phiYeO3-12							
Т3							
т7							
gh-1							
P.putida							
RP							
Aquifex							
B.subtilis							
Synechocystis							
Arabidopsis	578	RY					
P4	701	RNYRKYLYHAYLAYMEANGYRNVLS	SLKMFGLGLPVMLK	EYGLNY	EKRHI	KQGIQ	FNLTLK
PfPREXpri							
Т4							
S.typhimurium							
E.coli							
P.aeruginosa							
phiYeO3-12							
T3							
Т7							
ah-1							
P.putida							
RP							
Aguifex							
B.subtilis							
Synechocystis							
Arabidopsis							
P4	761	EESYGDWLPKCDDPTTA					
PfPREXpri							
TT 4							

Figure D.2. Full alignment of deduced amino acid sequence of DNA primases. Details of sequences used are shown in Table 6.3 alongside PfPREXpri. These sequences were aligned using DIALIGN. Black shading indicates identical amino acids; grey shading indicates conserved amino acid substitutions. The numberings on the left hand side are the relative positions of the amino acid in the respective sequences.

T7 T3 T4	1 1 1	LEGKHUW-NGGKKIVVIEGEILMETUMELQDCKYPVVSLEHGASAAKKTC- LELKHUW-SGGKKIVVIEGEILAITVMELQDCKYPVVSLEHGASAAKKTC-
E.COL1	1	RHF-TEMARIQESGSPHLLTTAAESLERQGQLDSWGGFAYLAELSKN
PfPRFYheli	1	
Mmtwinkle	1	LEGLERTSREDTERMINISER STATS-OSTGLETLS PRETVCLP
Hutwinkle	1	LEGLED ISRRDAEVVIN SREL STAIN-OSTGLETIN- PRETTCLP
C.elegans	1	VIGLEVATLSDRSAVENTNER ALALYEATKALVEAPHEEI-LD
D.melanogaster	1	NKTKAWWSNLWFWWATONIETHCVVCWPYELKTLP
S.tvphimurium	1	RH F-TEMGRINDESGSPHILLTTAESLEROGOL-DSVGGFAYLAELSKN
gh-1	1	LEGROWW-SGGRKIWWEGEFECTTWAOLOGGKYPVWSTCHGSKAAKKTC-
P22	1	RKAF-EVIRKOARNRNLILGIMWAEECGDEYATAVOMTARS
RP	1	LEGOOLW-GKGKALVVTEGEFLALSVFQARKN-WPVWSVPNGAQSAKKAL-
B.subtilis	1	QKIY-NAMLVIGDRGEPVILVTVTSELANTDLLEEVCGISYLTDIANS
P.putida	1	LEGONEWASGGKKEWWEEGELLAWSWSQVQNNKWPVWSEPNGAPAARKAI-
A.aeolicus	1	KYLPTGV-KGHISVLVYKSGVTKINAYQK
S.cerevisiae	1	QDLYPSL-
HuHEL308	1	TDFATENLCSESIKNKLS TTIGNLT LQTDKHTENQSG-YEGTTEPGADLLYDVP-
MmHEL308	1	
Arabidopsis	1	LEGLDDI-EKTSEMIIVEGEIKKAZEEAGFLNCVSEPDEAPAKVSSKEI-
т7	50	E
Т3	50	EYFEQFEQITLMFD-MDDAERKA-
T4	1	ETEKYWQQRA-MFNATSKI-
E.coli	48	TPSAANISAYADIVRERAVVREM

Figure D.3. Full alignment of deduced amino acid sequence of DNA helicases

phiYeO3-12	50	ENFRQFEQULMFD-MDDACRKA-
PfPREXheli	1	
Mmtwinkle	46	PUL OFRR WEG-DDLRSWEA-
Hutwinkle	46	PAL OFRR FWLG-DDLRSWEA-
Celegans	45	PILT PLUT
Deplanator	20	OFC DAT DEFEN FWI HVDASHSWDA-
D.meranogaster	39	
S.typhimurium	48	TPSAANISAIADIVREKAVVROM
gh-1	50	EFTTFDEmtLMFD-MDDAERAA-
P22	41	CPSAANLKGYAGMVADSYQRRQV
RP	49	TLLNFEE
B.subtilis	48	VPTAANIEYYAKIVEEKSILRRL
P.putida	51	ELSFEETLMFD-MDEPCREA-
A apolicus	29	
S cerevisiae	8	TMICKGKO MFGD-TNOSIXG
U.UET 200	57	
HUHELSUS	51	
MmHEL308	1	
Arabidopsis	50	PSEDKDTKYKFLWNCND
T 7	76	VEEAAQVIPAGVRUAVI
Т3	76	VEEAAOV
T4	19	LETOTNAELPPEKENKKN
E coli	96	
shiveon 10	00	
pnire03-12	10	VELAAQV
PfPREXheli	1	
Mmtwinkle	72	AKLFARKNPK CS VRPGNQQPRPLEALNQGLSLPRILRTALP
Hutwinkle	72	AKLFARKINPK CF VRPGDQQPRPLEALNGGFNLSRILRTALP
C.elegans	69	AKDWGSALNTLECLEVKNEERPIELVRNGDHRLIRNSLSN-
D.melanogaster	66	ARAFALKUDERECLUTRPTETEPAPHLALRRRLNI.RHTLAKATP
C. tumbimusium	00	
S. Cyphimarian	00	
gh-1	16	SQEAAEV
P22	87	DELVKRLSSIR PRNEVK
RP	75	AEECVSLFPADEVFTAFEDG
B.subtilis	85	EDEVEDLSEAEKT MEV
P. putida	77	AOECAELFSPGCCK AT
A apolicus	29	
C coroniciao	33	
S.CELEVISIAE	00	
HUHEL308	80	ERDWKSSSHNTVNEELPHNCIEQPQQNDESSSK
MmHEL308	1	
Arabidopsis	88	AEEIARRGKE CWRVKWPKKSEDEH
T7	94	PCBDBNECHLNGHDREEMEOVWNAGPWIPDGBV
T 3	200	PC DUNECHTMGEDKALLEOVWNANPWVPDGWV
ma	94	r of beneficiation branches and the board
1.44	94	PDUCHT PD_IMPONT SUS FDS YUCHDWMDDYFA
E sali	94 37	PDVGAIPD-IMRQALSUSFDSYVGHDWMDDYEA
E.coli	94 37 104	PDVGHIPD-IMRQALSUSFDSYVGHDWMDDYEAAESRENKDEGPKNIADVLDATVA
E.coli phiYeO3-12	94 37 104 94	PDVGHIPD-IMRQALSISFDSYVGHDWMDDYEA AESRONKDEGPKNIADVLDATVA
E.coli phiYeO3-12 PfPREXheli	94 37 104 94 1	PDVG-IPD-IMRQALSISFDSYVGHDWMDDYEA AESRINKDEGPKNIADVLDATVA
E.coli phiYeO3-12 PfPREXheli Mmtwinkle	94 37 104 94 1	PDVG IPD-IMRQALSISFDSYVGHDWMDDYEA
E.coli phiYeO3-12 PfPREXheli Mmtwinkle Hutwinkle	94 37 104 94 116 116	PDVGHIPD-IMRQALSISFDSYVGHDWMDDYEA- AESRUNKDEGPKNIADVLDATVA- PYEDDNECHIMGEDKALLEQVWNANPWVPDGVV- VKIS- QIL- AWEK- SIV-
E.coli phiYeO3-12 PfPREXheli Mmtwinkle Hutwinkle C.elegans	94 37 104 94 116 116 109	PDVGHIPD-IMRQALSISFDSYVGHDWMDDYEA
E.coli phiYeO3-12 PfPREXheli Mmtwinkle Hutwinkle C.elegans D.melanogaster	94 37 104 94 116 116 109 110	PDVGUIPD-IMRQALSISFDSYVGHDWMDDYEA- ABSRANKDEGEKNIAD/LDATVA- PY DD NECHIMGEDKALLEQVWNANPWVPDGVV- VK S- AW K- SV AW K- SV
E.coli phiYeO3-12 PfPREXheli Mmtwinkle Hutwinkle C.elegans D.melanogaster S.tuphimurium	94 37 104 94 116 116 109 110	PDVGHIPD-IMRQALSISFDSYVGHDWMDDYEA- ABESRANKDEGPKNIADVLDATVA- PYLDMECHIMGEDKALLEQVWNANPWVPDGVV- VKIS- QL AWEK- SV AWEK- SV VONK- AT- AEDRENKDEGPKSIDONLDATVA- AT-
