

**Characterisation of two *Plasmodium*
falciparum cell cycle related kinases and
the effect of kinase inhibitors on the
parasite**

Leonie Johanna Harmse

Thesis submitted to the Faculty of Health Sciences, University of the Witwatersrand,
Johannesburg, in fulfilment of the requirements for the degree of PhD

Johannesburg, 2006

Declaration

I , Leonie Johanna Harmse, declare that this dissertation is my own work. It is being submitted for the degree of PhD at the University of the Witwatersrand, Johannesburg. It has not been submitted for any other degree or examination at this or any other university.

L. J. Harmse

***Dedicated to
my husband Callie, my boys Kyle and Julian and
my late mother, Frances***

PUBLICATIONS AND PRESENTATIONS

Publications

Biochemical Pharmacology (2001) 62: 341-348

Structure activity relationships and inhibitory effects of various purine derivatives on the *in vitro* growth of *Plasmodium falciparum*.

Harmse L, Van Zyl R, Gray N, Schultz P, Leclerc S, Meijer L, Doerig C, Havlik I.

Presentations

South African Parasitology Society, Magoebaskloof, South Africa, September, 2005.

Cloning and characterization of *Plasmodium falciparum* cyclin G associated kinase.

L. Harmse. T.Coetzer, C Doerig.

British Society of Parasitology Conference, Manchester, United Kingdom, April 2003.

Cell cycle and signal transduction kinases in *Plasmodium falciparum*.

L. Harmse, N. Mtombeni, T. Coetzer, C. Doerig.

International Immuno-pharmacology congress, Sun City, South Africa, September 2001.

Plasmodium kinases as drug targets.

L. Harmse, R.van Zyl, I.Havlik, L.Meijer, C. Doerig.

ABSTRACT

Drug resistant malaria caused by *Plasmodium falciparum* remains a major health issue in under-developed countries. The *Plasmodium* genome project provides the facility to search for and identify novel drug targets. The regulation of the unique *P. falciparum* cell cycle remains poorly understood and a better understanding of the enzymes involved in this process may reveal parasite selective drug targets. Protein kinases are important in the control of the cell cycle and in signal transduction processes and are frequently deregulated in human disease. Recently inhibitors selective for specific kinases have been introduced to treat cancer. The objective of this study was to identify *Plasmodium* protein kinases that may be involved in the regulation of the cell cycle and to evaluate purine derivatives as potential antimalarial compounds.

Two genes potentially involved in the cell cycle were identified with a homology-based approach using data provided by the *Plasmodium* genome project. A putative homologue of cyclin G associated kinase (PfcGAK) and a NIMA related kinase (Pfnek-3) were identified. *P. falciparum* DNA was used to amplify the genes coding for the two kinases by PCR, and this DNA was cloned into expression vectors. The respective proteins were isolated and tested for *in vitro* kinase activity. The effect of purine inhibitors on *P. falciparum* survival was assessed by the ³H-hypoxanthine incorporation assay.

Both recombinant PfcGAK and Pfnek-3 were found to phosphorylate casein *in vitro*. PfcGAK displays structural features that indicate that the protein may be bound to membranes and is likely to be regulated by site-specific phosphorylation. PfcGAK has a unique nanomeric repetitive sequence that is rich in potential serine/threonine phospho-rylation sites. Pfnek-3 has a potential Pexel export sequence, which indicates that the protein is transported to the parasitophorous vacuole and functions outside the parasite.

A large number of purine related compounds (565) were screened for anti-malarial activity from which 19 compounds were selected for further analysis. There was considerable variation in the sensitivity of *P. falciparum* to the different inhibitors with a number of purines showing sub-micromolar activity against the parasite. Structure activity analysis showed that certain

structural features were associated with increased activity against *P. falciparum*. Further development of these compounds may offer useful alternatives to current anti-malarial drugs.

ACKNOWLEDGEMENTS

My supervisors, Prof Theresa Coetzer, and Prof Christian Doerig for their support and guidance throughout this study.

Prof I Havlik, for his financial support and continuous encouragement.

Financial support:

The INCO-DC program of the European Comission Contract no ERBIC18CT970217.

The French-South African Scientific and Technological Cooperation program.

The University of the Witwatersrand Research Committee.

TABLE OF CONTENTS	Page
DECLARATION	ii
DEDICATION	iii
PUBLICATIONS AND PRESENTATIONS	iv
ABSTRACT	v
ACKNOWLEDGEMENTS	vi
TABLE OF CONTENTS	vii
LIST OF FIGURES	xiv
LIST OF TABLES	xvii
LIST OF ABBREVIATIONS	xviii

Chapter 1 Introduction

1.1	Malaria is a problematic disease of developing countries	1
1.2	The <i>Plasmodium falciparum</i> life cycle	1
1.3	Protein kinases	4
1.3.1	Classification of the eukaryotic protein kinase superfamily	4
1.4	The eukaryotic cell cycle	6
1.4.1	Kinases and the cell cycle	6
1.4.1.1	Cyclin dependent kinases and the cell cycle	7
1.4.1.2	Other kinases involved in the cell cycle: Polo-, Aurora-, and NIMA kinases	8
1.4.2	The G ₁ and S phases of the cell cycle	9
1.4.2.1	The G ₁ phase	9
1.4.2.2.	The G ₁ -S checkpoint	10
1.4.2.3	The DNA synthesis (S) phase	12
1.4.3	The G ₂ and M phases of the cell cycle	13
1.4.3.1	The G ₂ phase	13

1.4.3.2	The G ₂ checkpoint	13
1.4.3.3	CDKs and the mitotic phase	14
1.5	Regulation of the cell cycle by the tumour suppressor p53	15
1.6	Cyclin G	16
1.7	Cyclin G associated kinase	17
1.8	NIMA kinases and cell cycle regulation	18
1.8.1	NIMA kinases in <i>Aspergillus nidulans</i>	19
1.8.2.	NIMA kinases in other organisms	19
1.8.3	Mammalian NIMA related kinases (Neks) and their functions	20
1.9	The cell cycle of <i>Plasmodium falciparum</i>	22
1.10	Kinases in <i>Plasmodium falciparum</i>	24
1.10.1.	<i>P. falciparum</i> cell cycle kinases and cyclins	25
1.10.1.1	<i>P. falciparum</i> protein kinase 5 (PfPK5)	25
1.10.1.2	<i>P. falciparum</i> CDK-related kinase 1 (Pferk-1)	26
1.10.1.3	<i>P. falciparum</i> CDK- related kinase 3 and 4 (Pferk-3, Pferk-4)	26
1.10.1.4	<i>P. falciparum</i> MO15 related kinase (Pfmrk)	27
1.10.1.5	<i>P. falciparum</i> protein kinase six (PfPK6)	27
1.10.1.6	<i>P. falciparum</i> cyclins	28
1.10.2	<i>P. falciparum</i> NIMA related kinases	29
1.10.3	<i>P. falciparum</i> kinases involved in signal transduction pathways	31
1.10.4	Other <i>Plasmodium falciparum</i> kinases	35
1.10.4.1.	Calcium dependent kinases in <i>P. falciparum</i>	35
1.10.4.2	<i>P. falciparum</i> casein kinases	36
1.10.4.3	Metabolic kinases	36
1.11	Non-kinase cell cycle regulatory proteins in <i>Plasmodium falciparum</i>	37
1.12	Kinases as drug targets	38
1.12.1	CDK selective inhibitors	39
1.13	Aims	42

Chapter 2 Methods

2.1	<i>Plasmodium falciparum</i> culture methods	43
2.2	Identification of kinase encoding open reading frames using the <i>P. falciparum</i> genomic database	43
2.3	DNA isolation	44

2.4	The Polymerase Chain Reaction (PCR)	45
2.4.1	Primer design for PCR of <i>in silico</i> identified kinases	45
2.4.2	Amplification of open reading frames	46
2.4.3	PCR cycling conditions	47
2.4.4	Colony screening by PCR	47
2.5	Agarose gel electrophoresis	48
2.6	Subcloning of PCR products	48
2.6.1	Preparation of the amplicons for ligation to vector	49
2.6.2	Ligation of the PCR products to the expression vector	49
2.6.3	Preparation of competent cells	50
2.6.4	Transformation of competent bacteria with plasmids	50
2.6.5	Plasmid isolation	51
2.6.6	Analysis of plasmids	52
2.7	DNA sequencing	52
2.7.1	Internal primers for the open reading frames	52
2.7.2	Sequencing of plasmid DNA	53
2.7.3	DNA sequence analysis	54
2.8	Expression of recombinant proteins	54
2.8.1	Induction of protein expression of the cloned <i>Plasmodium</i> inserts	54
2.8.2	Isolation of hexa-histidine tagged recombinant protein from bacteria	55
2.8.3	Isolation of GST-fusion proteins	56
2.9	Analysis of the recombinant protein	56
2.9.1	SDS-Polyacrylamide gel electrophoresis (SDS-PAGE)	56
2.9.2	Western blotting and detection of expressed proteins	57
2.9.3	Coomassie brilliant blue staining of PVDF membrane	58
2.9.4	Stripping and re-probing of PVDF membranes	58
2.9.5	Determination of the protein concentration	58
2.10	Kinase assays and densitometric analysis	59
2.10.1	Kinase assay	59
2.10.2	Densitometric analysis	59
2.11	Hypoxanthine incorporation assay	60
2.12	Stage experiments	60
2.13	Protein motif search	61

Chapter 3 Results

3.1	Identification of kinase encoding open reading frames	62
3.1.1	<i>P. falciparum</i> cyclin G associated kinase	62
3.1.2	<i>P. falciparum</i> Nima related kinase	64
3.2	Amplification of the PfcGAK and Pfnek-3 open reading frames	66
3.2.1	Annealing temperature	66
3.2.2	MgCl ₂ concentration	66
3.2.3	PCR for cloning	67
3.3	Cloning of PfcGAK and Pfnek-3 into expression vectors	68
3.4	Analysis of positive clones	69
3.4.1	Analysis of PfcGAK clones by plasmid preparation and restriction digestion	69
3.4.2.	Analysis of PfcGAK clones by PCR amplification	70
3.4.3	Analysis of Pfnek-3 clones by plasmid preparation and restriction digestion	71
3.5	Expression of recombinant PfcGAK and Pfnek-3	71
3.5.1.	PfcGAK recombinant protein expression	71
3.5.2	Determination of the nucleic acid sequences of the inserts of pHB6 PfcGAK clones 14 and 24	74
3.5.3	Pfnek-3 recombinant protein expression	79
3.5.4	Determination of the DNA sequence of the PfNek-3 insert	80
3.6	Kinase assays	82
3.6.1	Kinase activity of PfcGAK	82
3.6.2	Kinase activity of recombinant Pfnek enzymes	85
3.6.2.1	Kinase activity of Pfnek-1 and Pfnek-4	85
3.6.2.2	Kinase activity of Pfnek-3	85
3.7	Bio-informatic analysis of PfcGAK	87
3.7.1	Blastp analysis of PfcGAK	87
3.7.2	Comparative analysis of the catalytic domain of PfcGAK with its homologues	89
3.7.3	Comparative analysis of the carboxyl terminal region of PfcGAK with mouse, human and rat cGAKs	99
3.7.4	Comparative analysis of PfcGAK with BIKE kinases	101
3.7.5	Repetitive sequences	104

3.7.6	Prosite motif search of PfcGAK	106
3.8	Bio-informatic analysis of Pfnek-3	111
3.8.1	Blastp analysis of Pfnek-3	111
3.8.2.	Comparative amino acid sequence analysis of Pfnek-3	112
3.8.3	Prosite motif search of Pfnek-3	118
3.9	The effect of kinase inhibitors on <i>Plasmodium falciparum</i>	121
3.9.1	Screening and selection of compounds	121
3.9.2	Antimalarial activity of iso-olomoucine, olomoucine, roscovitine and isopentenyladenine	121
3.9.3	Antimalarial activity of the purvalanol series of compounds	122
3.9.4	Selective inhibitory activity of compounds 99 and 101 against <i>P. falciparum</i>	124
3.9.5	Antimalarial activity of R and S phenyl substitutions on the C2-hydroxyethyl side chain	124
3.9.6	The effect of halogen substitutions on C6 on antimalarial activity	125
3.9.7	Antimalarial activity of compound 40	126
3.9.8	The effect of purine inhibitors at an IC ₉₀ concentration on the development of the erythrocytic stages of parasite	126

Chapter 4 Discussion

4.1	Introduction	130
4.2	PfcGAK	130
4.2.1	PfcGAK gene structure	130
4.2.2	Cloning and expression of recombinant PfcGAK	131
4.2.3	PfcGAK kinase activity	132
4.2.4	Structural analysis of the putative PfcGAK protein (PFL2280w)	134
4.3	Pfnek-3	137
4.3.1	Pfnek-3 gene structure	137
4.3.2	Expression of recombinant Pfnek-3	137
4.3.3	Pfnek-3 kinase activity	137
4.3.4	Structural analysis of the putative Pfnek-3 protein (PFL0080c)	138
4.4	Identification of cell cycle regulatory kinases using a homology based approach	141

4.5	<i>Plasmodium falciparum</i> kinases as potential drug targets	141
4.6	Inhibitory effects and structure activity relationships of purine derived kinase inhibitors on the FCR-3 strain of <i>P. falciparum</i>	143
4.6.1	Antimalarial activity of iso-olomoucine, olomoucine, roscovitine and iso-pentenyladenine	145
4.6.2	Antimalarial activity of the purvalanol series of compounds	146
4.6.3	Selective inhibitory activity of compounds 99 and 101 against <i>P. falciparum</i>	147
4.6.4	Antimalarial activity of R and S phenyl substitutions on the C2-hydroxy-ethylamino side chain	148
4.6.5	The effect of halogen substitutions on C6 on antimalarial activity	149
4.6.6	Antimalarial activity of compound 40	149
4.6.7	The effect of purine inhibitors at an IC ₉₀ concentration on the erythrocytic stages of parasite	150
4.6.8	Concluding remarks regarding the purine derived inhibitors	150

Chapter 5 Summary

5.1	PfcGAK	152
5.1.1	PfcGAK expression	152
5.1.2	PfcGAK protein structure analysis	152
5.2	Pfnek-3	153
5.3	Purine derived inhibitors	154
5.4	Conclusion	155

References	156
------------	-----

Appendix

Appendix A	174
Appendix A1	Vector maps
Appendix A2	Solutions
Appendix B	175
Appendix B1	PCR primers and optimisation
Appendix B1.1	Primers for PCR
	175

Appendix B 1.2	Optimisation of PCR conditions	177
Appendix B1.2.1	Annealing temperature	177
Appendix B1.2.2	The effect of various MgCl ₂ concentrations	178
Appendix B2	The full length ORF of the cloned PfcGAK construct showing primers used for full-length sequencing	179
Appendix B3	Blastp analysis of PfcGAK - summary table	180
Appendix B4	Blastp analysis of Pfnek-3 - summary table	181
Appendix B5	Nucleic acid sequence alignment of PfcGAK inserts	183
Appendix B6	Nucleic acid sequence alignment of Pfnek-3 inserts	185
Appendix C	Clustal analysis	187
Appendix C1	Clustal alignment of PfcGAK with its most closely related putative and characterised proteins	187
Appendix C2	Clustal alignment of PfcGAK with cGAK and BIKE kinases	190
Appendix C3	Clustal alignment of the catalytic domain of PfcGAK with completely characterised cGAKs	193
Appendix C4	Clustal alignment of Pfnek-3	194
Appendix C5	Clustal alignment of human cGAK excluding PfcGAK and including bovine auxilin	197
Appendix C6	Clustal alignment of PfcGAK with AAK proteins	200
Appendix D1	Blastn analysis of cGAK nanomeric repeat region	202
Appendix D2	Blastp of a single nanomeric repeat of PfcGAK	205
Appendix E	Sequencing chromatograms	208
Appendix E1	PfcGAK sequencing chromatograms	208
Appendix E2	Pfnek-3 sequencing chromatograms	215

LIST OF FIGURES

Figure 1.1	The life cycle of the malaria parasite.	3
Figure 1.2	The conserved eukaryotic protein kinase catalytic domain.	5
Figure 1.3	Regulatory cascade of CDK/cyclin complexes during the G ₀ to S transition of the cell cycle.	10
Figure 1.4	The G ₁ -S checkpoint and the effects of the tumour suppressor p53.	11
Figure 1.5	Regulation of the G ₂ and mitotic phases of the cell cycle.	14
Figure 3.1	The ORF of PfcGAK.	63
Figure 3.2	Amino acid sequence of the putative PfcGAK kinase (PFL2280w) in PlasmoDB version 4.4.	64
Figure 3.3	The ORF of Pfnek-3 (PFL0080c) in PlasmoDB version 4.4.	65
Figure 3.4	PCR products obtained from <i>P. falciparum</i> DNA using PfcGAK and Pfnek-3 primers separated on a 1.5% agarose gel.	67
Figure 3.5	PCR products of Pfnek-3 and PfcGAK with human DNA and a spectrin positive control reaction separated on a 1.5% agarose gel .	68
Figure 3.6	Agarose gel electrophoresis (1.5%) used to estimate the amplicon concentration used in a typical ligation reaction.	69
Figure 3.7	Agarose gel electrophoresis (2%) to verify the cloning of PfcGAK inserts into pHB6.	70
Figure 3.8	Agarose gel electrophoresis (2%) of PfcGAK PCR products using pHB6 plasmid DNA as template for PfcGAK primers.	70
Figure 3.9	Agarose gel electrophoresis (2%) to verify the cloning of Pfnek-3 inserts into pQ80L.	71
Figure 3.10	Western blots of pHB6 PfcGAK recombinant protein purified by Ni ²⁺ -affinity chromatography from a number of clones, probed with anti-histidine (panel A) followed by anti-HA (panel B) antibodies.	72
Figure 3.11(a)	Partial sequence chromatogram of clone PfcGAK24 indicating the deletion site.	75
Figure 3.11(b)	Partial sequence chromatogram of PfcGAK clone 14 showing the normal sequence.	76

Figure 3.12	Figure 3.12 Portions of the sequence chromatograms showing the silent T→C mutations of PfcGAK14 and PfcGAK24	77
Figure 3.13	Amino acid sequence alignment of PfcGAK expressed from clone 24 and clone 14 with the PlasmoDB version 4.4 sequence.	78
Figure 3.14	Western blot of pQ80L Pfnek-3 recombinant protein, purified by Ni ²⁺ affinity chromatography from six clones, probed for expression by anti-histidine antibody.	79
Figure 3.15	A portion of the sequence chromatogram of Pfnek-3(4) showing the cytosine to guanosine mutation.	80
Figure 3.16	The amino acid sequence alignment of Pfnek-3 with clone 2 and 4.	81
Figure 3.17	Kinase assay of PfcGAK using different substrates (A) casein (lane 1-5) (B) myelin basic protein and (C) histone protein (lane 6-10) with different isolates of PfcGAK.	83
Figure 3.18	Casein kinase activity of different recombinant PfcGAK isolates and the GST-fusion proteins PfNek1 and PfNek4.	84
Figure 3.19	Prolonged incubation of kinase reaction mixture caused auto-phosphorylation of PfcGAK14.	85
Figure 3.20	Casein kinase activity of different recombinant Pfnek-3 isolates.	86
Figure 3.21	Analysis of the amino acid sequence alignment of the catalytic domain of PfcGAK.	92
Figure 3.22	Ribbon diagram of a typical serine/threonine kinase catalytic region.	97
Figure 3.23	Comparison of the major consensus amino acid motifs in the catalytic domain of Ser / Thr protein kinases with PfcGAK and cGAK homologous proteins.	98
Figure 3.24	An alignment of the carboxyl region of PfcGAK with human, mouse and rat cGAK proteins.	101
Figure 3.25	An alignment of PfcGAK with human and mouse Bike kinases.	104
Figure 3.26	The two repetitive sequences of PfcGAK.	104
Figure 3.27	Blastp analysis of the nanomeric repeat region of PfcGAK.	105
Figure 3.28	The hydrophobicity plot of PfcGAK.	106
Figure 3.29	The amino acid sequence of PfcGAK indicating the various motifs identified by the Prosite program.	108
Figure 3.30	Multiple amino acid sequence alignment of Pfnek-3.	114
Figure 3.31	Comparison of the major consensus amino acid motifs in the catalytic	117

	domain of Ser / Thr protein kinases with Pfnek-3 and a group of homologous Nek proteins.	
Figure 3.32	The amino acid sequence of Pfnek-3 indicating the various motifs identified by the Prosite program.	118
Figure 3.33	The hydrophobicity plot of Pfnek-3.	119
Figure 3.34	Dose response curves showing the effect of olomoucine, roscovitine and isopentenyladenine on the <i>P. falciparum</i> FCR-3 strain.	122
Figure 3.35	Dose response curves showing the effect of the various purvalanol derivatives on the <i>P. falciparum</i> FCR-3 strain.	123
Figure 3.36	Dose response curves showing the effect of compound 99 and 101 on the <i>P. falciparum</i> FCR-3 strain.	124
Figure 3.37	Dose response curves showing the effect of the various purine derivatives on the <i>P. falciparum</i> FCR-3 strain.	125
Figure 3.38	The effect of roscovitine and olomoucine at an IC ₉₀ on <i>P. falciparum</i> growth.	126
Figure 3.39	The effect of olomoucine and roscovitine at an IC ₉₀ on the stages of the erythrocytic cycle of the <i>P. falciparum</i> FCR3 strain.	127
Figure 3.40	Photographs showing the effect of olomoucine and roscovitine at IC ₉₀ concentrations on the <i>P.falciparum</i> .	128
Figure 4.1	The chemical structures of isopentenyl adenine, olomoucine and roscovitine.	145
Figure 4.2	The chemical structures of the purvalanol group of compounds.	147
Figure 4.3	The chemical structures of the compound 99 and 101.	148
Figure 4.4	The structures of various purine derivatives.	149

LIST OF TABLES

Table 2.1	Summary of expression vector characteristics	46
Table 3.1	Final PCR parameters	67
Table 3.2	Predicted properties of PfcGAK recombinant proteins compared to the predicted properties of the native protein	78
Table 3.3	Predicted properties of Pfnek-3 recombinant proteins compared to the predicted properties of the native protein	81
Table 3.4	Densitometric analysis of Coomassie stained gel and autoradiograph from Figure 3.17	83
Table 3.5	Densitometric analysis of Coomassie stained gel and autoradiograph from Figure 3.20	86
Table 3.6	Summary of PfcGAK Blastp results using WU-2 at EBI	88
Table 3.7	Summary of Pfnek-3 Blastp results using WU-2 at EBI	111
Table 4.1	Internal methionine residues as potential translation start sites	132
Table 4.2	Comparison of IC ₅₀ values of purine inhibitors tested on <i>P. falciparum</i> with those obtained from testing on the starfish oocytes CDK1/cyclin B complex	144

LIST OF ABBREVIATIONS

A	
AAK1	adaptor associated kinase
ATM	ataxia telangiectasia mutated
ATR	ATM and Rad3-related
AP1/2	adaptor associated protein
B	
BIKE	Bone morphogenic protein-2 inducible kinase
BLAST	Basic local alignment search tool
C	
cAMP	cyclic adenosine monophosphate
Cdc	cell division cycle proteins in yeast
CDK	cyclin dependent kinase
CDPK	calcium dependent protein kinase
CKI	CDK inhibitory proteins
CGAK	cyclin G associated kinase
Chk1/Chk2	checkpoint kinases
E	
E2F	transcription factor 2F
EGF	epidermal growth factor
EGFR	EGF-receptor
ERK	extracellular regulated kinase
G	
GST	glutathione S-transferase
GTP	guanosine triphosphate
GOLD	golgi dynamics
I	
IPTG	isopropyl- β -D-1 thiogalactopyranoside
M	
MAPK	mitogen activated protein kinase
MBP	myelin basic protein
Mdm	mouse double minute protein
MCM	mini chromosome maintenance protein
MPF	<u>mitosis promoting factor</u>
mTOR	mammalian target of rapamycin

N	
Nek	nima related kinase
NIMA	never in mitosis in <i>Aspergillus</i>
O	
ORC	origin of replication complex
P	
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffered saline
PCNA	proliferating cell nuclear antigen
PEST	proline-glutamic acid-serine- threonine
PEPCK	phosphoenol pyruvate carboxy kinase
PEXEL	<i>Plasmodium</i> export element
PfPK5	<i>P. falciparum</i> protein kinase 5
Pfcrk	<i>P. falciparum</i> CDK- related kinase
PfGSK-3	<i>P. falciparum</i> glycogen synthase kinase -3
Pfmrk	<i>P. falciparum</i> MO15 related kinase
PfCK1	<i>P.falciparum</i> casein kinase 1
PI3K	phosphatidyl-inositol-3- kinase
PKB	protein kinase B
PMSF	phenylmethyl-sulfonyl fluoride
pRB	retinoblastoma protein
PV	parasitophorous vacuole
PVDF	polyvinylidene fluoride
p53	tumour suppressor protein
p27	CDK inhibitor protein
p21	CDK inhibitor protein
R	
RCC1	regulator of chromosome domain 1
S	
SDS	sodium dodecyl sulphate
T	
TAE	Tris-acetate-EDTA
TBE	Tris-borate-EDTA
TBST	Tris buffered saline-Tween 20
V	
VTs	vacuolar transport signal

Chapter 1 Introduction

1.1 Malaria is a problematic disease of developing countries

Human malaria is caused by four protozoan parasites namely *Plasmodium falciparum*, *P. vivax*, *P. ovale* and *P. malariae*. Infections caused by *P. falciparum* are responsible for about 95% of the 1 to 3 million deaths per annum. Most of these deaths occur in children of under-developed countries. This is compounded by the fact that despite a huge research effort there is still no effective vaccine available for immunization against the disease. (Winstanley, 2000).

Infections caused by the parasite have become refractory to standard drug therapy mainly caused by the development of drug resistance in the parasite. Resistance to chloroquine is widespread, and resistance to the more recently introduced drugs such as pyrimethamine, atovaquone and mefloquine is developing rapidly. The mosquito vector for malaria has become resistant to many insecticides and this adds to the problem since under-developed countries rely heavily on vector control to prevent the spread of the disease. (Foote and Cowman, 1994, Mockenhaupt, 1995, Sina and Aultman, 2001, Ridley 2002).

There is thus a need for the development of cost effective non-toxic agents that will target the malaria parasite until such time as an effective vaccine becomes available. The completion of the *Plasmodium* genome-sequencing project has provided a tool for studying the parasite on a molecular level (Gardner *et al.*, 2002). This will aid the global understanding of the molecular mechanisms that drive parasite proliferation and accelerate the identification of parasite selective drug targets. Rational drug design can become a reality once a greater understanding of the parasite is achieved.

1.2 The *Plasmodium falciparum* life cycle

The parasite has a complex life cycle that is characterised by a succession of developmental stages that have different morphological forms. Some of the stages undergo extensive cell proliferation whereas in others the cell cycle is arrested. A *P. falciparum* infection in the human host is initiated when sporozoites are injected into the blood stream while the female *Anopheles* mosquito takes a blood meal. The malaria life cycle is shown in Figure 1.1. The sporozoites rapidly invade hepatocytes where they undergo asexual multiplication

(exoerythrocytic schizogony), generating several thousand merozoites. Merozoites are then released from the hepatocytes into the blood stream where they invade erythrocytes. Another round of asexual multiplication occurs in the erythrocyte and generates 8-32 new merozoites per infected cell. This 46 - 48 hour erythrocytic schizogony cycle is characterised by a succession of distinct morphological forms: the so-called ring form, apparent immediately after erythrocyte invasion, matures into a trophozoite which in turn forms a schizont. When nuclear divisions in the schizont are complete the parasite undergoes segmentation to form merozoites. This is followed by rupture of the erythrocyte and release of 8 - 32 merozoites into the bloodstream where they infect new erythrocytes. DNA synthesis is initiated in the trophozoite stage and continues during schizogony. This erythrocytic phase is responsible for the pathogenesis of malaria. (Garnham, 1988).

After erythrocyte invasion, a small number of the parasites undergo cell cycle arrest and sexual differentiation. Male and female gametocytes are essential for the transmission of the parasite back to the mosquito vector. This occurs when the mosquito takes a blood meal from an infected individual. In the mosquito midgut the male gametocytes exit their cell cycle arrest and undergo three rounds of cell division to form eight microgametes that penetrate the female macrogamete to form a diploid zygote. Zygotes develop into motile ookinets that enter the midgut epithelium where they are immobilized at the basal lamina and develop into oocysts. The oocysts undergo asexual multiplication and rupture to release thousands of cell cycle-arrested sporozoites, which accumulate in the insect's salivary glands. (Garnham, 1988, www.malaria.org).

The processes and molecules that control the progression of the parasite through its life cycle remain poorly understood. An improved understanding of these regulatory factors that control the passage of one morphological stage to the next is likely to reveal parasite specific processes that are amenable to parasite selective inhibition.

Protein kinases have been shown to play important regulatory roles in the regulation of the cell cycle in higher eukaryotes. Aberrant expression of these kinases is implicated in most types of cancer and also in neurological disorders like Alzheimer's disease. There is a huge interest worldwide to develop specific inhibitors for selected kinases. Such drugs are starting to enter the clinical arena with a number of agents currently in the various stages of clinical trials and some already in clinical use, like rapamycin (Young and Nickerson-Nutter, 2005) and imatinib (Drucker and Lydon, 2000).