E.coli phiYeO3-12 PfPREXheli Mmtwinkle Hutwinkle C.elegans D.melanogaster S.typhimurium ch-1	94 37 104 94 116 116 109 110 104	PDVGHIPD-IMRQALSISFDSYVGHDWMDDYEA- AESRGNKDEGFKNIADVLDATVA- PY DENECHINGEDKALLEQVWNANPWVPDGVV- VKIS
E.coli phiYeO3-12 PfPREXheli Mmtwinkle Hutwinkle C.elegans D.melanogaster S.typhimurium gh-1	94 37 104 94 1 116 116 109 110 104 94	PDVGHIPD-IMRQALSISFDSYVGHDWMDDYEA
E.coli phiYeO3-12 PfPREXheli Mmtwinkle Hutwinkle C.elegans D.melanogaster S.typhimurium gh-1 P22	94 37 104 94 1 116 116 109 110 104 94	PDVG IPD-IMRQALSISFDSYVGHDWMDDYEA- AESRENKDEGFKNIAD/LDATVA- PY D NECHIMGEDKALLQVWNANPWVPDGVV- VK SQIL- AW K
Pirecoli phiYeO3-12 PfPREXheli Mmtwinkle Hutwinkle C.elegans D.melanogaster S.typhimurium gh-1 P22 RP	94 37 104 94 116 116 109 110 104 94 105 95	PDVGHIPD-IMRQALSISFDSYVGHDWMDDYEA- AESRANKDEGFKNIADVLDATVA- PY DDNECHIMGEDKALLQVWNANPWVPDGVV- VKISQIL- AW K
E.coli phiYeO3-12 PfPREXheli Mmtwinkle Hutwinkle C.elegans D.melanogaster S.typhimurium gh-1 P22 RP B.subtilis	94 37 104 94 116 116 109 100 104 94 105 95 103	PDVGHIPD-IMRQALSISFDSYVGHDWMDDYEA- AESRONKDEGFKNIADVLDATVA- PY DENECHIMGEDKALEQVWNANPWVPDGVV- VKIS
<pre>Prive Content of the second seco</pre>	94 37 104 94 116 116 109 110 104 94 105 95 103 95	PDVG IPD-IMRQALSISFDSYVGHDWMDDYEA- AESRENKDEGFKNIAD/LDATVA- PY D NECHIMGEDKALLQVWNANPWVPDGVV- VK S
<pre>Privation of the second s</pre>	94 37 104 94 116 116 109 110 104 94 105 95 103 95 29	PDVG.IPD-IMRQALSISFDSYVGHDWMDDYEA- AESRENKDEGFKNIADVLDATVA- PY.D.NECHIMGEDKALLEQVWNANPWVPDGVV- VK.SQIL- AW KSV- AW K
E.coli phiYeO3-12 PfPREXheli Mmtwinkle C.elegans D.melanogaster S.typhimurium gh-1 P22 RP B.subtilis P.putida A.aeolicus S.cerevisiae	94 37 104 94 116 109 110 104 94 105 95 103 95 29 33	PDVG IPD-IMRQALSISFDSYVGHDWMDDYEA- AESRONKDEGFKNIADVLDATVA- PY DDNECHIMGEDKALLQVWNANPWVPDGVV- VKISQL- AW K
<pre>L. coli phiYeO3-12 PfPREXheli Mmtwinkle Hutwinkle C. elegans D.melanogaster S.typhimurium gh-1 P22 RP B.subtilis P.putida A.aeolicus S.ccrevisiae HuHEI.308</pre>	94 37 104 94 1 116 109 100 104 95 103 95 29 33	PDVG IPD-IMRQALSISFDSYVGHDWMDDYEA- ABSRANKDEGFKNIAD LDATVA- PY D NECHIMGEDKALLEQVWNANPWVPDGVV- VK S
E.coli phiYeO3-12 PfPREXheli Mmtwinkle Hutwinkle C.elegans D.melanogaster S.typhimurium gh-1 P22 RP B.subtilis P.putida A.aeolicus S.cerevisiae HuHEL308 MmuET.308	94 37 104 94 116 116 109 100 104 94 105 95 103 95 29 33 114	PDVG IPD-IMRQALSISFDSYVGHDWMDDYEA- AESRENKDEGFKNIADVLDATVA- PYID NECHIMGEDKALLQVWNANPWVPDGVV- VK S
E.coli phiYeO3-12 PfPREXheli Mmtwinkle Hutwinkle C.elegans D.melanogaster S.typhimurium gh-1 P22 RP B.subtilis P.putida A.aeolicus S.cerevisiae HuHEL308 MmHEL308	94 37 104 94 1 116 109 110 104 94 105 95 29 33 114	PDVG IPD-IMRQALSISFDSYVGHDWMDDYEA AESRANKDEGPKNIADVLDATVA PY DDNECHIMGEDKALLEQVWNANPWVPDGVV VK S
L. coli phiYeO3-12 PfPREXheli Mmtwinkle Hutwinkle C.elegans D.melanogaster S.typhimurium gh-1 P22 RP B.subtilis P.putida A.aeolicus S.cerevisiae HuHEL308 MmHEL308 Arabidopsis	94 37 104 94 116 109 110 104 94 105 95 103 95 29 33 114 114	PDVG IPD-IMRQALSISFDSYVGHDWMDDYEA- ABSRANKDEGPKNIAD LDATVA- PY DD NECHIMGEDKAILEQVWNANPWVPDGVV- VK 3
E.coli phiYeO3-12 PfPREXheli Mmtwinkle Hutwinkle C.elegans D.melanogaster S.typhimurium gh-1 P22 RP B.subtilis P.putida A.aeolicus S.ccrevisiae HuHEL308 MmHEL308 Arabidopsis	94 37 104 94 116 116 109 100 104 95 103 95 29 33 114 114	PDVG IPD-IMRQALSISFDSYVGHDWMDDYEA- AESRENKDEGFKNIAD LDATVA- PY D NECHIMGEDKALLEQVWNANPWVPDGVV- VK S
E.coli phiYeO3-12 PfPREXheli Mmtwinkle C.elegans D.melanogaster S.typhimurium gh-1 P22 RP B.subtilis P.putida A.aeolicus S.cerevisiae HuHEL308 Arabidopsis	94 37 104 94 116 116 109 110 104 94 105 95 103 95 29 33 114 114	PDVG IPD-IMRQALSISFDSYVGHDWMDDYEA AESRENKDEGFKNIADVLDATVA PY D NECHINGEDKALLEQVWNANPWVPDGV VK S
Tr Tr Tr Tr Tr Tr Tr Tr Tr Tr	94 37 104 1 116 106 109 110 104 95 103 95 29 33 114 114 114	PDVG IPD-IMRQALSISFDSYVGHDWMDDYEA AESRANKDEGPKNIADVLDATVA PY DDNECHIMGEDKALLEQVWNANPWVPDGV VK S
<pre>F. coli phiYeO3-12 PfPREXheli Mmtwinkle Hutwinkle C.elegans D.melanogaster S.typhimurium gh-1 P22 RP B.subtilis P.putida A.aeolicus S.cerevisiae HuHEL308 Arabidopsis T7 T3</pre>	94 377 104 94 11 116 109 100 104 95 29 33 31 114 114 127 127	PDVG IPD-IMRQALSISFDSYVGHDWMDDYEA- AESRENKDEGFKNIAD LDATVA- PYID NECHIMGEDKALLEQVWNANPWVPDGVV- VK S
<pre>F. coli phiYeO3-12 PfPREXheli Mmtwinkle Hutwinkle C.elegans D.melanogaster S.typhimurium gh-1 P22 RP B.subtilis P.putida A.aeolicus S.cerevisiae HuHEL308 Arabidopsis T7 T3 T3 T4</pre>	94 37 104 94 11 1166 109 110 104 95 103 95 29 33 114 114 114 1277 69	PDVG IPD-IMRQALSISFDSYVGHDWMDDYEA AESRENKDEGFKNIADVLDATVA PY D NECHIMGEDKALLQVWNANPWVPDGV VK S
<pre>F. coli phiYeO3-12 PfPREXheli Mmtwinkle Hutwinkle C.elegans D.melanogaster S.typhimurium gh-1 P22 RP B.subtilis P.putida A.aeolicus S.cerevisiae HuHEL308 MmHEL308 Arabidopsis T7 T3 T4 E.coli</pre>	94 37 104 94 11 116 109 110 104 95 29 95 103 95 29 95 33 114 114 127 127 69 927	PDVG IPD-IMRQALSISFDSYVGHDWMDDYEA AESRANKDEGFKNIADVLDATVA PY D NECHINGEDKALLEQVWNANPWVPDGV VK S
<pre>F. coli phiYeO3-12 PfPREXheli Mmtwinkle Hutwinkle C.elegans D.melanogaster S.typhimurium gh-1 P22 RP B.subtilis P.putida A.aeolicus S.cerevisiae HuHEL308 MmHEL308 Arabidopsis T7 T3 T4 E.coli phiYeO3-12</pre>	94 37 104 94 11 1166 1199 110 104 95 299 5103 95 299 33 314 114 114 127 127 69 127	PDVG IPD-IMRQALSISFDSYVGHDWMDDYEA- AESRANKDEGPKNIAD LDATVA- PY D NECHIMGEDKALLEQVWNANPWVPDGVV- VK 3
<pre>F. coli phiYeO3-12 PfPREXheli Mmtwinkle Hutwinkle C.elegans D.melanogaster S.typhimurium gh-1 P22 RP B.subtilis P.putida A.aeolicus S.cerevisiae HuHEL308 Arabidopsis T7 T3 T4 E.coli phiYeO3-12 pfpPrbali</pre>	94 37 104 94 11 1166 109 110 104 95 103 95 29 33 114 114 114 1277 227 27 27 27 27 27 27	PDVG IPD-IMRQALSISFDSYVGHDWMDDYEA- AESRENKDEGFKNIAD LDATVA- PYID NECHIMGEDKALLEQVWNANPWVPDGVV- VK S
<pre>F. coli phiYeO3-12 PfPREXheli Mmtwinkle Hutwinkle C.elegans D.melanogaster S.typhimurium gh-1 P22 RP B.subtilis P.putida A.aeolicus S.cerevisiae HuHEL308 MmHEL308 Arabidopsis T7 T3 T4 E.coli phiYeO3-12 PfPREXheli Wathigher </pre>	94 37 104 94 11 116 109 110 104 95 103 95 29 95 103 31 114 114 127 79 127 127 127 8 22 95	PDVG IPD-IMRQALSISFDSYVGHDWMDDYEA AESRENKDEGFKNIADVLDATVA PY D NECHINGEDKALLEQVWNANPWVPDGV VK S
Tr E.coli phiYeO3-12 PfPREXheli Mmtwinkle Hutwinkle C.elegans D.melanogaster S.typhimurium gh-1 P22 RP B.subtilis P.putida A.aeolicus S.cerevisiae HuHEL308 MmHEL308 Arabidopsis T7 T3 T4 E.coli phiYeO3-12 PfPREXheli Mmtwinkle	94 37 104 94 11 116 116 109 110 104 95 29 95 103 95 29 95 33 3114 114 114 127 127 27 27 27 27 27 27 27 27 27	PDVG IPD-IMRQALSISFDSYVGHDWMDDYEA- ABSRANKDEGPKNIAD LDATVA- PY D NECHIMGEDKALLEQVWNANPWVPDGVV- VK S
<pre>F. coli phiYeO3-12 PfPREXheli Mmtwinkle Hutwinkle C.elegans D.melanogaster S.typhimurium gh-1 P22 RP B.subtilis P.putida A.aeolicus S.cerevisiae HuHEL308 Arabidopsis T7 T3 T4 E.coli phiYeO3-12 PfPREXheli Mmtwinkle Hutwinkle</pre>	94 377 104 94 11 1166 1199 100 104 95 103 95 299 33 3114 114 127 127 69 127 127 8 123 123	PDVG IPD-IMRQALSISFDSYVGHDWMDDYEA- AESRENKDEGFKNIAD LDATVA- PYID NECHIMGEDKALLEQVWNANPWVPDGVV- VK S
<pre>F. coli phiYeO3-12 PfPREXheli Mmtwinkle Hutwinkle C.elegans D.melanogaster S.typhimurium gh-1 P22 RP B.subtilis P.putida A.aeolicus S.cerevisiae HuHEL308 MmHEL308 Arabidopsis T7 T3 T4 E.coli phiYeO3-12 PfPREXheli Mmtwinkle Hutwinkle C.elegans</pre>	94 37 104 94 11 116 109 110 104 95 103 95 29 33 114 114 114 127 127 69 127 127 8 123 123 123 109	PDVG IPD-IMRQALSISFDSYVGHDWMDDYEA AESRENKDEGFKNIADVLDATVA PY D NECHINGEDKALLEQVWNANPWVPDGV VK S
<pre>F. coli phiYeO3-12 PfPREXheli Mmtwinkle Hutwinkle C.elegans D.melanogaster S.typhimurium gh-1 P22 RP B.subtilis P.putida A.aeolicus S.cerevisiae HuHEL308 MmHEL308 Arabidopsis T7 T3 T4 E.coli phiYeO3-12 PfPREXheli Mmtwinkle Hutwinkle Hutwinkle Hutwinkle K.elegans D.melanogaster</pre>	94 37 104 94 11 116 109 110 104 95 29 95 103 95 29 95 103 95 29 95 103 114 114 114 127 127 127 127 127 8 123 123 123 123 123	PDVG IPD-IMRQALSISFDSYVGHDWMDDYEA- ABSRANKDEGPKNIAD LDATVA- PY D NECHIMGEDKA LEQVWNANPWVPDGVV- VK S
<pre>F. coli phiYeO3-12 PfPREXheli Mmtwinkle Hutwinkle C.elegans D.melanogaster S.typhimurium gh-1 P22 RF B.subtilis P.putida A.aeolicus S.cerevisiae HuHEL308 MmHEL308 Arabidopsis T7 T3 T4 E.coli phiYeO3-12 PfPREXheli Mmtwinkle Hutwinkle Hutwinkle C.elegans D.melanogaster S.typhimurium</pre>	94 37 104 94 11 1166 109 110 104 95 29 95 103 95 29 95 103 95 29 95 103 114 114 114 127 127 27 27 27 127 127 127 127 127 12	PDVG IPD-IMRQALSISFDSYVGHDWMDDYEA- AESRENKDEGFKNIAD LDATVA- PYID NECHIMGEDKALLEQVWNANPWVPDGVV- VK S

P22	110	EIINDYTDTLDRR
RP	127	DGRT 5 V
B.subtilis	110	AFQN KEV
P.putida	128	NVRDILEE
A.aeolicus	29	
S.cerevisiae	84	QIIDDTDETPSELVRKLPSGVSPQIMTFDDLAAESEFIIDKITQLICSSAKFSDIALLSE
HuHEL308	121	RRKSIKDH
MmHEL308	1	
Arabidopsis	146	SFKDFF
Т7	135	REHUSS-EESVGLLFSGLFSG
т3	135	WKEANTS-EDAVGLLFDGLFDG
Τ4	77	ARKVPFKLR
E.coli	128	EOLFOOPHDGVTGVNTGVNTGVNTG
phiVe03-12	135	KEANTS-EDAVGLLFDG
PfPREXheli	16	LEFIKYPDRINGVKSKTKSKT
Mmtwinkle	131	WI.GEUSNVEOAAGVRWSRRWSR
Hutwinkle	131	WLGEUSNVEOAAGLBWSRBWSR
Cologans	117	GERSUTD-LDMREGIRSDLINSTNRVVGFGOWKR
C.eregans	125	GIRGITE DEMENDINGTROPERING THE TOTOL CONTRACTOR CONTERNATION OF THE TOTOL CONTERNATION OF THE TOTOL CONTERNATION OF THE TOTOL CONTENT OF THE TOTOL CONTENT.