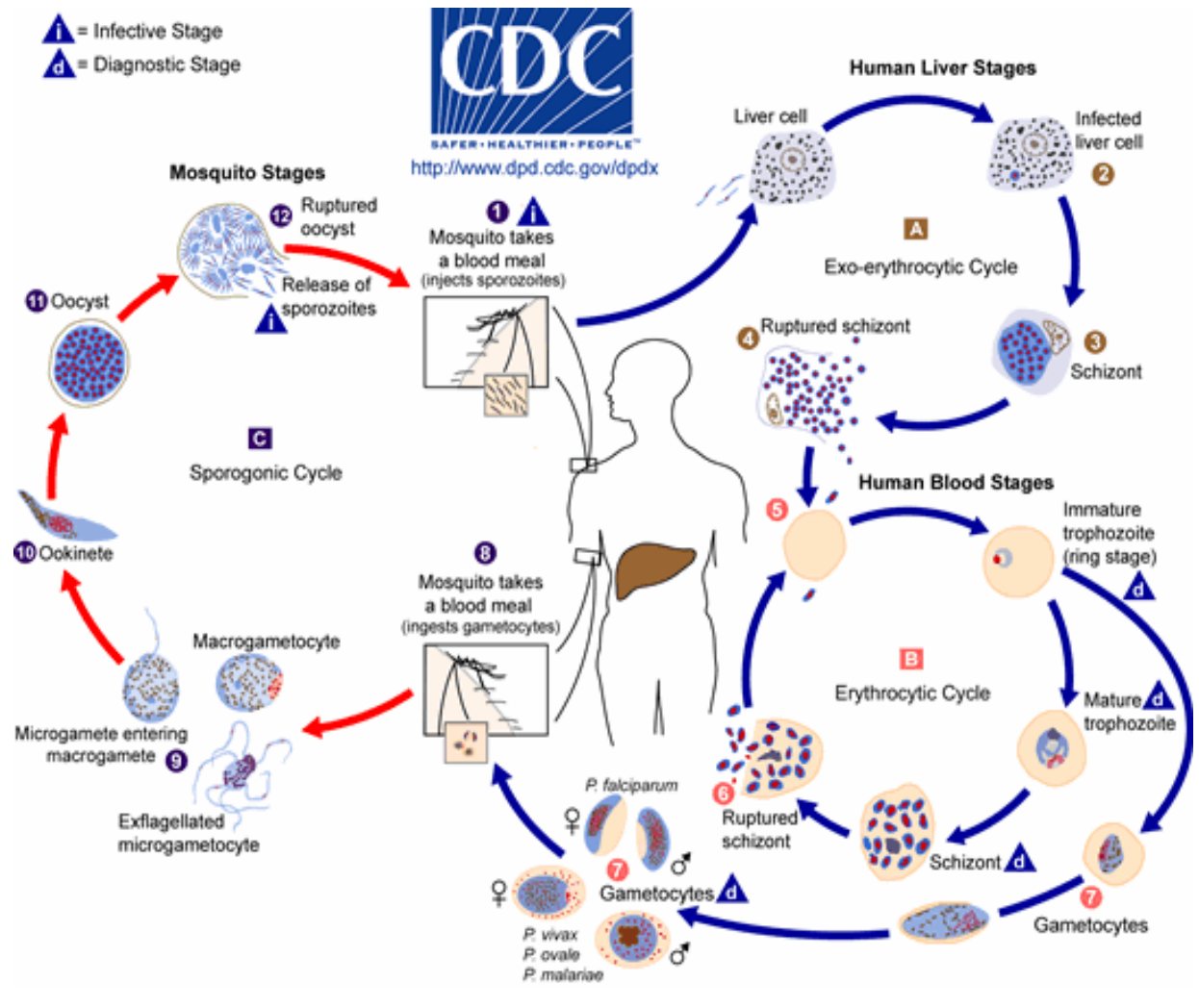


Figure 1.1 The life cycle of the malaria parasite. A malaria infection starts when an infected *Anopheles* mosquito inoculates sporozoites into the human host **1**. The sporozoites infect hepatocytes **2**, mature into schizonts **3** that rupture to release merozoites **4** into the blood stream where they infect erythrocytes **5**. In the erythrocytes the ring stage progresses to a trophozoite, which forms a mature schizont that ruptures **6** to release merozoites that invade more erythrocytes. Some parasites exit the erythrocytic cycle **7** and differentiate into male and female gametocytes **7**. The mosquito ingests gametocytes when taking a bloodmeal and the sporogonic cycle **8** then takes place in the mosquito. In the stomach of the mosquito a microgamete (male) penetrates a macrogamete (female) to generate a zygote **9** which becomes motile and elongated (ookinete) **10**. Ookinetes develop into oocysts **11** in the midgut wall where they grow and rupture to release sporozoites **12** which find their way to the mosquito's salivary glands to perpetuate the parasite's life cycle. (www.dpd.cdc.gov/dpdx).

The commitment of a small number of parasites to exit the replicative cycle and become gametocytes also provides an attractive process for drug development. It is not unlikely that protein kinases are involved in the switch from active cell division to replicative senescence. If an agent can be identified that will prevent the formation of gametocytes it will cause a disruption in the life cycle of the parasite. Hypothetically such an agent can be added to the traditional malaria treatment regimen and will prevent transmission back to the mosquito vector.

1.3 Protein kinases

Phosphorylation is a reversible post-translational covalent modification of a protein and is a ubiquitous cellular regulatory mechanism throughout living organisms. Phosphorylation controls processes such as metabolism, cell motility, assembly of the nucleus, cell cycle progression, signal transduction and operation of membrane channels. Phosphorylation of the hydroxyl groups of serine, threonine or tyrosine amino acid residues may either activate or deactivate a specific protein by inducing a conformational change in the three dimensional structure of the protein. The presence or absence of phosphate groups at specific residues may change the enzymatic activity, stability, binding properties and the subcellular localization of enzymes. (Pollard and Earnshaw, 2004: p 401-5). Non-protein molecules such as lipids and sugars are also subject to phosphorylation by kinases.

Phosphorylation is carried out by protein kinases, which catalyse the transfer of the gamma phosphate group of ATP to its specific acceptor site. The energy that accompanies the phosphorylation reaction is used to maintain the protein in a conformation that is energetically less favourable. Phosphatases remove phosphate groups from specific proteins and so modulate the activity of these proteins. This process is conserved throughout evolution with resultant high degrees of structural similarity between protein kinases as well as between phosphatases of divergent species. (Pollard and Earnshaw, 2004: p 401-5)

1.3.1 Classification of the eukaryotic protein kinase superfamily

The catalytic site of the eukaryotic protein kinases is conserved and has a number of invariant amino acids and conserved motifs as shown in Figure 1.2. Eukaryotic protein

kinases (ePKs) are classified into seven major groups (Hanks, 2003) and a group of diverse kinases that does not clearly fit into a specific group but has been shown to have kinase activity:

1. AGC group : cyclic nucleotide and calcium or phospholipid dependent kinases. This group includes the PKA, PKG, and PKC families involved in signal transduction pathways.
2. The CMGC group – this group includes the cyclin dependent – (CDK), cyclin dependent like (CDK-like), glycogen synthase (GSK) and mitogen-activated kinases (MAPK). This group is involved in the control of cell proliferation and development.
3. The calmodulin dependent kinases (CaMK).
4. The tyrosine kinases (TyrK).
5. The casein kinase 1 group.
6. The STE group (ste=sterile) includes enzymes functioning in MAPK pathways.
7. The tyrosine kinase - like enzymes - this group is structurally related to the tyrosine kinases but phosphorylates serine and threonine.
8. The “other” group - this group has been shown to have protein kinase activity but lacks high sequence similarity to the classical ePK domain. (Hanks, 2003).

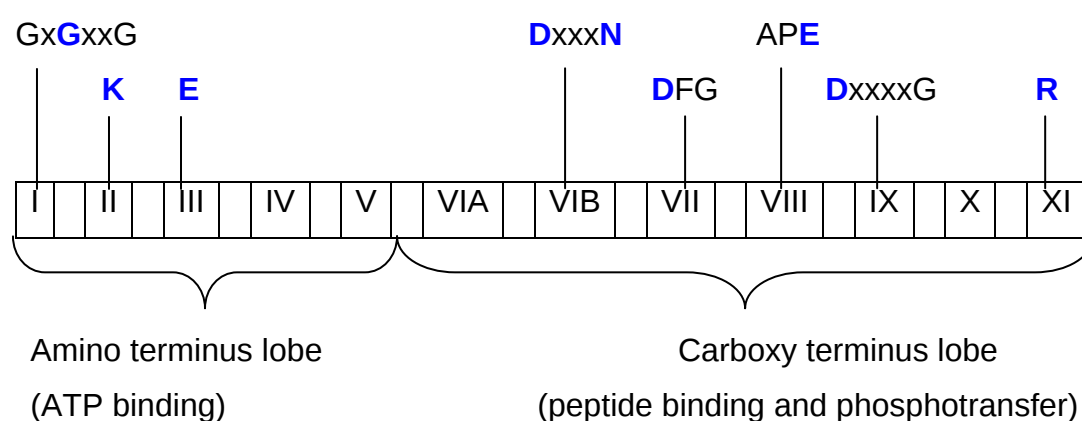


Figure 1.2 The conserved eukaryotic protein kinase catalytic domain. Invariant amino acid residues are shown in blue. The catalytic domain consists of 12 conserved subdomains indicated by Roman numerals. Highly conserved amino acid residues of the ePK superfamily are indicated by single letter abbreviations above the domain and “x” is any amino acid (Hanks, 2003).

With the exception of the tyrosine kinases (class 4) all the other eukaryotic kinases phosphorylate proteins at serine or threonine residues. The tyrosine kinases specifically phosphorylate tyrosine residues. In prokaryotes and plants, phosphorylation also occurs at histidine and aspartate residues. Protein kinase activity is in turn tightly controlled by a number of strategies: (1) reversible phosphorylation of serine or threonine residues by a regulatory kinase; (2) interactions with regulatory proteins (3) targeting to a specific cellular location where the kinase activity is required. (Pollard and Earnshaw, 2004: 401-5).

1.4 The eukaryotic cell cycle

In proliferating eukaryotic cells, the cell cycle consists of four phases, namely the G₁, S, G₂ and M phases. When cells are not proliferating they exit to the G₀ phase. The G₁ phase (Gap 1) is the interval between mitosis and DNA replication and is characterised by cell growth and protein synthesis. During this phase the cell accumulates nutrients and synthesizes the necessary enzymes required for replicating the DNA. (Grana and Reddy, 1995).

The transition that occurs at the restriction point (R) in G₁ commits the cell to the proliferative cycle. If the conditions that signal this transition are not present, the cell exits the cell cycle and enters G₀, a non-proliferative phase during which growth, terminal differentiation or apoptosis occurs. During the S (synthesis) phase the DNA complement of the cell is copied. This is followed by a second G₂ (Gap) phase during which preparation for cell division takes place. Mitosis and the production of two daughter cells are the main events of the mitotic phase (M) (Sherr and Roberts, 1999).

To control the orderly progression of the cell cycle, mammalian cells have developed regulatory pathways termed cell cycle checkpoints. Checkpoints ensure that critical events such as the correct replication of the DNA component and chromosome segregation are completed. Checkpoints halt the progression of the cell cycle until the faults are repaired or cause the activation of the apoptotic pathway of the faulty cell. Checkpoint failure has been associated with neoplastic transformation. (Malumbres and Barbacid, 2001). The overview that follows refers to the mammalian cell cycle unless otherwise stated.

1.4.1 Kinases and the cell cycle

Kinases are intimately involved in the orderly progression of the cell cycle. Several different types of kinases are important in the regulation of the cell cycle including cyclin

dependent kinases, Polo and Aurora kinases as well as NIMA kinases. Polo, Aurora and NIMA kinases are important for the completion of the mitotic process. Since NIMA kinases are an important part of this study they will be reviewed in section 1.8. Cyclin dependent kinases belong to the CMCG class of kinases while Polo, Aurora and NIMA kinases all belong to the 'other' class of kinases (Hanks, 2003).

1.4.1.1 Cyclin dependent kinases and the cell cycle

Cyclin dependent kinases (CDKs) are serine/threonine kinases and display kinase activity only when they are complexed with cyclins. They are highly conserved among eukaryotes and are present in many organisms. Currently there are eleven identified CDKs, which are expressed constitutively throughout the cell cycle whereas cyclin levels are restricted by the transcriptional regulation of cyclin encoding genes and by ubiquitin - mediated degradation (Morgan, 1997; Koepp *et al.*, 1999; Grana and Reddy, 1995). All CDKs have the bi-lobed structure of protein kinases, with the N-terminal region forming the smaller lobe and the C-terminal region forming the larger C-lobe. The catalytic site is found in a deep cleft between the two domains and is partially obstructed by the T-loop. The N-lobe also contains a helix consisting of a PSTAIRE motif, which is responsible for binding cyclin. (Pollard and Earnshaw, 2004: 679).

Structural constraints prevent monomeric CDKs from having kinase activity. In the absence of cyclin, substrates cannot access the active site of the CDK and the residues involved in the transfer of the γ -phosphate from the ATP to the substrate are misaligned. The T-loop, which obstructs the entry of ATP to its binding site, moves away when cyclin binds to the CDK. Phosphorylation of the conserved Thr160 residue in the T-loop further activates the CDK by stabilizing the kinase in its most active conformation. (De Bondt *et al.*, 1993; Jeffrey *et al.*, 1995; Morgan, 1995; Koepp *et al.*, 1999).

In contrast the CDK is inactivated by phosphorylation of Thr14/Tyr15 by the Wee1-like kinases and activated by dephosphorylation of these residues by the cell division cycle 25 (Cdc25) protein phosphatase (Malumbres and Barbacid, 2001; Lee and Yang 2001). The activity of the active CDK/cyclin complex is attenuated by targeted degradation of the cyclin by the ubiquitin proteolytic pathway or by the association of protein cyclin dependent kinase inhibitors (Willems *et al.*, 1996, Morgan, 1997).

Two families of CDK inhibitory proteins (CKI) exist, the INK4 inhibitors and the Cip/Kip inhibitors. These proteins inhibit the activity of CDK/cyclin complexes. The INK family inhibits CDK4 and CDK6 activity during the G₁ phase while the Cip/Kip inhibition of CDKs are not phase specific. The four members of the INK4 family are called p16^{INK4a} (p16), p15^{INK4b} (p15), p18^{INK4c} (p18) and p19^{INK4d} (p19). The three members of the Cip/Kip family are p21^{Waf1/Cip1}, p27^{Kip1} (p27) and p57^{Kip2} (p57). (Sherr and Roberts, 1999). Active CDK/cyclin complexes drive progression through the phases of the cell cycle. In early to mid G₁ the activity of the CDK/cyclin complexes are modulated by extracellular signals. This demonstrates a direct link between extracellular signal transduction and the cell cycle (Grana and Reddy 1995, Sherr and Roberts 1999). The successive waves of active cyclin/CDK complexes are tightly controlled by post-translational modifications, intracellular trans-locations and proteolytic degradation of cyclin (Morgan, 1997).

Not all cyclin/CDK complexes are directly involved in the control of the cell cycle. CDK5 is important in normal neuronal functioning (Maccioni *et al.*, 2001). In mammals CDK7 associates with cyclin H and function as a CDK activating kinase since the complex causes phosphorylation of the T-loop Thr 160 of several CDKs (Fisher and Morgan, 1994). CDK7/Cyclin H together with a RING finger protein MAT1 forms a component of the transcription factor TFIIH (Tassan *et al.*, 1995). Strong evidence suggests that CDK8, 9 and 11 are essential in regulating transcription and RNA processing (Loyer *et al.*, 2005).

1.4.1.2 Other kinases involved in the cell cycle: Polo-, Aurora-, and NIMA kinases

Polo-kinase is a critical regulator of multiple mitotic events and conserved from yeast to humans and is also known as Cdc5. It is involved in the regulation of a number of mitotic events including the G₂/M transition, centrosome duplication, bipolar spindle formation, chromosome separation, removal of cohesion complexes from chromosomes, activation of the anaphase promoting complex/cyclosome (APC/C), mitotic exit and cytokinesis (Lee and Amon, 2003). They are localized to centrosomes early in mitosis and accumulate on the central region of the spindle where they persist at the midbody through cytokinesis in addition to locating to kinetochores (reviewed in Ohkura, 2003). Polo-kinase binds to the mitotic regulator Cdc25 that is phosphorylated by the cyclin dependent kinase, CDK1 (Elia *et al.*, 2003). Polo kinases also regulate a number of events that are specific to meiosis (Lee and Amon, 2003).

Three aurora kinases namely A, B and C, have been described in mammals. Aurora kinase A and B are essential for the correct execution of chromosome segregation and completion of the cell division process. They are involved in the centrosome cycle, chromosome condensation, microtubule-kinetochore attachment, cytokinesis and the execution of the spindle checkpoint. Aurora B kinase mediates accurate chromosome segregation ensuring the proper orientation of the sister chromatids. Aurora C is selectively expressed in testicular cells and presumed to function during meiosis. (Reviewed in Andrews *et al.*, 2003, Giet *et al.*, 2005, Meraldi *et al.*, 2004).