D. meranoyaster	120	
S.typhimurium	120	
gn-1	100	REARVIPSEPEVAP
P22	123	KN-GEESDI
RP	135	WTTPHHGRDAEYP
B.subtilis	118	VQTYDNIEQLYNRKGDITGIPTG
P.putida	136	URKALVMGLPWF
A.aeolicus	29	
S.cerevisiae	144	TNSHITA-IASILKKYGIPY
HuHEL308	129	KNAWTGNAKAQTPIFSRSKQLKDTLLSEEINVAKKTIESSSNDLGPFYSLPSKVRDLYA
MmHEL308	1	
Arabidopsis	154	TDAYYDRTHGHEYGVSTGVSTG
Т7	152	CTG MDKI-LEARGEDVIWVISGSCMCHSTWCVRQQALQWGTAMGK
T3	152	CQGINDRI-LCANGGENVNVISGSIMEVSNEVRQQALAWGKRMGK
Т4	86	IMKINKGEAETTINVIMEGVNVELSLGECSLAADYLQLG-H
E.coli	146	YDD WKKH-ACLQPS HILVAWRPSM HIVE AMNLVENAAMLQD
phiYeO3-12	152	CQG WDRI-LEARGERY INVISCEMENS THY RQQALAWGKRMGK
PfPREXheli	34	IPS KYL-Y LOMONISIW CSTSVEWILLSQLSLDYCIQG-V
Mmtwinkle	149	FPD WRLL-K HEKED TOFTEPRES HEWET SEYALDLCTQG-V
Hutwinkle	149	FPD WRIL-KCH2KEN TVFVCPUCSCAUUPUSEYALDLCSQG-V
C.elegans	150	FAVWKYL-GELEPERMTVITEGUENENEECEYSLDLFTQG-V
D.melanogaster	143	FPVWKLL-KEHEREPETHONGPWSSEVSLDLAMQG-V
S.tvphimurium	146	YODINKKI -ALORSOLI WAMRPSMOLTH AMNLCENAAMLOD
ch-1	150	HENKKMN-KDCDCBCBONTINVISCESSING RONVYNLFHNESI
911 1	136	TEELATE-GENNAEDIVELABROCHELALKIAEGVASRV-IPGSGV
PD	148	FAD STR-CHIRT WITH ACSET HER LCGE LAVSLINOD-O
RF D subtilis	140	
B.SUDUIIIS	142	
P.putida	148	
A.aeolicus	29	KAVE-RELEAEKEFEEHEPPEIEVECIKKLAQEG-I
S.cerevisiae	163	
HuHEL308	189	QFKGIEKLYEWQHTCL-M-LNSVQERKNOLISLFTSMGKULMAEILMLQELLCC
MmHEL308	1	THE THE REPORT OF THE PARTY OF THE OWNER
Arabidopsis	172	WKNMMN-L-YSVVPEDD TWONGIPNS MALCHLNHSVGW
m7	100	CI D
17	190	PHOT 3
T3	196	
T4	128	NWYISMEMAEEVCAKRIDANMLDVSLDDIDD-GH
E.coli	189	-KPV_IFSLEMPSEQIMMRSLASLSRVDQTKIRT-GQ
phiYeO3-12	196	RUGLA
PfPREXheli	77	STWWGSFEINNVKLGKVMLNQFCG-KN
Mmtwinkle	192	2NTLWGSFEISNVRLARVMLTQFAV-TR
Hutwinkle	192	2NTWWGSFEISNVRLARVMLTQFAE-GR
C.elegans	193	3RT FCSFEMPEKKILHWMLVQYAG-YDDLVFFLFPSHLRSLARTNSYKNG
D.melanogaster	186	5NTWGSFEIRNTRLAATLLRQYVG-YP
S.typhimurium	185	-KPVLIFSLEMPGEQIMMRMLASLSRVDQTRIRT-GQ
gh-1	193	PCGVA
P22	184	RRGVLIFSMEMSAIQVVERGIAGAGMMSVSVLRNPSR
RP	191	RWGYI
B. subtilis	185	-ESWAIFSLEMGAEOLVMRMLCAEGNINAONLRT-GN
Poutida	193	
A apolique	192	KWATADS
R. acorrician	165	
S.CELEVISIAE	243	- PKDUINTI DYVAT
HUHEL308	241	-KED DET DE I VAT
MINHEL308		

Arabidopsis	215	KFALC
т7	201	MINDSWEETAEDLIGLHNRVRLRQS
T3	201	MIDDISMEDTIQDMMGLNNKVRLRQS
T4	162	SYA
E.coli	224	
phiYeO3-12	201	MICONSWEDTIQDMMGLNNKVRLRQS
PfPREXheli	103	
Mmtwinkle	218	
Hutwinkle	218	
C.elegans	242	IKLPLHRVEYSNSINSWLDRFERSSSALTMLDSEEFDERSINEIAIRIHVENSGIQHVVI
D.melanogaster	212	
S.typnimurium	224	MANAZARETUODIUCI UICABUDON
gn-1	198	MARKAN CONTRACTOR OF THE OWNER
PZZ	100	NUMBER OF THE NUMBER OF THE NEW AND THE NE
RP	190	AND SMARIGLELMI VAANAPLELN
B. SUDTIIIS	220	
P.putida	197	FINDORFIETARRVAGRIAGRREHVP
A.aeolicus	10	
S.Cerevisiae	103	
HUHEL308	254	QCEKISGLSSFGIELGFFVE
MMHEL308	1	
Arabidopsis	220	Sun NAVRDHARALLERHIKAPPPDA
m7	225	
T/	220	DSLKREITENGKFDQWFBELFGNDTFHLIDSFMEA-ETERMAAN
Т3	226	DEVKKAIAEDGRFDEWYBELFGDDTFHLYDSFAEAEARDAK
T4	100	EYKGKMEKWREKSTLGRLIVKQYPTGGDANTFRSELNE
E.COLI	228	DWARISGTMGILLEKRNHYIDDSSELTPTAVRSRAR
phiYeO3-12	226	DEVKKAIAEDGRFDEWYBELFGDDTFHLYDSFTEAEARNFAKE
PIPREXneli	109	LPDIYARK-FELLPEKFLKFHSTNILOW DAG
Mmtwinkle	224	LPEYFMTFHOQQ-SIRSUDTO
Hutwinkle	224	LPEYFMTFHEQQSIRTEDT
C.elegans	302	DNLQFLINQGMMADEKFLINQGMMADEK
D.melanogaster	218	LPHYFMTFHOQQPLKPWEA
S.typhimurium	228	DWARISGTMGILLERRNWYIDDSSELTPTWRSRAR
gn-1	223	PDETTEEVFDRAFTEIFESDKEFLYDAFTEA-AERKTAK
222	225	GWA
RP	221	NELPTDELKTAFISTLGTGRWFLRDGFCSVDPUSE ND
B.SUDTIIIS	224	DwgKLTMAMGS-LSNSGHYIDDTPGIRVSBIRARCR
P.putida	222	KDTAGWTDEELDAAVWALGENWVMYDAFEETEWWIMKRKW
A.aeolicus	10	A DE LA
S.cerevisiae	10/	
Multer 200	2/4	
Prehidensis	245	
Arabidopsis	245	DIGKSVQKMSVEERDEGKKWLNDIFIPIKCEMDSLPSI
		And a second secon
T/	270	AYMRSGLGCDV11L0H1STVVSA-SG-ES-
T3		
T4	270	AYMRTGLGCDW WITHH SIXVSA-SE-ES-
E.COli	270	AYMRTGLGCDV VVI HISTVVSA-SE-ES-L
	270 205 264	AYMRTGLGCDVIVE HIST VVSA-SE-ES-E- LKLKKNFVPTITTV YEGTCKSCRIRVYSE-N RIAREHGGIGLIMITYCOMRVP-ALSD-N
phiYeO3-12	270 205 264 270	AYMRTGLGCDVIVU HISTVVSA-SE-ES-L LKLKKNFVPTITTV YUGTCKSCRIRVYSE-N RIAREHGGIGLIMI YUQDWRVP-AL-SD-N AYMRTGLGCDVIVU HISTVVSA-SE-ES-L
phiYeO3-12 PfPREXheli	270 205 264 270 141	AYMRTGLGCDVIVU HISTVVSA-SE-ES-L LKLKKNFVPTITIVY YUGTCKSCRIRVYSE-N RIAREHGGIGLIMI YUQDWRVP-AL-SD-N AYMRTGLGCDVIVU HISTVVSA-SE-ES-T DYAVYAYDWKHITTIN QF2LNI-NK-FS-T
phiYeO3-12 PfPREXheli Mmtwinkle	270 205 264 270 141 256	AYMRTGLGCDVIVULHISI VVSA-SE-ES-L
phiYeO3-12 PfPREXheli Mmtwinkle Hutwinkle	270 205 264 270 141 256 256	AYMRTGLGCDVIVU HIST VVSA-SE-ES-L
phiYeO3-12 PfPREXheli Mmtwinkle Hutwinkle C.elegans	270 205 264 270 141 256 256 318	AYMRTGLGCDVIVU HIST VVSA-SE-ES- LKLKKNFVPTIIIVY YIGTCKSCRIRVYSE-N- RIAREHGGGLMIY OLMRVP-AL-SD-N- AYMRTGLGCDVIVL HISIVVSA-SE-ES- DYAVYAYDVKHTTIN QFWLNI-NK-FS-F
phiYeO3-12 PfPREXheli Mmtwinkle Hutwinkle C.elegans D.melanogaster	270 205 264 270 141 256 256 318 250	AYMRTGLGCDVIVII HIST WVSA-SE-ES- LKLKKNFVPTIIIVIYIGTCKSCRIRVYSE-N
phiYeO3-12 PfPREXheli Mmtwinkle Hutwinkle C.elegans D.melanogaster S.typhimurium	270 205 264 270 141 256 256 318 250 264	AYMRTGLGCDVVVI HIST VVSA-SE-ES- LKLKKNFVPTIITVIYIGTCKSCRIRVYSE-N
phiYeO3-12 PfPREXheli Mmtwinkle Hutwinkle C.elegans D.melanogaster S.typhimurium gh-1	270 205 264 270 141 256 256 318 250 264 263	AYMRTGLGCDVIVL HISTVVSA-SE-ES- LKLKKNFVPTIITVIYIGTCKSCRIRVYSE-N
phiYeO3-12 PfPREXheli Mmtwinkle Hutwinkle C.elegans D.melanogaster S.typhimurium gh-1 P22	270 205 264 270 141 256 318 250 264 263 264	AYMRTGLGCDVVVUHSIVVSA-SE-ES- LKLKKNFVPTIIVYYGTCKSCRIRVYSE-N- RIAREHGGGGLMTYYGTCKSCRIRVYSE-N- AYMRTGLGCDVVVIHSIVVSA-SE-ES- DYAVYAYDVKHIIIN QFVLNI-NKSS- QHAVYVYDVCHVVIN QFVMGH-EQL-SS- QHAVYVYDVCHVVIN QFVMGH-EQL-SS- EHASYVHDVMHVTIN QFVMGH-EQL-SS- EHASYVHDVMHVTIN QFVMGV-STFRG RIFREHGGISIMTY QFVMGV-STFRG RIFREHGGISIMTY QFVMGV-SL-SD-N- AYMVEAEGCRVVLHISIVVSAMDGDQ-N- RHKQEHPNISLIKA YGLIEKP-KAER-N-
phiYeO3-12 PfPREXheli Mmtwinkle Hutwinkle C.elegans D.melanogaster S.typhimurium gh-1 P22 RP	270 205 264 270 141 256 256 318 250 264 263 260 260	AYMRTGLGCDVIVULHISIVVSA-SE-ES- LKLKKNFVPTIIVY YIGICKSCRIRVYSE-N- RIAREHGGIGLIMIY YOLMRVP-AL-SD-N- AYMRTGLGCDVIVLHISIVVSA-SE-ES- DYAVYAYDVKHIIIN OFMGH-EQL-SS- OHAVYVYDVCHVVIN OFMGH-EQL-SS- OHAVYVYDICHVIIN OFMGH-EQL-ST-
phiYeO3-12 PfPREXheli Mmtwinkle Hutwinkle C.elegans D.melanogaster S.typhimurium gh-1 P22 RP B.subtilis	270 205 264 270 141 256 256 318 260 264 263 260 260 259	AYMRTGLGCDVIVU HIST VVSA-SE-ES- LKLKKNFVPTIIIVY YIGICKSCRIRVYSE-N RIAREHGGGLIMIY Y QLMRVP-AL-SD-N AYMRTGLGCDVIVL HIST VVSA-SE-ES- DYAVYAYDVKHITIN QFVLNI-NK-FS- QHAVYVYDVCHVVIN QFVMGH-EQL-SS- QHAVYVYDTCHVTIN QFVMGH-EQL-SS- EHASYVHDVMHVTIN QFVMGH-EQL-ST- EHASYVHDVMHVTIN QFVMGH-EQL-ST- RIFREHGGISLIMIY QUMRVP-SL-SD-N AYMVEAEGCRVIVL HISTVVSAMDG-DQ- RHKQEHPNISLIMA YIGITEKP-KA-ER-N RFLTKAHEVQWIVL HISTLLSG-NE-ST- RLKQE-SGIGMILIY QLIQGS-GRS-KD-N-