1.4.2 The G₁ and S phases of the cell cycle

1.4.2.1 The G₁ phase

The G₁ phase is responsive to extracellular signaling that can either stimulate the cell to enter the proliferative cycle or cause the cell to withdraw from the cell cycle and become terminally differentiated or quiescent. In G₀ and early G₁ high levels of CDK inhibitory proteins (CKI) and low levels of cyclins suppress activity of CDKs as seen in Figure 1.3. In response to external signals cyclin D which normally has a halflife of 10 minutes starts to accumulate and associates with CDK4/6 to form a partially active complex CDK4/6/cyclin D that is activated fully by CDK7/cyclin H. (Grana and Reddy 1995, Sherr and Roberts 1999).

As indicated in Figure 1.3 the retino-blastoma protein (pRB) and its related protein p107 are important substrates for activated cyclin-CDK complexes in the G₁ phase (reviewed in Cobrinik, 2005). During the progression of the G₁ phase pRB is sequentially phosphorylated by CDK4,6/cyclin D and CDK2/cyclin E. Depending on its phosphorylation status pRb can either activate or repress transcription. In a state of low phosphorylation pRb is tightly bound to the transcription factor E2F, inhibiting its activity.

E2F is responsible for the transcription of many genes that are needed to drive the cell through the G₁/S transition and initiate the replication of the DNA complement of a cell. E2F further stimulates the transcription of CDK2/cyclin E. (Malumbres and Barbacid, 2001). In the absence of appropriate mitogenic signals, faulty DNA or anti-proliferative signals, the G₁/S checkpoint is activated and stops the progression into S phase.

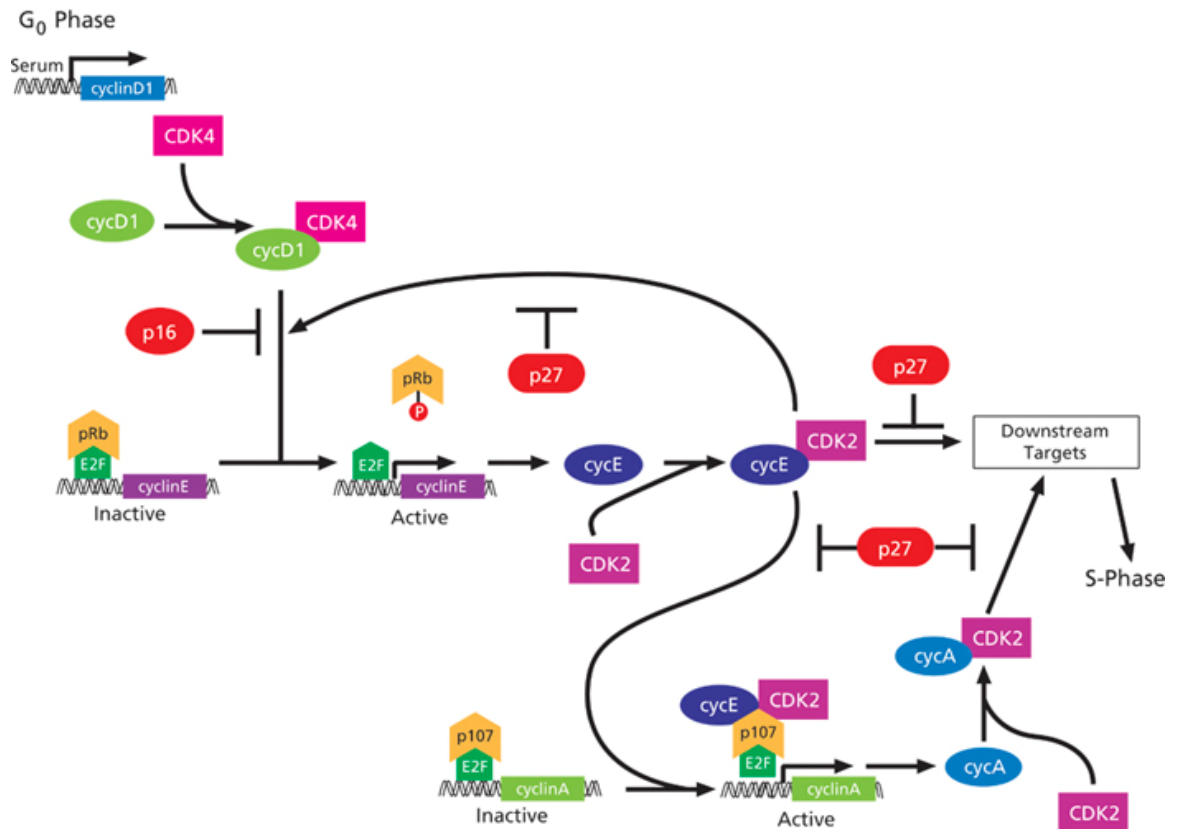


Figure 1.3 Regulatory cascade of CDK/cyclin complexes during the G₀ to S transition of the cell cycle. Cyclin D transcription is stimulated by external signals and it associates with CDK4 in the cell. The CDK4/cyclin D complex causes hyper-phosphorylation of the retinoblastoma protein, which dissociates from the transcription factor E2F which then allows transcription of cyclin E. This can be inhibited by the CDK inhibitor p16. Cyclin E associates with CDK2 to form an active complex which associates with the p107/E2F complex to activate the transcription of cyclin A. The CDK2/cyclin E complex also promotes the phosphorylation of pRb to further stimulate cyclin E transcription. Cyclin A also associates with CDK2 and together with the CDK2/cyclin E complex activates downstream targets to progress to the S-phase and initiate DNA synthesis. The CDK inhibitor p27 inhibits the action of CDK2/cyclin E. pRb = retinoblastoma protein, E2F = transcription factor 2F, p27 = CDK inhibitory protein. (Cell signalling and neuroscience, Sigma-aldrich.com)

1.4.2.2 The G₁-S checkpoint

The G₁ phase cyclin/CDK complexes are pivotal in the G₁-S checkpoint. This checkpoint makes extensive use of CDK/cyclin inhibition by the protein CDK inhibitors of both the Cip/Kip and INK classes and prevents the progression of damaged cells to the S phase. (Sherr and Roberts 1999). When cells are exposed to genotoxic stress the p53 tumour suppressor protein accumulates in the cell and activates the transcription of p21. In turn p21

inactivates CDK4/cyclin D as well as CDK2/cyclin E complexes and inhibits the phosphorylation of pRb as shown in figure 1.4. E2F remains associated with pRb and prevents the synthesis of cyclin E and subsequently cyclin A.

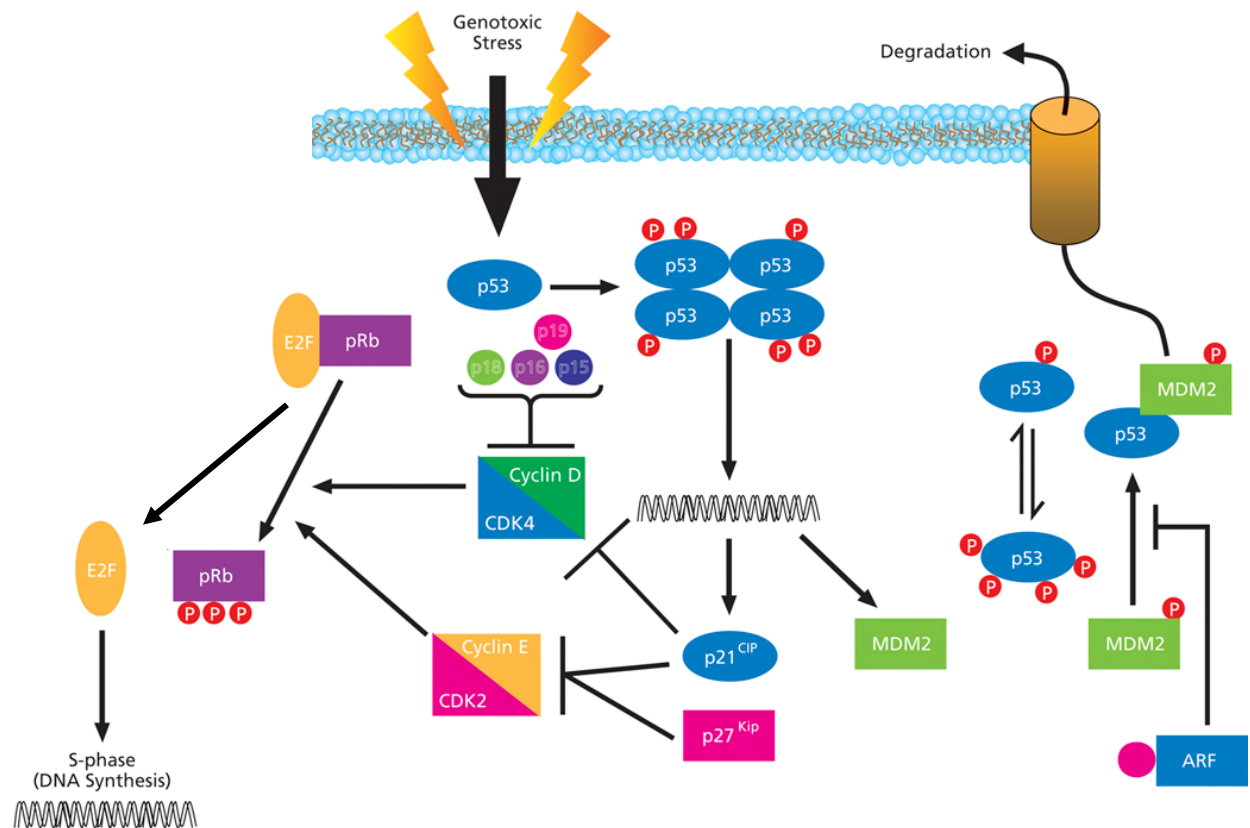


Figure 1.4 The G₁-S checkpoint and the effects of the tumour suppressor p53.

Genotoxic stress induced DNA damage causes accumulation of p53 as well as phosphorylation by ATM (ataxia telangiectasia mutated) kinase. Active p53 causes transcription of the CDK inhibitor p21^{cip}, and the ubiquitin ligase MDM2. Active CDK4/cyclin D and CDK2/cyclin E are inhibited by p21^{cip} while p27^{kip} inhibits the CDK2/cyclin E complex. Inhibition of either cdk/cyclin complexes prevents the hyper-phosphorylation of pRb, the dissociation of E2F and transcription of cyclin D which occurs in the S-phase of the cell cycle. MDM2 associates with p53 and promotes the rapid degradation of p53 by the ubiquitin/protease system and can overcome p53 mediated arrest. ARF prevents MDM2 mediated degradation of p53 and functions to sustain the late G₁ block. CDK inhibitors of the INK4 family (p15, p16, p18, p19) inhibit CDK4/cyclin D independently of p53. Legend: spiral = DNA; pRb = retinoblastoma protein; P= phosphorylation; E2F = transcription factor E2F; p53 = tumour suppressor protein p53; p21, p27 = CDK inhibitory proteins of the CIP class; p14, p16, p18, p19 = CDK inhibitory proteins of the INK subclass, MDM2 = p53 regulatory protein, ARF = inhibitor of MDM2 (Cell signalling and neuroscience, Sigma-Aldrich.com)

p21 also binds CDK1/cyclinB complexes directly and blocks mitotic entry in the G₂ checkpoint as shown in figure 1.5 (Steward and Pietsch, 2001). The p16 inhibitor acts independently of p53 and causes G₁ arrest by inhibiting CDK4/cyclin D and CDK6/cyclin D mediated phosphorylation of pRB (Shapiro *et al.*, 2000). Arrest in G₁ mediated by p21 is unblocked when p53 is exported from the nucleus and inactivated by Mdm2-mediated ubiquitination and proteasomal degradation as shown in figure 1.4 (Oren *et al.*, 2002).

The association of Mdm2 with p53 and its subsequent degradation is prevented by the ARF protein. Activation of TGF- β receptors induces the inhibition of CDK4/cyclin D by p15, while cAMP inhibits the CDK4/cyclin D complex via p27 (see Figure 1.4).

1.4.2.3 The DNA synthesis (S) phase

The entire genome is replicated during the S phase. Replication is initiated simultaneously on multiple sites on the chromosome. It is critical that the chromosomes are copied once only and that a restart of replication is not initiated before cell division is complete. (Kelly and Brown, 2000). Minichromosome maintenance proteins (MCMs) bind at origins of replication and function as helicases ahead of each replication fork. Their displacement from the replication fork during synthesis prevents the restart of replication. (Blow and Hodgson, 2002).

CDK2/cyclin A causes the phosphorylation of many proteins involved in DNA replication and is required for the faithful duplication of the genetic material. The synthesis and accumulation of CDK1/cyclin B also begins during S phase, but the complex is inactive. CDK1 is also referred to as CDC2 after the first CDK characterized in yeast (cell division cycle 2). Cyclin A accumulates during the S phase and its binding to CDK2 marks transition to G₂, a phase characterised by the accumulation of cyclin B/CDK1, the inhibition of DNA replication, cell growth and new protein synthesis. Cyclin A synthesis is controlled by E2F but is negatively regulated by CDK2/cyclin A activity via phosphorylation of an E2F dimerization protein. Cyclin A is destroyed by ubiquitin-mediated proteolysis before metaphase. (Johnson and Walker, 1999; Yam *et al.*, 2002).

1.4.3 The G₂ and M phases of the cell cycle

1.4.3.1 The G₂ phase

Cells at the G₂ phase contain the fully replicated DNA in the form of two sister chromatids. If the DNA complement is not correct the ATM and ATR (ATM and Rad3-related) dependent signals induce cell cycle arrest via inhibition of CDK1 (CDC2). The transition from the G₂ phase to mitosis requires an active CDK1/cyclin B complex. This process is summarized in Figure 1.5.

Activated CDK1/cyclin B is phosphorylated on Thr160 (Thr162 in Figure 1.5) by the CDK7/cyclinH complex and dephosphorylated on Thr14 and Tyr15. The actual amino acid residue number of Thr160 varies slightly according to species. During the G₂ phase, Wee1 and Myt1 phosphorylate CDK1/cyclin B complex and keep it in an inactive state (Smits and Medema 2001; Fisher and Morgan, 1994). CDK7/cyclin H is essential for entry into mitosis and activation of CDK1/cyclin A and CDK1/cyclin B complexes (Larochelle *et al.*, 1998). CDK1/cyclin B activation is triggered by the dephosphorylation of Tyr15 by Cdc25, as shown in Figure 1.5 (Smits and Medema, 2001).

1.4.3.2 The G₂ checkpoint

Genotoxic stress or incorrectly copied DNA causes cell cycle arrest in the G₂ phase. DNA damage activates the ATM and ATR dependent signalling pathways. This in turn activates the checkpoint kinases called Chk1 and Chk2 that phosphorylate Cdc25 and cause the nuclear export and cytoplasmic sequestration of Cdc25 (Abraham 2001). Removal of Cdc25 from the nucleus causes inhibition of CDK1/cyclin B as seen in Figure 1.5 (Abraham, 2001).

Activation of the tumour suppressor p53 up-regulates the transcription of the 14-3-3 σ proteins as well as p21. The 14-3-3 σ protein sequesters Cdc25 to the cytoplasm and p21 inactivates CDK1/cyclin B (Bunz *et al.*, 1998; Flatt *et al.*, 2000; Hermeking *et al.*, 1997; Innocente, 1999, Chan *et al.*, 2000) as shown in Figure 1.5. The inhibitor p21 can also disrupt the interaction between proliferating cell nuclear antigen (PCNA) and Cdc25 to cause G₂ phase arrest (Kawabe *et al.*, 2002).