phiYeO3-12 PfPREXheli Mmtwinkle Hutwinkle C.elegans D.melanogaster S.typhimurium gh-1 P22 RP B.subtilis P.putida	270 205 264 270 141 256 318 250 264 263 260 260 259 262	AYMRTGLGCDVIVU HIST WVSA-SE-ES- LKLKKNFVPTIIVV YIGTCKSCRIRVYSE-N RIAREHGGGLMIY YOMRVP-AL-SD-N AYMRTGLGCDVIVL HISTVSA-SE-ES- DYAVYAYDVKHTTIN QFMGH-EQL-SS- QHAVYVYDVCHVVIN QFMGH-EQL-SS- QHAVYVYDTCHVIIN QFMGH-EQL-ST-
phiYeO3-12 PfPREXheli Mmtwinkle Hutwinkle C.elegans D.melanogaster S.typhimurium gh-1 P22 RP B.subtilis P.putida A.aeolicus	270 205 264 270 141 256 256 318 250 264 263 260 260 259 262 76	AYMRTGLGCDVIVU HISIVVSA-SE-ES- LKLKKNFVPTIIVV YIGTCKSCRIRVYSE-N RIAREHGGGLGIMIY QUMRVP-AL-SD-N AYMRTGLGCDVIVL HISIVVSA-SE-ES- DYAVYAYDVKHVTIN QFMGH-EQL-SS- QHAVYVYDVCHVVIN QFMGH-EQL-SS- QHAVYVYDTCHVTIN QFMGH-EQL-ST-
phiYeO3-12 PfPREXheli Mmtwinkle Hutwinkle C.elegans D.melanogaster S.typhimurium gh-1 P22 RP B.subtilis P.putida A.aeolicus S.cerevisiae	270 205 264 270 141 256 258 250 264 263 260 260 259 262 76 179	AYMRTGLGCOVIVE HIST WVSA-SE-ES- LKLKKNFVPTIIIVY YGICKSCRIRVYSE-N- RIAREHGGGGLMIYY QLWRVP-AL-SD-N- AYMRTGLGCDVIVL HIST VVSA-SE-ES- DYAVYAYDVKHIIIN QFWLNI-NK-FS- QHAVYVYDVCHVVIN QFWMGH-EQL-SS- QHAVYVYDVCHVVIN QFWMGH-EQL-SS- GLA- EHASYVHDWHVTIN QFWMGV-ST-FRG EHASYVHDWHVTIN QFWMGV-ST-FRG RIFREHGGISLIMIY QLWRVP-SL-SD-N- AYMVEAEGCRVIVL HISTVVSAMDG-DQ- RHKQEHPNISLIVA YGLIEKP-KA-ER-N- RLKQE-SGIGULIY QLJQGS-GRS-KD-N- RYMAVSEGIKLIYI HITAMAD-TA-
phiYeO3-12 PfPREXheli Mmtwinkle Hutwinkle C.elegans D.melanogaster S.typhimurium gh-1 P22 RP B.subtilis P.putida A.aeolicus S.cerevisiae HuHEL308	270 205 264 270 141 256 258 250 264 263 260 260 259 262 76 179 309	AYMRTGLGCDVVVU HIST VVSA-SE-ES- LKLKKNFVPTIIIVY YGICKSCRIRVYSE-N- RIAREHGGGGLIMIY QUMRVP-AL-SD-N- AYMRTGLGCDVIVL HISTVVSA-SE-ES- DYAVYAYDVKHIIIN QFMLNI-NK-FS- QHAVYVYDVCHVVIN QFMGH-EQL-SS- QHAVYVYDVCHVVIN QFMGH-EQL-SS- ELASYVHDMHVTIN QFMGH-EQL-SS- ELASYVHDMHVTIN QFMGV-ST-FRG RIFREHGGISLMIY QUMRVP-SL-SD-N- AYMVEAEGCRVIVL HISTVVSAMDG-DQ- RHKQEHPNISLIMA YGLTEKP-KA-ER-N- RFLTKAHEVQWIVL HISTVLSG-NE-ST- RIFLTKAHEVQWIVL HISTLLSG-NE-ST- RIFLTKAHEVQWIVL HISTLLSG-NE-ST- RIFLTKAHEVQWIVL HISTLLSG-NE-ST- RIKQEISLIMIY QUNQGS-GRS-KD-N- RYMAVSEGIKLIYI HITAMAD-TA-
phiYeO3-12 PfPREXheli Mmtwinkle Hutwinkle C.elegans D.melanogaster S.typhimurium gh-1 P22 RP B.subtilis P.putida A.aeolicus S.cerevisiae HuHEL308 MmHEL308	270 205 264 270 141 256 256 318 250 264 263 260 269 262 76 179 309	AYMRTGLGCOVIVUE HIST WVSA-SE-ES- LKLKKNFVPTIIIVY YG TCKSCRIRVYSE-N- RIAREHGGIGLIMIY OLMRVP-AL-SD-N- AYMRTGLGCDVTVLHSTVVSA-SE-ES- DYAVYAYDVKHIIIN QFVLNI-NK-FS- QHAVYVYDVCHVVIN QFVMGH-EQL-SS- QHAVYVYDTCHVTIN QFVMGH-EQL-SS- EHASYVHDVMHVTIN QFVMGH-EQL-SS- EHASYVHDVMHVTIN QFVMGH-EQL-SS- EHASYVHDVMHVTIN QFVMGH-EQL-SS- RIFREHGGISLIMIY QUMRVP-SL-SD-N- AYMVEAEGCRVIVLHISIVVSAMDG-DQ- RHKQEHPNISLIMA YGLIEKP-KA-ER-N- RFLTKAHEVQWIVTHSILLSG-NE-ST- RKQE-SGIGMILTY QLIQGS-GRS-KD-N- RYMAVSEGIKLIYIH TAMAD-TA-

	0.0.0	and and a second second
Т7	298	CTCT
T3	298	
m A	236	SYTTEKA STATES
14	230	
E.coli	292	SR
phiYeO3-12	298	
DEDDDVhali	100	TYPEONTA
PIPREAMELL	109	TIELQNIA
Mmtwinkle	285	WGWGWGWG
Hutwinkle	285	RIAAODY
C . l	202	DENLODE
C.elegans	323	KINLODKI VG
D.melanogaster	279	AKFFEQDS
S typhimurium	292	RTLE AESBSB
D. Cypriniar Lan	0.00	
gn-1	292	ERKI DR.