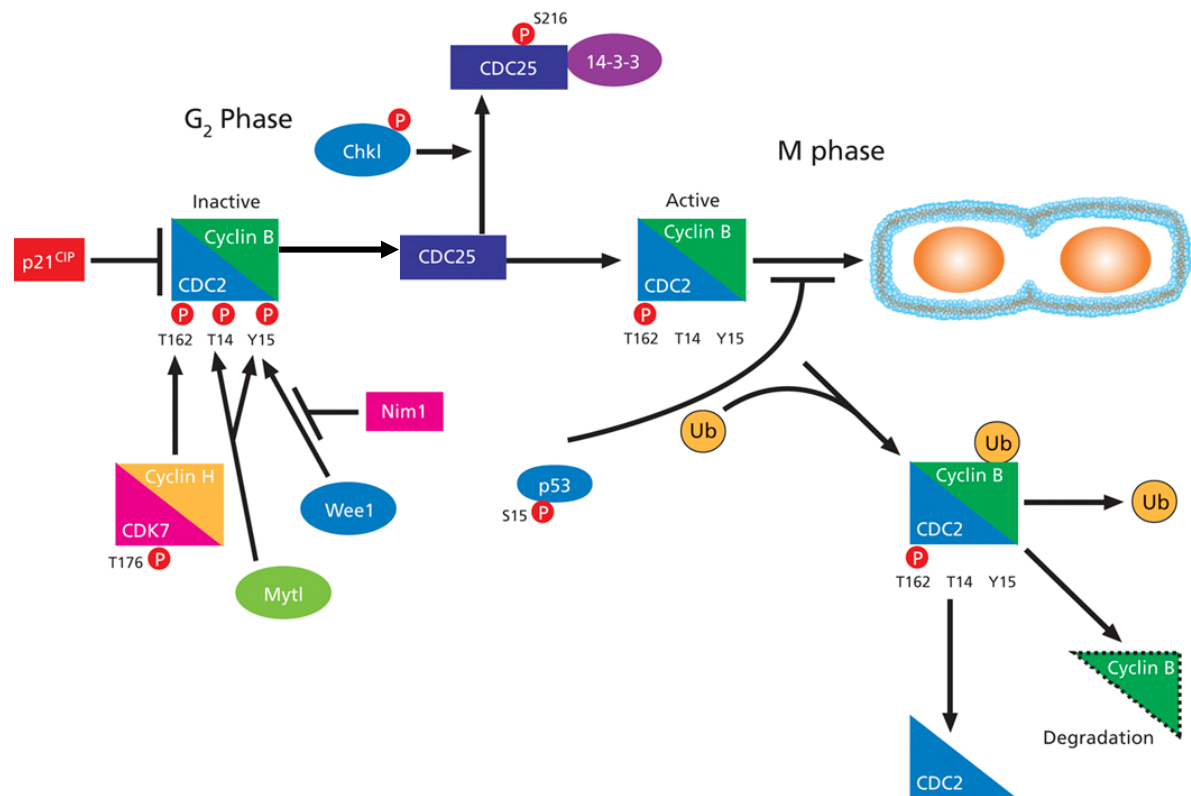


Figure 1.5 Regulation of the G₂ and mitotic phases of the cell cycle. CDC2/cyclin B complex (mitosis promoting factor) controls the transition from G₂ to mitosis. The CDK7/cyclin H complex phosphorylates CDC2/cyclin B at Thr₁₆₂ and the kinases Myt1 and Wee1 phosphorylate CDC2/cyclin B at Thr₁₄ and Tyr₁₅. Wee1 is inhibited by Nim1. In addition, the complex is inhibited by p21^{cip}. The inactive CDC2/cyclin B complex is activated by the phosphatase CDC25 which dephosphorylates Thr₁₄ and Tyr₁₅. In turn the activity of CDC25 is tightly controlled by inhibitory and stimulatory phosphorylations. Phosphorylation by a checkpoint kinase (Chk1) at Ser₂₁₆ causes the inhibition of CDC25 resulting in G₂ arrest of DNA damaged cells. Furthermore the 14-3-3 adapter protein sequesters CDC25 to the cytoplasm away from the CDC2/cyclin B complex. Active CDC2/cyclin B in turn is inhibited by high levels of Ser₁₅ phosphorylated p53 and ubiquitin mediated proteolytic degradation of cyclin B as the cell progresses from metaphase to anaphase during mitosis (M phase). Legend: cdc2 = CDK1; p21 = CDK inhibitory protein, CDC25 = cell division cycle 25 (phosphatase); Chk1 = checkpoint kinase 1; 14-3-3 = protein sequestering CDC25 to cytoplasm; Ub = ubiquitination (Cell signalling and neuroscience, Sigma-aldrich.com).

1.4.3.3 CDKs and the mitotic phase

The active CDK1/cyclin B holoenzyme, also known as mitosis promoting factor (MPF), catalyzes the phosphorylation of lamins and histone 1, and is involved in the regulation of events preceding cell division, such as spindle formation, chromatin condensation, and

fragmentation of the nuclear envelope and of organelles such as the Golgi and endoplasmic reticulum. The metaphase to anaphase transition is triggered by inactivation of the complex due to the ubiquitination and the proteolytic degradation of cyclin B.

The anaphase promoting complex (APC) is an ubiquitin-protein ligase that targets key proteins for proteolysis. This induces the separation of chromatids and their movement to the poles of the mitotic spindle, after which the mitotic apparatus disappears, the nuclear membrane reforms and the nucleoli reappear. During cytokinesis, the cytoplasm divides and the resulting daughter cells enter G₁. (Stewart *et al.*, 2003). Cyclin B is inactivated by ubiquitin mediated proteasomal degradation as shown in Figure 1.5.

The mitotic spindle checkpoint prevents entry into anaphase until kinetochores of every chromatid pair have attached correctly to the spindle microtubules. Chromosome segregation during anaphase is delayed until defects are repaired. (Nigg, 2001; Musacchio and Hardwick, 2002, Scholey *et al.*, 2003).

1.5 Regulation of the cell cycle by the tumour suppressor p53

As seen above, the tumour suppressor protein, p53, plays a central role in cell cycle checkpoint control (Stewart and Pietenpol 2001). Mutated p53 has been found in many tumours. In a normal unstressed cell p53 is inactive with its activity controlled by the murine double minute 2 (Mdm2) protein or the Hdm2 homologue in humans. Activated p53 protein functions as a transcription factor and directly stimulates the transcription of many genes (Michael and Oren 2003). The majority of genes activated by p53 causes transient growth arrest, replicative senescence, terminal differentiation, cell cycle arrest or apoptosis (Chen, 2002). One of the earliest p53 targets to be identified was cyclin G (Okamoto and Beach, 1994; Petrocelli *et al.*, 1996) but its precise role in the p53 pathway and in the cell cycle remains elusive (Okamoto *et al.*, 2002, Ohtsuka *et al.*, 2004).

Numerous studies on p53 activation showed that p53 is pivotal in an extensive network of signal transduction pathways (Harris and Levine, 2005). The activity of p53 requires tight control since it is harmful to normal cell growth and must be deactivated once the exogenous stress has subsided or the DNA repaired (Chen, 2002). Changes in its phosphorylation status are one of the most important control mechanisms for p53 (Oren *et al.*, 2002, Michael and Oren, 2003).

Mdm2 mediates ubiquitination of p53 and targets p53 for proteolytic degradation (Michael and Oren, 2003). Mdm2 activity itself is controlled by phosphorylation via protein kinase B (PKB) that phosphorylates Mdm2 on two sites and enables it to translocate from the cytoplasm to the nucleus where it binds to p53 (reviewed in Oren *et al.*, 2002).

1.6 Cyclin G

Cyclin G has a highly conserved cyclin box and displays similarity to the fission yeast cyclin, Cig1, the B cyclins and human cyclin A and I (Horne *et al.*, 1996). An anomaly of cyclin G is the absence of the PEST sequence (Tamura *et al.*, 1993) that mediates ubiquitin degradation of other cyclins. A specific CDK partner for cyclin G has yet to be identified although it associates with CDK5 and a kinase called cyclin G associated kinase (cGAK). Active CDK5 is only found in neuronal cells and is activated by the p35 protein (Maccioni *et al.*, 2001). Cyclin G does not affect the activity of CDK5 (Kimura *et al.*, 1997). Increased activity of CDK5 is implicated in neuronal death in Alzheimer's disease and cytoskeletal abnormalities (Maccioni *et al.*, 2001).

Cyclin G has been shown to have growth promoting activities and expression is increased in regenerating hepatocytes and motor neurons (Morita *et al.*, 1996). When normal or cancer cells are induced to over-express cyclin G, cell proliferation is stimulated (Smith *et al.*, 1997). Cyclin G is over-expressed in a number of cancers like osteogenic sarcoma, (Skotkzo 1995), breast and prostatic carcinomas (Reimer *et al.*, 1999) and colorectal tumours (Perez *et al.*, 2003). In nude mice cyclin G antisense oligonucleotides inhibit tumour growth *in vitro* (Chen *et al.*, 1997).

The precise role of cyclin G in the cell cycle remains obscure. In doxorubicin treated cells cyclin G expression was increased when the cells arrested during the G₂/M phase (Shimizu *et al.*, 1998). Cyclin G interacts with active PP2A phosphatase as well as Mdm2 and stimulates the ability of PP2A to dephosphorylate Mdm2 at Thr216. The PP2A enzyme dephosphorylates Mdm2 and the human homologue of Mdm2, (Hdm2) (Okamoto *et al.*, 2002). Okamoto postulated that cyclin G functions as a negative regulator of p53 by modulating the phosphorylation status of Mdm2 and Hdm2. The increase in p53 levels in response to DNA damage is inhibited by the over-expression of cyclin G (Ohtsuka *et al.*, 2003).

1.7 Cyclin G associated kinase

In the search for a CDK partner for cyclin G, a serine/threonine protein kinase was identified (Kanaoka *et al.*, 1997) using a rat kidney cDNA library. This kinase could not be identified as a CDK since it lacked the typical PSTAIRE motif of CDKs and was therefore called cyclin G associated kinase (cGAK). cGAK was found to associate with CDK5 (Kanaoka *et al.*, 1997). However, cGAK is active without the presence of cyclin G and is therefore not considered to be a true CDK- association partner for cyclin G (Kanaoka *et al.*, 1997). cGAK displays a typical kinase domain as well as similarity in the central region to tensin and auxilin. A potential tyrosine kinase phosphorylation site, as well as a leucine zipper motif was identified (Kanaoka *et al.*, 1997). Human and rat cGAK have been shown to be similar in amino acid sequences and share all the motifs indicated above (Kimura *et al.*, 1997). Human cGAK is ubiquitously expressed and displays kinase activity in the absence of cyclin G (Kimura *et al.*, 1997).

cGAK is similar to the brain specific protein auxilin involved in uncoating of clathrin coated vesicles in neuronal tissues. Clathrin coated vesicles isolated from the liver and testes contain cGAK protein and it appears to be involved in the assembly of clathrin into baskets (Greener *et al.*, 2000). Umeda *et al.*, (2000) found that cGAK associates with clathrin and phosphorylates the plasma membrane adaptor protein AP-1, and the Golgi adaptor protein AP-2. Receptor mediated endocytosis of transferrin is impaired in cells that over-express cGAK (Umeda *et al.*, 2000). Korolchuk and Banting, 2002, found that cGAK causes the phosphorylation of the $\mu 2$ -subunit of the AP-2 protein in intact clathrin coated vesicles as well as in solution after uncoating has taken place. This phosphorylation is essential for the clathrin mediated endocytosis of transferrin (Olusanya *et al.*, 2001). However, cGAK is not the only kinase able to phosphorylate $\mu 2$ and a more likely candidate was shown to be the adaptor associated kinase 1 (AAK1) protein (Smythe, 2002). A recent study using RNA interference techniques to silence PfcGAK mRNA confirms that cGAK is essential in the early stages of clathrin coated mediated endocytosis (Zhang *et al.*, 2005).

The effect of cGAK on epidermal growth factor receptor (EGFR) mediated signalling was studied since it makes use of clathrin-coated vesicles for receptor trafficking. cGAK was down-regulated by using RNA interference techniques in mammalian cells. The absence of cGAK caused a 50-fold increase of EGFR expression on the cell surface. In cells stimulated

with EGF the EGFR remained associated with the cell surface for an extended period and showed an increase in tyrosine kinase activity. cGAK knockdown cells also showed changes in downstream EGF signalling. There was approximately a 100-fold increase in the abundance of extracellular signal-regulated kinase 5 (ERK5), a kinase that is involved in cell survival and proliferation. The activity of a regulatory subunit of phosphatidylinositol 3-kinase was elevated in cGAK knockdown cells and this caused an increase in active PKB, a kinase that plays a central role the phosphatidyl inositol -3 - kinase (PI3K) pathway. It was also observed that down regulation of cGAK promoted cell proliferation even in the absence of growth factors (Zhang *et al.*, 2004).

cGAK has recently been found to bind to the cAMP response element binding protein (CBP) in a yeast two hybrid screen of a cDNA library derived from mouse embryonic tissue. CBP is a co-activator of transcription factors and required in the transactivation of many genes (Yin *et al.*, 2005). The implications of this interaction are not yet understood.

In a study to determine the cellular targets of the protein kinase inhibitor SB 203580 it was found that the supposedly p38 kinase selective inhibitor inhibited cGAK and casein kinase I activity with IC₅₀ values comparable to that of p38 (Godl *et al.*, 2003). cGAK was shown to interact with the androgen receptor and enhance its activity and may play a role in prostate cancer (Ray *et al.*, 2006). It is clear that the current information on the function of cGAK is incomplete and that this kinase requires further investigation.

1.8 NIMA kinases and cell cycle regulation

The NIMA kinases (never in mitosis in *Aspergillus nidulans*) and their homologues, Nima related kinases (NRKs) play important roles in cell cycle regulation and the successful entry and completion of mitosis. In mammals the kinases homologous to NIMA kinases are called NRKs or Nek proteins. The term Nek protein is currently loosely used for all eukaryotic NIMA related kinases. In humans 11 NRKs (Neks) have been identified (reviewed in O'Connell *et al.*, 2003). These proteins vary in size from 302 - 1258 amino acids with the kinase catalytic domain being the most conserved feature. The non-catalytic regions show very low sequence similarity that may indicate a diversity of function of this kinase family (O'Connell *et al.*, 2003). NIMA and related kinases have a conserved N-terminal catalytic domain followed by a centrally located coiled-coil domain with at least one PEST sequence in the C-terminal region (O'Connell *et al.*, 2003). Phosphorylation of

internal sequences rich in proline, glutamic acid, serine and threonine (PEST sequences) target proteins for rapid degradation via ubiquitination and proteasomal degradation.

1.8.1 NIMA kinases in *Aspergillus nidulans*

In *A. nidulans* NIMA kinase mediates chromatin condensation (Osmani *et al.*, 1988), functions in DNA checkpoint control (Ye and Osmani, 1997) and is involved in the nuclear localization of Cdc2/cylin B (Wu *et al.*, 1998). NIMA is required in conjunction with Cdc2/cylin B to enter mitosis in *A. nidulans* (Ye and Osmani, 1997, Ye *et al.*, 1999). Cells deficient in NIMA arrest in the late G₂ phase. Overexpression of NIMA causes chromosome condensation and the transient appearance of mitotic spindles. Transfection of human cells with *A. nidulans* NIMA caused premature condensation of chromosomes (O'Connell *et al.*, 1994). In *A. nidulans* histone-3 phosphorylation is dependent on the activation of NIMA and correlates with the transient localization of NIMA on chromatin early in mitosis (De Souza *et al.*, 2000).