P22	288	D_LAMAH
PP	288	BKMIDTB
RE LALA	200	
B.subtilis	287	SkSkSk
P.putida	287	EKGS EO
3 applique	05	PUCNDURU
A. aeoffcus	100	
S.cerevisiae	196	NREFNTGDKWQSNFS1LVTgS
HuHEL308	367	OKFLOAEYYTSOFRPVELKEYLKINDTIYEVDSKAENGMTFSRLLNYKY
Mertinit 200	70	ON FUNDER VERY VERY VERY VERY VERY VERY VERY VE
PIIIFIEL 300	10	QAELKAETTISQEREVELKEELKVNDTITEVDSQA
Arabidopsis	319	ETEYNSQUTEYNSQU
	1000	
T7	308	
тЗ	308	
m 4	245	
14	240	
E.coli	302	
phi Ve03-12	308	
phileos iz	500	
PfPREXheli	179	
Mmtwinkle	295	
Historia la la	205	
Hutwinkie	293	
C.elegans	333	
D_melanogaster	289	
C tombinguocor	202	
S. Cyphimurium	302	
gh-1	302	
P22	297	
1 4 4	000	
RP	298	
B.subtilis	297	
P putida	297	
r.putida	231	
A.aeolicus	105	
S.cerevisiae	217	
UNUET 200	116	COMT FEMODOUT UNIT UMPOUT DIVICOT UPCDOEFFUCEDIVARMTCEFT CEPT FUEFFECE
HUHELSUO	410	SDILKRMDPDHLVALVIEVIPNISCLVFCPSKRCENVAEMICKFLSKEIDKRKEKERCE
MmHEL308	119	SEALKKMDPDRLVALVTEAIPNYSCLVFCPSKKNCENVAEMLCKFLSKDYLNHREKEKCE
Arabidopsis	329	
T7	308	K
m 2	200	PL SCHARCES
15	300	- Ringt Maj 10
T4	246	E ALVETE
E.coli	302	STATES STAT
-b-iv-02 12	200	V CTWETC
phile03-12	300	A GI ASIG
PfPREXheli	179	KFSFTNKN
Mmtwinkle	295	AF@KF#TDNS
Untwinklo	205	
nucwinkie	220	VE KEET DAN
C.elegans	333	Y@MQL#TQNQ
D.melanogaster	289	AFUSFUTKHN
S tunhimurium	302	STAT AKELO-
o.cypninurium	302	Shinhartany
gh-1	302	KingAFuKTKN
P22	297	SUMAM_KDLK
PD	200	KING FURPTC
KF	230	NE SEVERIG
B.subtilis	297	ELESIMELQ
P.putida	297	EAGLENELG
A applique	105	
A. deolicus	100	1 Quality of the biological statement of the second st
S.cerevisiae	217	AngoleDasi
HuHEL308	476	VIKNLKNIGNGNLCPVLKRTIPFGVAYHHSGLTSDERK LEEAYSTG
MmHET.308	170	VIKSLENICNCKUCDULKETUDEGIAVHHSCITSFFPKTIFFMVSTC
PHILIE JOO	119	VIRGENIIGNURVUEVERKIVEFUIRIIGUESEEKKNEEEIIOIG
Arabidopsis	329	KampanaKampana
m 7	310	
111-1	518	
1 /	976	
T3	318	
T3 T4	318	
T3 T4	318	
T3 T4 E.coli	318 256 312	
T3 T4 E.coli phiYeO3-12	318 256 312 318	
T3 T4 E.coli phiYe03-12	318 256 312 318	
T3 T4 E.coli phiYeO3-12 PfPREXheli	318 256 312 318 189	
T3 T4 E.coli phiYeO3-12 PfPREXheli Mmtwinkle	318 256 312 318 189 305	

C.elegans	343	HITWV
D.melanogaster	299	
s typhimurium	312	WPVVA
ab-1	312	
911-1 D22	207	
F22	200	
RP	308	
B. subtilis	307	
P.putida	307	
A.aeolicus	128	YEEIDNLKEEQKKYVKPEPRYRRGLSDEEILKRAKTGTPORG
S.cerevisiae	227	QAUYKACS
HuHEL308	523	ULCLFTCTSTLAAGVNLPARRVI
MmHEL308	226	WLCL TCTSTLAAGVNLPARRVI
Arabidopsis	339	CHWWFWA
m7	226	UNDDKCKAHFFCPD
17	320	DEVCKAUEEDA
T3	320	
Τ4	263	
E.coli	317	LNRSLEQRADKR
phiYeO3-12	326	PEKGKAHEEGRA
PfPREXheli	197	PKKEDNNL
Mmtwinkle	313	EDDDKE
Hutwinkle	313	EDDDKE
C.elegans	351	TDGDTE
D melanogastor	307	PERKFROEDF
C tuphimusium	217	
S. Cyphimurium	311	
gn-1	320	PDKGKPHEEGRP
P22	312	LSRDVEKRPNKR
RP	316	NQGDKGHEDGAQ
B.subtilis	312	LSRGVEQRQDKR
P.putida	315	PE-GKPHEEGGR
A.aeolicus	170	V
S cerevisiae	235	KKKMENYT
HUNEL308	546	LEADYVAKEFUNBNOVK
MmUEI 308	249	
Arabidoncic	247	
Arabidopsis	341	Ly hwood
	~	
T7	341	CSTIDER SCAFRELS TO ALEENQQGDMPNLVL-CRU
T3	341	WSWIDER CALRELS THEATENQQGDMPN-LVL-WR UCCETTED-T
Т4	274	WNMSDIAESAGEPATAEFWHANIETEELAAAEQQL-HKQTESEMGDKNKWNKFL
E.coli	332	PVNSDLRENGSEEDDALLMFTYEDEVYHENSDLKGI-AEDTIGEQR-N
phiYeO3-12	341	VSTUDRESCARRESTINALEPNQQGDMPHLVL-BRE
PfPREXheli	205	SISSIFT VKSTREANVFILOPHVSKTNETVFF-ID
Mmtwinkle	322	MOTAS FERENAS FERENAL FORRELYTOPGKRY-MONSINNEDCO-V
Hutwinkle	322	TOTASE FERRICASE FOR NEW THOORKLYTGPGKRY-LOWSINNED CO-V
Cologano	360	
C.elegans	300	
D.melanogaster	310	TTOSOF TARATTER NUTI ODERLISVEG-ERI-DOANNESSD-L
S.typhimurium	332	PVNSDLRESCSIECDALLIGFNYEDEVYHENSDLKGI-AE
gh-1	335	IVANDIREEGGRELS TIMALEENQQGAFPHIIL-FREMACHETEE-T
P22	327	PTNALIRDES EDDATS I MEYBEAVYDENSSAAPF-AE
RP	331	SISQURENHS ALLS LVIA EPDISKGDNRSQ RU MNR NEQ-T
B.subtilis	327	PMUSDIRESCS ENDALIVAFY HODYYDKETENKNI-IEIIIAKQR-N
P.putida	329	WTIRHFK RAIGFWSYFWFGEEDQQAEDPVVROTTTFR TODEVTOAT
A apolicus	214	CTTNSTA SEVICOLNED THDEATOATEPSCLI-PL
S cereviciae	267	SNUYBE TENDOWH LD POEL FEVASTTNNLNLNDTYF
S.CELEVISIAE	507	
HUHEL308	203	-ONIGRATRAGEDTIGES TI QEKDROOVLE-LII-RPHENC
MmHEL308	266	-OWVGRALREGIDTAEDSHILLQEKDKQQVLELIS-GPHETC
Arabidopsis	358	PNWY BUSCHARFINKCUNGUINHUNRDENAGPLDL-WQUGVRKWWNWVAQ-I
т7	388	
Т3	388	IAGY EY
T4	327	MGVOK
E coli	379	PIGTWRI
nhi Ye03-12	388	TACVEV
DEDERALS	250	TAGIDEI
FIFREAREII	202	KIPY YN
MMTWINKIE	369	
Hutwinkle	369	VFPLEFN
C.elegans	409	VESDQ EM
D.melanogaster	363	CIMPLEFD
S.typhimurium	379	PIGTWRI
ah-1	382	
P22	374	CT CTU
PD	377	DICCLEV
R cubtilia	274	EPAGGIS I
B.SUDUITIS	3/4	
r.putlda	380	EVLY A

A seclicus	267	PTVLSOFAOFALSYTLFERLDLYGEEIYEILBIOYRMNKKIMEFSNKMFYE KLIAD	KS
S cerevisiae	306		
HuHEL308	603		
MmHEL308	306		
Arabidopsis	409	DAYLC	YD
		한번 영문 그는 것이 같은 것이 가지 않는 것이 같이 가 많아.	
T7	396	N	
Т3	396	RR	
Т4	340	EQQ	
E.coli	387		
phiYeO3-12	396	RR	
PfPREXheli	260	KENMTIKEISINNF	
Mmtwinkle	377	KNSLTFSIPPKSKARLKKKKK	
Hutwinkle	377	KNSLTFSIPPKNKARLKKKKK	
C.elegans	417	VFNPSTYSHTVVEF	
D.melanogaster	371	KDGLSYSTQIQNAKRKRRR	
S.typhimurium	387	FF	
gh-1	390	DKK	
P22	380		
RP	385	SVV	
B.subtilis	382	AFF	
P.putida	388	DRRR	
A.aeolicus	327	VENHTIKDLINPEKLKEIPEPFKSVLEPEKVVVFINVRGKEKQRRGSTSFYNEEEA	
S.cerevisiae	306		
HuHEL308	603	YSHLVQEFTKGIQTLFLSLIGLKIATNLDDIYHFMNGTFFGVQQKVLLKEKS	LW
MmHEL308	306	CSHLVEEFTKGIQALFLSLIGLK	
Arabidopsis	417	RTTGSYSDSPVTPGMPEEEEEE	
		전 전 것은 동안에서 가슴을 다 가지 않는 것 같아. 것 같아. 나라 같아.	