NIMA activity is regulated at the level of gene transcription (Osmani *et al.*, 1987), phosphorylation and by proteolytic degradation in *A. nidulans* (Pu and Osmani, 1995). NIMA levels start to increase during late interphase and reach the highest levels at the G₂/M transition (Osmani *et al.*, 1991). Davies *et al.*, (2004) suggest that NIMA is required for the regulation of nuclear membrane fission during mitotic exit in *A. nidulans* prior to its degradation.

1.8.2. NIMA kinases in other organisms

In fission yeast the NIMA related kinase, Fin1, induces chromosome condensation and is required for a functional bipolar spindle and nuclear envelope integrity (Krien *et al.*, 2002). Functional Fin1 is also required for the normal association of Polo kinase with spindle pole bodies (Grallert and Hagan, 2002). The KIN3 NIMA related kinase of *Saccharomyces cerevisiae* is more closely related to mammalian Nek2 than to *A. nidulans* NIMA (O'Connell *et al.*, 2003). In *Chlamydomonas* the FA2 gene codes for a NIMA kinase that is essential for basal-body/centriole associated microtubule severing (Mahjoub *et al.*, 2002). A homologue of NIMA of unknown function has been cloned from the protozoan organism *Tetrahymena pyriformis* (Wang *et al.*, 1998). In *Drosophila* a Nek2 homologue is required for the normal functioning of the centrosome (Prigent *et al.*, 2005). Apart from the NIMA

homologue of *Neurospora crassa*, functional complementarity to *A. nidulans* NIMA has not been demonstrated for any of the NIMA related enzymes (reviewed in O'Connell *et al.*, 2003).

1.8.3 Mammalian NIMA related kinases (Neks) and their functions

Nek1 is expressed in meiotic germ cells and is thought to play a role in meiosis (Letwin *et al.*, 1992). Nek1 has an extensive coiled-coil region and a number of PEST sequences located at the C-terminus (O'Connell *et al.*, 2003). Mutations of this kinase gene cause developmental disorders in mice and it is mutated in polycystic kidney disease (Upadhyaya *et al.*, 2000). Nek1 interacts with the phosphatase PP2A and protein 14-3-3, which inhibit cell growth. (Surpili *et al.*, 2003). Nek1 is involved in the DNA damage response pathway and associates with DNA damage response proteins (Polci *et al.*, 2004).

Nek2 is a mammalian protein kinase that is structurally but not functionally most closely related to the NIMA kinase found in *A. nidulans*. It is an important regulator of centrosome function and is required for the assembly and maintenance of the centrosome (Fry *et al.*, 1998) Nek2 kinases, together with CDKs, polo- and aurora kinases, are all involved in centrosome function and regulation (Fry, 2002).

Nek2 remains associated with the centrosome throughout the cell cycle, which suggests that Nek2 is involved in more than just regulation of centrosomal function. (Kim *et al.*, 2002). Recent evidence suggest that in mouse spermatocytes, Nek2 is phosphorylated by the p90Rsk2 kinase that functions in the MAPK signal transduction pathway and that Nek2 is required for chromosome condensation (Di Agostino *et al.*, 2002). Nek2 has also been shown to undergo autophosphorylation that increases its activity (Helps *et al.*, 2000).

Phosphorylation targets of Nek2 have been identified, namely a centrosome associated protein called C-Nap1 (Fry *et al.*, 1998), the Hec1 (highly expressed in cancer) protein (Chen *et al.*, 2002) and more recently the Golgi protein named NIP1 (Yoo *et al.*, 2004). *In vitro*, Nek2 decreases the phosphatase activity of protein phosphatase 1 (PP1) by phosphorylation (Helps *et al.*, 2000). This interaction is thought to be unique for Nek2. Nek2 levels decrease upon mitotic entry as result of proteasomal degradation (Hames *et al.*, 2001).

Apart from its structural similarity mammalian **Nek3** kinase is vastly different from Nek2 and NIMA. There is no direct evidence that Nek3 is involved in the cell cycle (Tanaka and Nigg 1999) although a different study showed that murine Nek3 was preferentially expressed in tissues undergoing mitosis (Chen *et al.*, 1999). Nek3 does not have the conserved coiled coil region C-terminus of the catalytic site but does possess two PEST sequence regions (O'Connell *et al.*, 2003).

Nek4 is highly expressed in germ cells and shows no cell cycle regulated expression. Murine Nek4 shows variation in expression levels in developing testicular germ cells (Chen *et al.*, 1999). Apart from the kinase catalytic domain Nek4 does not share any of the other NIMA kinase conserved features (Belham *et al.*, 2003). To date no Nek5 kinase has been described.

Murine **Nek6** and **Nek7** are highly homologous with 87% identical amino acids in the catalytic domain. They are unusual since their catalytic domains are found at the C-terminus of the protein. Nek6 has been shown to phosphorylate both histone H1 and H3 *in vitro* (Hashimoto *et al.*, 2002). Both Nek6 and Nek7 phosphorylate p70 ribosomal S6 kinase (S6K) *in vitro* at threonine-412 which causes the activation of this kinase. S6K controls a group of mRNAs that encode components of the protein translation machinery (Belham *et al.*, 2001). Nek6 and Nek7 phosphorylate glucocorticoid-induced kinase (SGK) and protein kinase B isoforms (Lizcano *et al.*, 2002).

There is considerable confusion around the nomenclature of **Nek8** and **Nek9**. The human Nek8 related kinase originally described by Holland *et al.*, in 2002 is the same kinase as Nercc1 described by Roig *et al.*, in 2002, which has been renamed Nek9 by Genbank since it was not homologous to murine Nek8. For the purpose of clarity this protein will be referred to as Nercc1/Nek8/Nek9. Nek8 is distinct from this protein and was first described in mice. Nek8 and Nek9 share the presence of a regulator of chromosome condensation domain (RCC1) in the C-terminus (O'Connell *et al.*, 2003).

Murine **Nek8** like Nek1, was found to be mutated in the mouse model of polycystic kidney disease. Antisense oligonucleotides to Nek8 caused the formation of pronephric cysts in zebrafish (Lui *et al.*, 2002). Nek8 has the typical catalytic domain at the N-terminus but no coiled-coil domain (O'Connell *et al.*, 2003). Nek8 was found to be over-expressed in human breast cancer. Over-expression of a mutated kinase domain causes a decrease in

actin protein, which indicates that human Nek8 may be associated with the maintenance of the cytoskeleton (Bowers and Boylan, 2004).

Nek9 (Nercc1/Nek8/Nek9) has an N-terminal kinase domain similar to the Nek family of kinases, a central domain with similarity to RCC1, a guanine nucleotide exchange factor for the GTPase Ran and a C-terminal coiled coil domain. Nercc1/Nek8/Nek9 undergoes oligomerisation and phosphorylates Bcd2 *in vitro*, a microtubule associated protein. Nercc1/Nek8/Nek9 activity is independent of the cell cycle and is marginally higher in G₀ arrested cells (Holland *et al.*, 2002). It binds specifically to the Ran GTPase and has been shown to be a substrate for Cdc2 phosphorylation.

Nercc1/Nek8/ Nek9 regulates chromosome alignment and segregation in mitosis. Over-expression of active and inactive Nercc1/Nek8/Nek9 causes inhibition of cell division and abnormal nuclear morphology (Roig *et al.*, 2002). Nercc1/Nek8/Nek9 has recently been shown to associate with FACT, a chromatin modulator that is important in replication and transcription (Tan and Lee, 2004). Nercc1/Nek8/Nek9 can also phosphorylate a critical site in the activation loop of Nek6 and Nek7 *in vitro* with a resultant increase in activity of both (Belham *et al.*, 2003).

Nek10 is poorly characterized. It is unusual in that the kinase domain is located centrally, with a coiled-coil motif on both sides and a single PEST sequence (O'Connell *et al.*, 2003). **Nek11** functions mainly as a DNA replication/damage stress responsive kinase. Nek11 was activated when cells were treated with DNA damaging agents and replication inhibitors (Noguchi *et al.*, 2000).

1.9 The cell cycle of *Plasmodium falciparum*

Although correlations have been made between the intraerythrocytic phase of the *Plasmodium* life cycle and the cell cycle phases of higher organisms, it is not clear exactly which stages of the erythrocytic parasite cycle correlate with the typical G₀, G₁, S, G₂ and mitotic phases. In the typical eukaryotic cell cycle the DNA complement is copied once only and yields a duplicate set of chromosomes. During schizogony in the parasite, multiple rounds of DNA replication occur and several sets of chromosomes are formed from one parasite, which diverges dramatically from that of the typical eukaryotic cell. (Doerig *et al.*, 2002).

The erythrocytic cycle is normally between 46 and 48 hours long and divided into four stages based on morphological and biochemical criteria. When a merozoite invades an erythrocyte it forms a ring structure that slowly progresses to an early trophozoite. During this stage the parasite is present in a mononucleated form and has low metabolic activity. Digestion of host cytoplasm takes place by endocytosis and protein and nucleic acid synthesis is negligible. This phase normally lasts until around 18 hrs after invasion (Graeser, *et al.*, 1996b). It has been suggested that the ring and early trophozoite forms constitute the G₁ phase of the cell cycle (Leete and Rubin 1996; Doerig *et al.*, 1999). During the trophozoite stage (18-28 hrs post invasion) extensive RNA and protein synthesis takes place and haemozoin crystals are formed. It is generally accepted that the trophozoite and early schizont stage represents the S-phase of the cell cycle. Multiple rounds of DNA replication in the parasite result in a single cell containing 8-32 nuclei. It is not known if there is a short G₂ phase between successive genome replications. (Arnott and Gull 1998; Leete and Rubin, 1996; Doerig *et al.*, 1999; Doerig *et al.*, 2000).

Schizogony is vastly different from mitosis. Some of the atypical features of *Plasmodium* schizogony include the absence of chromosome condensation and the maintenance of the nuclear membrane (cryptomitosis). Spindle protein formation also diverges from the higher eukaryotic cells. A single centriolar plaque gives rise to daughter spindles that form from a hemispindle. The daughter spindles migrate in opposite directions across the nuclear membrane inner face. There is also a variation in the number of nuclei produced from a single schizont with asynchronous division being a common occurrence. The chromatin from each nucleus remains attached to a partially disassembled half spindle that represents an autonomous unit which functions independently of other units. (Read *et al.*, 1993; Read and Hyde, 1996).

When parasites exit schizogony to form gametocytes the cell cycle is arrested. It is not clear at exactly what stage of the erythrocytic development the gametocytes exit the cell cycle. Gametocytes from *P. berghei* contain more DNA than the haploid merozoite, which indicates that the arrest takes place after DNA replication has been initiated. The female gametocytes also contain intra-nuclear spindles that are consistent with cell cycle arrest occurring at a later stage. When male gametocytes reach the mosquito midgut they are activated and undergo rapid DNA replication to produce eight flagellated gametes. (Janse *et al.*, 1988; Sinden *et al.*, 1996).

There is very little evidence that *Plasmodium* employs similar cell cycle checkpoints as are observed for most eukaryotes. Checkpoints in the late G₁ and G₂ phases ensure the correct progression of the cell cycle events during cell division. Experiments designed to study the effect of various compounds on nuclear division cycles in *P. falciparum* did not allow any conclusion as to the existence of checkpoints. Damage to DNA by γ -irradiation of synchronised parasite cultures at the start of the schizont phase did not decrease parasite DNA synthesis despite lethal chromosome damage. (Reviewed in Doerig *et al.*, 1999). In checkpoint controlled cells of higher eukaryotes similar damage causes a halt in DNA synthesis to allow for DNA repair or stimulates the activation of apoptotic pathways.

1.10 Kinases in *Plasmodium falciparum*

The completion of the *P. falciparum* genome project (Gardner *et al.*, 2002) revealed the presence of 65 predicted protein kinase sequences, plus 20 “FIKK” sequences (see below). These predictions were made using an approach based on homology between the active sites of proteins across species. The analysis of the *P. falciparum* kinome by Ward *et al.*, (2004) revealed some unusual features. The most notable feature is the absence of the classical GXGXXG triad in subdomain I of many *P. falciparum* kinases. These residues are normally important in the binding of the ATP molecule during the transfer of the γ -phosphate group of ATP to its substrate. Apart from the relatively low percentage of conservation in subdomain I all the other essential kinase consensus residues are highly conserved. The analysis identified CDK related kinases, NIMA related kinases as well as a putative homologue of cGAK. The other important fact is that several *P. falciparum* kinase sequences do not fit with any of the known kinase families. (Ward *et al.*, 2004).

The identification of a unique family of 20 putative kinases given the name FIKK on the basis of a conserved amino acid motif may be important in the design of *P. falciparum* selective kinase inhibitors, although to date no kinase activity has been demonstrated for any of the FIKK kinases (Ward *et al.*, 2004). All the essential residues required for kinase activity (blue residues in Figure 1.2) are conserved in the FIKK kinases. A unique feature of the putative FIKK kinases is the presence of a *Plasmodium* export signature known as the PEXEL (*Plasmodium* export element) motif, which occurs downstream of a signal/anchor sequence close to the amino terminus of the protein. This group of kinases are also referred to as the R45-FIKK kinases (Schneider and Mercereau-Puijalon, 2005).

1.10.1. *P. falciparum* cell cycle kinases and cyclins

Since the cell cycle and its regulation in *Plasmodium* are poorly understood, the fact that the kinase catalytic domain is highly conserved among species, was used to identify *Plasmodium* kinases, which may be involved in the cell cycle. Prior to the completion of the *Plasmodium* genome, homology based approaches made extensive use of degenerate primers to identify selected kinase homologues.

1.10.1.1 *P. falciparum* protein kinase 5 (PfPK5)

Plasmodium falciparum protein kinase 5 (PfPK5) was the first *P. falciparum* CDK-like kinase to be identified and has structural similarity to both CDK1 and CDK5. It has a PSTTIRE cyclin binding motif that is closely related to the PSTAIRE motif normally found in CDKs. (Ross-MacDonald *et al.*, 1994). The kinase has regulatory phosphorylation sites at similar positions to the Thr14, Tyr15, Thr160 of CDKs. PfPK5 has 58% sequence identity and > 65% sequence similarity with CDKs. It is expressed in the asexual stages with maximal expression at the schizont stage. (Graeser *et al.*, 1996 a,b). Monomeric PfPK5 shows low levels of histone H1 kinase activity, a property that is not usually observed in CDKs.

PfPK5 is unique in that its kinase activity is increased by a number of different cyclin molecules. Human cyclin H (a CDK7 cyclin), p25 (a CDK5 cyclin), and Pfcyc-1 (a *P. falciparum* cyclin) all increase the activity of PfPK5 (Le Roch *et al.*, 2000). In a separate study bovine cyclin A (CDK1/CDK2 cyclin) also increased PfPK5 activity (Holton *et al.*, 2003). Unlike other eukaryotic CDKs, PfPK5 undergoes autophosphorylation in the presence of a cyclin (Doerig *et al.*, 2002). PfPK5 activity stimulated by Pfcyc1 was inhibited by the human CDK inhibitory protein of the cip/kip subclass (p21^{cip1}) but not by the INK- class (p16^{INK4}) (Li *et al.*, 2001).

In a more recent study PfPK5 was activated by the newly identified Pfcyc-3 at sub-picogram levels. This indicates that Pfcyc-3 may be the true *in vivo* cyclin of PfPK5 (Merckx *et al.*, 2003). Considering the combined information on PfPK5 it is not unlikely that it is the true *P. falciparum* homologue of CDK1, although it also shows similarity to neuronal CDK5. A recent study has developed a three-dimensional model of PfPK5 with the aim to define the structural and spatial prerequisites for the binding of small molecule

inhibitors to the ATP binding sites. This may guide the design of PfPK5 selective inhibitors (Keenan and Welsh, 2004).

1.10.1.2 *P. falciparum* CDK-related kinase 1 (Pfcrk-1)

The second *P. falciparum* CDK-related kinase (Pfcrk-1) was identified using a degenerate PCR primer approach. Pfcrk-1 has 46% of the p58-GTA -‘PITSLRE’ family of enzymes (Doerig *et al.*, 1995), which belongs to the CDK11 subgroup. The similarity is restricted to the catalytic domain. The CDK11 group of kinases functions in the regulation of transcription and mRNA processing. The CDK11^{p58} subgroup is thought to be an important regulator of mitosis (Loyer *et al.*, 2005). The cyclin partner for CDK11 is cyclin L, previously known as Ania - 6a and b proteins (Loyer *et al.*, 2005). Pfcrk-1 has an AMTSLRE cyclin binding motif but to date no *in vitro* kinase activity has been demonstrated for this enzyme. Pfcrk-1 mRNA has been detected in gametocytes but not in asexual stages. At the N-terminal region Pfcrk-1 has a unique 370 amino acid region which is rich in charged amino acids; lysine, arginine, aspartic acid and glutamine. This region also includes seven degenerate repeats of a 10 mer motif (HNNNNHHTDQ) of unknown function (Doerig *et al.*, 1995).

1.10.1.3 *P. falciparum* CDK-related kinase 3 and 4 (Pfcrk-3, Pfcrk-4)

Pfcrk-3 and Pfcrk-4 were identified *in silico* utilizing *Plasmodium* genome sequence data. Pfcrk-3 is a putative kinase encoding a protein with similarity to CDK1 and is expressed in asexual and sexual erythrocytic stages. The typical activation residues, Thr14 of CDK1, is not conserved in Pfcrk3 although Tyr15 and Thr160 are present. Thr14 is replaced by an alanine residue in Pfcrk-3 and it has the sequence AKTYIRE as a cyclin binding domain (Doerig 2002). Pfcrk-4 is similar to Pfcrk-3 and shows similarity to both CDKs and mitogen activated protein kinases (MAPKs). Pfcrk-4 has the poorly conserved cyclin binding sequence EEFAVNE. Thr14 is substituted by a valine residue in Pfcrk-4, but a Thr160 residue is present. Asparagine rich N-terminal extensions of approximately 100 amino acids containing repeated motifs are found in Pfcrk-3 and Pfcrk-4. This is a common feature of many *Plasmodium* proteins and its significance is currently poorly understood. (Doerig *et al.*, 2002).

1.10.1.4 *P. falciparum* MO15 related kinase (Pfmrk)

Pfmrk (MO15-related kinase) has maximal identity (47%) to human CDK7 and 36-43% identity to other CDKs (MO15 is another name for CDK7). Pfmrk has a NFVLLRE cyclin binding motif, which is similar to the NRTALRE motif found in human CDK7. CDK7 combined with its cyclin partner cyclin H, is the main CDK activating kinase (CAK) in eukaryotic cells. Active CDK7 phosphorylates CDK on an equivalent Thr160 residue. CDK activating kinases display short insertions before their C-helices which is important for cyclin binding; Pfmrk has a similar insertion at this site. (Li *et al.*, 1996).

Pfmrk, measured as mRNA, is expressed at higher levels in gametocytes than in the asexual stages (Li *et al.*, 1996). *E. coli*- expressed Pfmrk displayed weak activity towards histone H1 but activity was enhanced when human cyclin H was added to the kinase reaction. This increased activity was unaffected by the addition of the CDK inhibitors olomoucine and roscovitine to a concentration of 100 μ M (Waters *et al.*, 2000).

The *P. falciparum* cyclin, Pfcyc-1, only moderately increased the activity of Pfmrk (Le Roch *et al.*, 2000). The carboxyl terminus of RNA polymerase II was a more efficient substrate for Pfmrk than histone H1 protein. Pfmrk/Pfcyc-1 activity was inhibited by the human protein CDK inhibitor, p21^{Cip} but not by p16^{INK4}. Compared to PfPK5, Pfmrk/Pfcyc-1 was more sensitive to p21^{Cip} inhibition than PfPK5. (Li *et al.*, 2001). Pfmrk-cyclin complexes tested *in vitro* for CDK activating activity against PfPK5 showed no increase in PfPK5 kinase activity (Le Roch *et al.*, 2000). This may be due to the fact the Pfmrk is not a true CAK or that additional regulators are required for activity.

1.10.1.5 *P. falciparum* protein kinase 6 (PfPK6)

PfPK6 has similarity to the catalytic domain of CDK1 as well as to mitogen activated protein kinases (MAPKs), with a SKCILRE cyclin binding motif. The regulatory phosphorylation sites found in CDKs (Thr14, Tyr15 and Thr160) are conserved in PfPK6. *In vitro* PfPK6 displays cyclin independent kinase activity towards histone H1 and also undergoes autophosphorylation. PfPK6 is unusual in that it prefers Mn²⁺ to Mg²⁺ as a divalent cation, which is a characteristic of tyrosine kinases. (Bracchi-Richard *et al.*, 2000). PfPK6 protein expression peaks during the trophozoite and schizont stages and the protein is located in both the nucleus and the cytoplasm (Bracchi-Richard *et al.*, 2000). PfPK6

interacts with elongation factor-1 α that is involved in protein biosynthesis (Bracchi-Richard and Chakrabarti, unpublished, reported in Doerig, 2002).

Molecular modelling of the kinase suggests that it is more closely related to active human CDK2 than to inactive human CDK2 (Doerig *et al.*, 2002). This is consistent with the fact that kinase activity is not enhanced by the presence of cyclins. The human protein CDK inhibitors, P16^{INK4} and p21^{cip1} do not inhibit PfPK6 kinase activity (Li *et al.*, 2001). PfPK6 is sensitive to the CDK inhibitor roscovitine in the micromolar range (IC₅₀: 30 μ M) (Bracchi-Richard *et al.*, 2000), which indicates that roscovitine is not an efficient inhibitor of this enzyme.

Data obtained from mutational analysis of PfPK6 suggests that autophosphorylation of Thr178 is responsible for the maximal activation of the enzyme by enhancing the interaction with its substrate. Mutation of the Thr178 to Ala178 results in a loss of both substrate and autophosphorylation activity (Bracchi-Richard *et al.*, 2000).

1.10.1.6 *P. falciparum* cyclins

The identification of the putative *P. falciparum* cyclin dependent kinases stimulated a search for their putative cyclin partners. Compared to their mammalian counterparts, *P. falciparum* cyclins are poorly conserved, which explains why homology based PCR based approaches followed in many prominent laboratories failed to identify them. This was changed by the availability of *P. falciparum* genomic sequence data and to date four *P. falciparum* cyclins have been characterised. PfPK5, the putative malaria homologue to CDK1 was used extensively to characterize the four malaria cyclins.

The first *Plasmodium* cyclin to be identified was Pfcyc-1. Pfcyc-1 has a low level of similarity to mammalian cyclins (17%) but has been shown to efficiently enhance the activity of PfPK5 as described in section 1.10.1.1 (Le Roch *et al.*, 2000). Pfcyc-1 is most closely related to cyclin H with 20% identical amino acids. Pfcyc-1, as well as mammalian cyclin H and p25 (the activator of CDK5) activates PfPK5 *in vitro*. In the presence of Pfcyc-1, but not Pfcyc-3, PfPK5 underwent autophosphorylation (Merckx *et al.*, 2003). Pfcyc-1 also activates the kinase Pfmrk, the *Plasmodium* CDK with maximal homology to CDK7.

The characterization of the three additional *Plasmodium* cyclins (Pfcyc-2, -3 and -4) is described in Merckx *et al.*, 2003. The Pfcyc-2 gene consists of two exons and codes for a protein of 2281 amino acids with an expected molecular weight of 273.5 kDa. In contrast the other cyclin genes, Pfcyc-1, Pfcyc-3 and Pfcyc-4 code for proteins consisting of 327, 229, and 226 amino acids respectively. Their predicted molecular weights are 39.2, 26.8 and 31.1 kDa respectively. Pfcyc-2 has a maximal 13% identity with plant mitotic cyclins, Pfcyc-3 has a maximal identity of 26% to *T. brucei* CYC2 and Pfcyc-4 shares 33% identity to ania-type cyclins. The genes for Pfcyc-1 and Pfcyc-3 consist of one exon only whereas the gene for Pfcyc-2 consists of 5 exons.

Pfcyc-2 and Pfcyc-3 mRNA was present in both the asexual and sexual erythrocytic stages of the *P. falciparum* 3D7 strain. In contrast Pfcyc-1 mRNA was much more abundant in gametocyte RNA than in RNA obtained from the asexual erythrocytic stage although the Pfcyc-1 protein was detected in the erythrocytic stage of the parasite. Pfcyc-4 was expressed in the erythrocytic stages of the parasite. Analysis of protein expression of the cyclins by using antibodies showed that the expression of Pfcyc-1, Pfcyc-2 and Pfcyc-4 varied during the asexual cycle. Pfcyc-1 expression was more abundant in the trophozoite stage, Pfcyc-2 expression in the schizont stage and Pfcyc-4 expression in the schizont and ring stages. (Merckx *et al.*, 2003). However, the data on Pfcyc-2 protein expression are not convincing since the Western blot did not show a band corresponding to the predicted size of 273.5 kDa.

When recombinant cyclins were tested against PfPK5, Pfcyc-3, at an amount of less than 1 picogram in the 30 µl kinase assay, stimulated PfPK5 kinase activity whereas Pfcyc-2 and Pfcyc-4 failed to have any effect on PfPK5 (Merckx *et al.*, 2003). Pull down experiments with parasite lysates and recombinant cyclins immobilized on GST beads, followed by kinase assays showed increased phosphorylation activity of Pfcyc-1 and Pfcyc-4 towards histone H1 protein. In contrast, kinase assays using immunoprecipitated cyclins showed kinase activity towards histone protein with all four cyclins. (Merckx *et al.*, 2003)

1.10.2 *P. falciparum* NIMA related kinases

As indicated in section 1.8 NIMA related kinases play important roles during mitosis. The *P. falciparum* genome encodes four NIMA related kinases, which have been called Neks. The first *Plasmodium* Nek to be described was Pfnek-1, which displays maximal similarity

to mammalian Nek2 (Dorin *et al.*, 2001). Pfnek-1 is expressed both in the asexual and sexual erythrocytic stage of parasite development. Three other NIMA related kinases have been predicted in the *Plasmodium* genome (Doerig and Chakrabarti 2004).

In vitro kinase assays showed that Pfnek-1 preferred β -casein and myelin basic protein as substrates compared to α -casein and histone H1 that were poor substrates. Pfnek-1 was able to phosphorylate Pfmap-2, an atypical *P. falciparum* MAPK (mitogen activated protein kinase) *in vitro*. When the two enzymes were incubated together in a kinase assay, a synergistic phosphorylation of exogenous substrate was observed. It is not known if this was an artefact of the *in vitro* experimental design or if it is a true *in vivo* association. However, this was only observed for Pfmap-2 and not with any of the other MAPKs. In many NIMA kinases a phosphorylation motif, FXXT, is found. In Pfnek-1 this motif has been replaced by a SMAHS motif, which is more closely related to a MEK activation site. It is therefore possible that Pfnek-1 functions in a *Plasmodium* MAPK pathway (Dorin *et al.*, 2001). MAP kinase relays signals from a number of different stimuli and receptors to the cell nucleus. Active MAP kinase can cross the nuclear membrane and cause the phosphorylation of transcription factors, thereby regulating gene transcription (Pollard and Earnshaw, 2004: p 455).

Pfnek-4 has been characterised by Reininger *et al.*, 2005. It displays maximal similarity to the catalytic domain of human NEK4 (43% identity). Pfnek-4 messenger RNA as well as protein were detected mainly in gametocytes. The coding region of the Pfnek-4 gene is composed of 7 exons. The recombinant enzyme was able to phosphorylate α - and β - casein but not histone 1 or MBP in *in vitro* kinase assays. Mutation of Thr163, which is a target for activating phosphorylation in NIMA kinases, decreased the activity of the recombinant enzyme. The rodent *Plasmodium* parasite *P. bergeri* has an almost identical gene called Pbnek-4 and is more amenable to gene knockout investigations than *P. falciparum*. Using this approach it was demonstrated that Pbnek-4 was essential for the transformation of zygotes to ookinetes and so blocked transmission to the mosquito. Further analysis using knockout mutants showed that Pbnek-4 is specifically required for the replication of the diploid genome of the zygote prior to meiosis. (Reininger *et al.*, 2005)

1.10.3 *P. falciparum* kinases involved in signal transduction pathways

Cyclic adenosine mono phosphate (cAMP) is a ubiquitous second messenger in eukaryotic cells. G-protein coupled surface receptors activate membrane-bound adenylate cyclase to form cAMP which in turn activates protein kinase A (PKA), a key element of this pathway. PKA classically consists of more than one subunit, normally two catalytic and two regulatory subunits. Binding of cAMP to the regulatory subunits decreases their affinity for the catalytic units and releases them to take part in phosphorylation reactions (reviewed in Skalhegg and Tasken, 2000).

The presence of *Plasmodium* adenylate cyclase activity has been reported as early as 1991, and changes in cAMP levels affect differentiation and growth in the parasite (Read and Mikkelsen, 1991). An open reading frame (ORF) with high similarity to the catalytic subunit of PKA family members (46-88% identity) has been cloned from *P. falciparum* and called PfPKAc by two independent groups (Li and Cox, 2000(a), Syin *et al.*, 2001). PfPKAc is expressed in the asexual stage of the erythrocytic cycle with the highest level in the schizont stage (Syin *et al.*, 2001). PfPKAc has most of the essential amino acid residues required for the interaction with regulatory subunits and protein kinase inhibitors with the exception of the FxxF motif and the conserved Trp196 and Lys213 residues (Li and Cox, 2000(a)). A gene encoding the regulatory subunit (PKAr) has been identified and characterised (Merckx and Doerig, unpublished, reported in Doerig and Chakrabarti 2004). An inhibitor of protein kinase A, H89, inhibited parasite proliferation with an IC₅₀ of 3 µM (Syin *et al.*, 2001)

The cyclic GMP pathway is an important signal transduction pathway responsible for a number of physiological effects mediated by the second messenger cGMP. cGMP-dependent protein kinases form an integral part of the post-second messenger effects (Deng and Baker, 2002). There is ample evidence that this pathway functions in *P. falciparum* since two guanyl cyclases have been identified (Carucci *et al.*, 2000). A *P. falciparum* homologue of cGMP-dependent protein kinases (**PfPKG**) has also been identified (Deng and Baker, 2002). The protein is divergent from its mammalian counterpart since it features four cGMP binding sites instead of two. It is expressed predominantly in the ring stage of the parasite, undergoes autophosphorylation and displays *in vitro* kinase activity with histone H1 as substrate (Deng and Baker, 2002). cGMP at a Ka of 0.17 µM stimulated the kinase activity of PfPKG whereas there was no effect with cAMP. Staurosporine, a non-

selective inhibitor of protein kinases inhibited PfPKG with an IC₅₀ of 10nM (Deng and Baker, 2002). An analysis of the putative cGMP binding sites showed that the C-terminus binding site is the most important for PfPKG activation by cGMP and that the degenerate fourth site is important for the maximal activation of the enzyme but unlikely to bind cGMP (Deng *et al.*, 2003). The cGMP pathway appears to be involved in gametogenesis in *Plasmodium* (Kawamoto *et al.*, 1993).