T7	398	ETGWSYSGEEE	SH
T3	398	ETGWSYTGGEG	EG
Т4	342	EVNEVAGSQQ	IQ
E.coli	389	SRFDNYAGPQY	DD
phiYeO3-12	398	ETGWSYTGEEG	EG
PfPREXheli	274		
Mmtwinkle	394	KIKDDNGLVAKKSSSGKKGAAHQNPEICLGQDPSPAQPDTSKS	SG
Hutwinkle	394	KIKDDTGPVAKKPSSGKKGATTQNSEICSGQAPTPDQPDTSKR	SK
C.elegans	431		
D.melanogaster	388	SEKTPSEN	
S.typhimurium	389	RFDNYAGPQY	DD
gh-1	392	EPMEPM	
P22	380	QRFCNGHFVAC	
RP	387	MIEALDFEDDKPTAPTD	YG
B.subtilis	384	KFVNLERRFD	DA
P.putida	390	LSLTEAPEPSSPFKDES	EF
A.aeolicus	383	KVAVKIVEYLMKIGLSEHIGVISPYEDQV	
S.cerevisiae	306		
HuHEL308	657	EITVESLRYLTEKGLQKDTIYKSEEEVQY	
MmHEL308	329	EKGLSCGDNEG	LE
Arabidopsis	434	KRRSPKK	
	Same		
т7	415	SESTDWSNDTDF	
Т3	415	DTGWTEQDGQSDF	
T4	359	AEQNRYQRNESTRAQLDALANELKF	
E.coli	406	E	
phiYeO3-12	415	DTGWTEQDGQSDF	
PfPREXheli	274		
Mmtwinkle	439		
Hutwinkle	439		
C.elegans	431		
D.melanogaster			
S.tvphimurium	406	E	-
gh-1	400		
P22	392		
RP	410	DF	
B subtilie	401	CV	
P. putida	401		Contraction of the second
P. pucida	413	NET PET I VDP	
A.aeolicus	413	-NE PEEPPVDE	
S.Cerevisiae	306		
HUHEL308	087	-NEHITKLGKASEKGTIDLAYCDILYKDLKKGLEGLVLESLLHLIYLTTPYDLVSQCN	PD
PHILHEL308	346	CHEKIIKLGQASEKGAIDLAICDTLIKDLKKGLEGLVLESLLHLIYLTTPYDLAAQSE	ED
ALADIGODSIS			

т7		
т3		
m 4		
T4		
E.coli		
nhiVeO3-12		
property 12	074	NEURICIO
PIPREANELL	614	NEAV VON 1
Mmtwinkle		
Hutwinkle		
nucwinkie	2.2.2	
C.elegans	431	PNVIIQFI
D.melanogaster		
C tombimumium		
S. Lyphimurium		
gh-1	400	
P22	392	
RP		
B.subtilis	403	PPGAPPGA
P putida		
r.puttua		
A.aeolicus	423	
S.cerevisiae	306	OSLSDAOSSLEFKTHLOEMAOVMKVSKSNKPPG
UNUET 200	746	NUTYEDOPOOL SDAPONUAATI CUSESET CKKASCOATCKKUDKNUNDI
HUHELSUS	140	WHITERQESQUSPALQUVANILGVSESFIGRASGQATGARVDAUVVAL
MmHEL308	406	WMVYFKQVTEQVSWAETSLDFVT
Arabidonsis		
ntuordoporo		
77		
17		
тз		
T4		
D		
E.COL1		
phiYeO3-12		
DEDDDVhali	202	VI DONNE
PIPREAREII	202	ILPSNKF
Mmtwinkle		
Hutwinkle		
nucwinkie	120	
C.elegans	439	FLNI
D.melanogaster		
e turbimurium		
S. Cyphilmarran		
gh-1	400	
P22	392	
DD	0.00	
RP		
B.subtilis		
P putida		
r.pucida		
A.aeolicus	423	
S.cerevisiae	339	
UNUET 200	706	VI CEVIT VALT PERMITMANUCEPPENINDECVICNILI ACAN CECCCUI DECETT FEDRUVDALI
HUHEL308	190	IP26AP11PPresuim1A2Pri umbkei 16uppleigupt 220Apt CEPPEE MAIKUP
MmHEL308		
Arabidopsis		
the second second second		
m7		
T3		
T4		
F coli		
D.COLL		
phiYeO3-12		
PfPREXheli		
Menturi plala		
MUCWINKIG		
Hutwinkle		
C.elegans	443	
D molancesato		
D.melanogaster		
S.typhimurium		
ab-1	400	
gii i	100	
P22	392	
RP		
P subtilis		
B. SUDUIIIS		
P.putida		
A.aeolicus	423	EVEVKTVDGFOG
C commining	220	
S.Cerevisiae	222	
HuHEL308	856	VELTKKLTYCVKAELIPLMEVTGVLEGRAKQLYSAGYKSLMHLANANPEVLVRTIDHLSR
MmHEL308		
1.		
Arabidopsis		
т/		
ТЗ		
mA		
1.4		
E.coli		
phiYeO3-12		
DEDDEVIS		
FIFREXNELL		

Hutwinkle		
C.elegans	443	FFCFFLSPYYADT
D.melanogaster		
S.typhimurium		
gh-1	400	
P22	392	
PD		
P subtilis		
B. SUDCITIS		
P.putida		
A.aeolicus	435	REKEVIIISFVRSNEKGEI
S.cerevisiae	339	ISFVKWFLETYFDQTMVFHQSQQALQTTGPGTVKLSTIHSAKG
HuHEL308	916	RQAKQIVSSAKMLLHEKAEALQEEVEELLRLPSDFPGAVASSTDKA
MmHEL308		
Arabidopsis		
T 7		
m 2		
13		
14		
E.COll		
phiYeO3-12		
PfPREXheli		
Mmtwinkle		
Hutwinkle		
C.elegans		
D melanogaster		
C turbimurium		
s. cyprimarium	400	
gn-1	400	
P22	392	
RP		
B.subtilis		
P.putida		
A.aeolicus	454	GFLKDYRRLNVALTRARRKLITLGN
S.cerevisiae	382	LEFPIVFLTNGSMSNFPMDTNALYVGITRARNLLYMCNMKHERLVSKSSPYSRNIMSNNL
HuHEL308		
MmHEL308		
Arabidoneie		
ATADIGOPSIS		
Τ/		
T3		
T4		
E.coli		
phiYeO3-12		
PfPREXheli		
Mmtwinkle		
Hutwinkle		
Cologana		
C.elegans		
D.melanogaster		
S.typhimurium		
gh-1	400	PEGWKPEDTSGADEAWKDQQEPDF
P22	392	DQDEARQICTASNAPAGRRKRYAQGADV
RP		
B.subtilis		
P.putida		
A.aeolicus	479	EKTLSSDEVYKOFIGYVKSIGGYINI
S cerevisiae	442	FWTYYNKDLKESUCDVKVTHGYNVORYNOLEKNEGEYRAYSSLEGCKSVERET
HUNEL309	174	THE THE DEFINITION OF THE
Meurer 200		
MINHEL308		
Arabidopsis		

Figure D.3. Full alignment of deduced amino acid sequence of DNA helicases. Details of sequences used are shown in Table 6.4 alongside PfPREXheli. These sequences were aligned using DIALIGN. Black shading indicates identical amino acids; grey shading indicates conserved amino acid substitutions. The numberings on the left hand side are the relative positions of the amino acid in the respective sequences.

GLASGOW UNIVERSITY 282