Two *Plasmodium* mitogen activated protein kinases (MAPKs) have been characterised. MAPKs are also known as extracellular signal regulated kinases (ERKs) and are ubiquitous in eukaryotic organisms. The activation of the mitogen activated protein kinase pathway results in the phosphorylation of transcription factors that regulate gene expression. One of the many diverse targets is the transcription of the D-type cyclins required for the progression of the cell cycle (Pollard and Earnshaw, 2004: p 690). Using a degenerate primer approach, Pfmap-1 was identified simultaneously by Doerig *et al.*, 1996 and Lin *et al.*, 1996. The catalytic domain of **Pfmap-1** has the common TXY motif that acts as an activation site in most MAPKs and the conserved VATRWYRAPE sequence in subdomain VIII of the kinase catalytic domain (see Fig.1.2). The C-terminal extension is large and contains a charged domain with two repetitive sequences of either a tetra- or octapeptide that are not shared by other members of the MAPK family. Pfmap-1 mRNA was detected in gametocytes and gametes/zygotes which make it likely that the kinase play a role in the sexual differentiation of the parasite (Lin *et al.*, 1996).

Pfmap-2, the second MAPK homologue has maximal similarity to MAPKs from various organisms. The recombinant enzyme displays a number of functional characteristics that strengthen the argument that it is a true MAPK. In kinase assays Pfmap-2 undergoes autophosphorylation and is able to phosphorylate the classical MAPK substrate, myelin basic protein (MBP). In the presence of a MAPK specific phosphatase the kinase activity of Pfmap-2 towards MBP was decreased. Pfmap-2 is expressed in gametocytes and its kinase activity was stimulated seven fold after incubation with gametocyte extracts. Pfmap-2 diverges from the MAPKs since the conserved TXY activation site is replaced by a TSH site. Mutational analysis showed that the threonine residue is essential for kinase activity towards (MBP) and that the histidine residue is required for autophosphorylation and MBP phosphorylation (Dorin *et al.*, 1999).

In higher eukaryotes MAPKs are activated by MAPK kinases (MAPKK). The *Plasmodium* genome database does not appear to have a gene that is highly homologous to MAPKKs. The closest relative is **PfPK7**, which displays 34% identity to vertebrate MAPKKs. The similarity is restricted to the C-terminal region of the kinase domain, with the N-terminal region closer to the fungal protein kinase A enzymes. PfPK7 is expressed during all the stages of the erythrocytic cycle, gametocytes, as well as sporozoites and infected hepatocytes. Recombinant PfPK7 undergoes autophosphorylation and phosphorylates the exogenous kinase substrates MBP and histone H2 *in vitro*. However no kinase activity was displayed towards Pflmap-1 and Pflmap-2 although the enzyme may require an unknown activator that was absent in the *in vitro* kinase assay. (Dorin *et al.*, 2005). However, the authors concluded that PfPK7 is most likely not a MAPKK since it lacks the typical MEK activation site (SVAKT or SMANS in mammals and SMGQRDT in plants). PfPK7 activity was not inhibited by the mammalian MEK2 inhibitor U0126, and was poorly inhibited by hymenialdisine and staurosporine (Dorin *et al.*, 2005). This study provides evidence that the three-component MAPK module does not function in malaria parasites. The fact that PfMap-2 was phosphorylated by Pfnek-1 strengthens this argument (Dorin *et al.*, 2001).

PfPKB is a homologue of protein kinase B (also termed Akt/PKB) that forms part of the phosphatidylinositol 3 - kinase (PI3K) signal transduction pathway that is found in many higher eukaryotes. In higher organisms, activation of cell surface receptors produces second messengers that activate PI3K. PI3K generates phosphorylated phosphatidylinositides that bind to the pleckstrin homology domain of PKB. The phosphorylated inositides also activate phosphoinositide-dependent kinase (PDK), which phosphorylates PKB. Activated PKB causes the phosphorylation of a wide variety of proteins including MDM2, mTOR, GSK3 β , transcription factors and proteins involved in the regulation of apoptosis. PKB activation promotes cell survival (reviewed in Vivanco and Sawyers, 2002).

The *Plasmodium* PKB homologue shares high sequence similarity with regards to the kinase catalytic domain but poor similarity outside the catalytic domain (Kumar *et al.*, 2004). The PfPKB catalytic domain also displays high similarity to the catalytic domain of protein kinase C and was inhibited by a protein kinase C selective inhibitor. The PKC inhibitor Go6983 inhibited the kinase activity of the enzyme with an IC₅₀ of approximately 1 μ M and also inhibited parasite proliferation in the schizont phase. Furthermore, human PDK1, an activator of PKB, was not able to phosphorylate or activate PfPKB. PfPKB was able to phosphorylate histone protein and Crosstide, a synthetic protein kinase peptide

substrate, and underwent autophosphorylation. PfPKB has a signal sequence at its N-terminus, which appears to downregulate the activity of recombinant PfPKB by inhibiting its ability to undergo autophosphorylation. PfPKB is expressed in schizonts and merozoites and PfPKB immuno-precipitated from schizont lysates was able to phosphorylate Crosstide. (Kumar *et al.*, 2004).

Although a PI3K homologous gene sequence exists in the *P. falciparum* genomic database, and has been verified in our laboratory (Mtombeni & Harmse, unpublished) it is not known if this signal transduction pathway functions in a similar manner in the parasite since there is no homologue to PDK (www.PlasmoDB.org). A critical feature of PKB enzyme function is the presence of a pleckstrin homology domain at the N-terminus, which provides a binding site for phosphatidyl-inositol-3-phosphate and causes translocation of PKB to the inner face of the cell membrane. This domain is absent in the *P. falciparum* homologue (Kumar *et al.*, 2004). No other kinase with high homology to protein kinase C is predicted in the *Plasmodium* kinome and it is therefore possible that PfPKB fulfils the function of protein kinase C in *P. falciparum*.

A homologue of the PKB target, **glycogen synthase kinase-3 (GSK-3)**, has been characterised in *P. falciparum* (Droucheau *et al.*, 2004). In higher organisms GSK-3 is involved in many diverse cellular processes including the insulin signalling pathway to regulation of transcription factors. It is also one of the downstream substrates of PKB (Vivanco and Sawyers 2002). GSK-3 has lately become a potential drug target in diseases like cancer due to its involvement in many cellular processes.

The *P. falciparum* homologue, PfGSK-3, has a well-conserved catalytic site and an unusual N-terminal domain extension of unknown function. The domains required for substrate binding of human GSK-3 are highly conserved suggesting similar mechanisms of substrate recognition. Recombinant PfGSK-3 phosphorylated GSK specific substrates, glycogen synthase, recombinant axin and the microtubule binding protein Tau. Axin is a scaffold protein found in the proteolytic complex with GSK-3 and the tumour suppressor protein APC (adenomatous polyposis coli) that degrades β -catenin (Alberts *et al.*, 2002, p 895-6). In *Plasmodium* PfGSK-3 mRNA was present during all stages of the erythrocytic cycle but the highest protein levels were found in the early trophozoite stage (Droucheau *et al.*, 2004). PfGSK-3 was exported from the parasite to the red blood cell where it associated

with cytoplasmic membranous structures suggesting a function in protein trafficking (Droucheau *et al.*, 2004).

1.10.4 Other *Plasmodium falciparum* kinases

1.10.4.1. Calcium dependent kinases in *P. falciparum*

Calcium dependent protein kinases (CDPKs) have only been identified in plants and protists, therefore they represent a unusual *P. falciparum* selective target. CDPKs consist of a catalytic Ser/Thr kinase domain, followed by a linker region and four calcium binding EF-hands and are dependent on Ca^{2+} for catalytic activity. CDPKs are thought to be the functional analogues of protein kinase C in plants. PfCDPK1 was the first CDPK isolated from *P. falciparum* and it phosphorylates erythrocyte membrane proteins as well as MBP, casein and histone protein in *in vitro* kinase assays (Zhao *et al.*, 1994).

A second *Plasmodium* CDPK (PfCDPK2) was identified and cloned by Farber *et al.*, in 1997. In the presence of Ca^{2+} the recombinant PfCDPK2 underwent autophosphorylation and phosphorylated MBP as an exogenous substrate (Farber *et al.*, 1997). Li *et al.*, (2000b) identified a gene encoding a third CDPK, displaying all the conserved CDPK features. The N-terminal region is rich in serine, threonine and lysine and contains a number of consensus phosphorylation sites for a range of protein kinases. PfCDPK3 was shown to be expressed in the sexual stage of the erythrocytic cycle (Li *et al.*, 2000b). A fourth CDPK was identified by Billker *et al.*, 2004, and found to regulate cell cycle progression in the male gametocyte. CDPK4 is essential for sexual reproduction and mosquito transmission of *P. bergeri*. CDPK4 regulates exflagellation in gametocytes induced by either xanthurenic acid or a rise in extracellular pH. CDPK4 converts the increase in cytosolic Ca^{2+} into a cellular response required for the differentiation of gametocytes to gametes (Billker *et al.*, 2004).

Recent information regarding PfCDPK1 suggests involvement in multiple signal transduction pathways since the protein is located in various subcellular locations. A consensus sequence for myristoylation, a palmitoylation site and a basic motif is found close to the N-terminus and is presumably involved in anchoring the protein to membranes. Endogenous as well as *in vitro* translated PfCDPK1 is myristoylated *in vivo* and the protein has been shown to be anchored to membranes. (Möskes *et al.*, 2004). Immunogold electron microscopy shows that the protein is found in the parasite vacuole and the tubovesicular system of the parasite. Dual acylation of the basic cluster is required for the correct

targeting of the protein to the parasite vacuole since mutations prevented its association with the vacuole (Möskes *et al.*, 2004).

1.10.4.2 *P. falciparum* casein kinases

Casein kinase activity of parasite origin has been reported to phosphorylate proteins located at the infected erythrocyte membrane (Chishti *et al.*, 1994). Two homologues of casein kinase have been identified and characterised in *P. falciparum*. PfCK1 has maximal activity in the ring and trophozoite stages of the parasite and is small in size compared to other casein kinases (Barik *et al.*, 1997). Recombinant PfCK1 has a high affinity for the CDK inhibitor purvalonol B (Knockaert *et al.*, 2000). Expression of *Plasmodium* casein kinase 2, PfCK2, peaks during the schizont stage and prefers ATP over GTP as a phosphate donor (Bracchi-Richard, unpublished, reported in Doerig and Chakrabarti 2004).

1.10.4.3 Metabolic kinases

Plasmodium enzymes involved in the transfer of phosphate groups to metabolites in the glycolytic pathway have also been described and include phosphoenolpyruvate carboxy-kinase, 3-phosphoglycerate kinase and pyruvate kinase. These kinases do not phosphorylate other proteins but rather transfer phosphate groups from ATP to metabolic intermediates.

The gene encoding **phosphoenolpyruvate carboxykinase (PEPCK)** was identified in *P. falciparum* and cloned. PfPEPCK is ATP dependent and more closely related to plant and yeast PEPCK than to human PEPCK. Expression is increased during the extracellular zygote phase in the mosquito midgut. PEPCK catalyses the reversible formation of oxalacetate and ATP from phosphoenolpyruvate and carbon dioxide and may represent a metabolic switch in carbohydrate metabolism towards gluconeogenesis in the glucose-poor medium of the mosquito midgut (Hayward, 2000).

The enzyme **3-phosphoglycerate kinase** is important in the production of ATP from the glycolytic pathway. The *Plasmodium* counterpart of this enzyme has been identified, cloned and expressed in *E. coli*. The recombinant enzyme is active and was inhibited by suramin with an IC₅₀ value of 7µM (Pal *et al.*, 2004).

Chan and Sim (2005) cloned and expressed *P. falciparum* **pyruvate kinase**, an enzyme important in the glycolytic pathway. The recombinant enzyme is active. Pyruvate kinase

exists as a homotetramer and transfers a phosphate group from phosphoenolpyruvate to ADP to yield pyruvate and ATP.

1.11 Non-kinase cell cycle regulatory proteins in *Plasmodium falciparum*

Homologues of the protein inhibitors of CDKs of the INK and cip/kip subclasses have not been identified in the *Plasmodium* genome database. However the mammalian inhibitor p21 was able to inhibit PfPK5-Pfcyc-1 activity *in vitro*, implying that a *P. falciparum* functional homologue may exist (Li *et al.*, 2001). No structural homologues of the activators CDC25 and Wee1 have been identified in the genome database (Doerig and Chakrabarti 2004). Furthermore, the ability of PfCDK5 to autophosphorylate and its promiscuity with respect to cyclins suggest that PfPK5 and its *in vivo* activators may be vastly different in structure from those of classical eukaryotic CDKs.

Origins of replication complex (ORC) proteins are a group of high molecular weight proteins that bind to origins of replication *in vivo* throughout the cell cycle and are a prerequisite for initiation of DNA synthesis. The ORC consists of six polypeptides with molecular weights ranging from 50 to 120 kDa. ORCs interact with the major groove of DNA and bind within 10 Å of the origin of replication (Kelly and Brown, 2000). A *P. falciparum* homologue of the ORC1 protein has been identified and is expressed in the sexual stage of the parasite (Li and Cox 2003).

The MCM (mini-chromosome maintenance) complex is a group of proteins that binds to the origin of replication. Chromatin binding of MCM proteins is strictly cell cycle regulated and depends on the replication status of the cell. They associate with CDC6 and Cdt1, which are responsible for loading the MCM complex onto the origin of replication site on DNA (Lei and Tye, 2001). A *P. falciparum* MCM4 homologue transcript has been detected in the sexual erythrocytic stage (Li and Cox, 2001). Several other MCM related proteins are predicted in the *P. falciparum* database (Doerig and Chakrabarti, 2004) and some of these have been characterised (Patterson *et al.*, 2006). No obvious homologues of the other pre-replication complex proteins like CDC6, and Cdt have been identified in the *P. falciparum* genome database.

The genes responsible for DNA replication in *P. falciparum* have been identified and cloned. *P. falciparum* homologues of DNA polymerase α and δ and the polymerase δ

accessory factor, PCNA, have been isolated (White *et al.*, 1993, Ridley *et al.*, 1991, Kilbey *et al.*, 1993). A second *P. falciparum* homologue to PCNA has recently been identified (Patterson *et al.*, 2002). A homologue to the large subunit of replication protein A, has also been recently characterised (Vos *et al.*, 2002).

A *Plasmodium* homologue for the 14-3-3 protein has been identified. Messenger RNA levels peak during the early trophozoite stage. (Al-Khedery *et al.*, 1999). In the mammalian cell cycle this protein is important in modulating the activity of CDC25 and also controls its cellular location. They are important in cellular signalling since they mediate interactions between various proteins.

1.12 Kinases as drug targets

Since protein kinases are pivotal in the regulation of cellular events they have been the subject of intensive study for the last three decades. Kinase inhibitors have been used extensively as tools in signal transduction and cell cycle research. Aberrant expression of kinases has been identified in many types of cancer and recently in neurodegenerative diseases such as Alzheimer's disease (Ahn *et al.*, 2005). This stimulated a huge interest in the clinical application of kinase inhibitors. The research has had clinical impact since several inhibitors are currently approved for clinical use, including sirolimus (rapamycin) and imatinib (Drucker and Lydon, 2000; Young and Nickerson-Nutter, 2005). Furthermore, a number of kinase inhibitors are in various phases of clinical trials with encouraging results to date (Meijer and Raymond 2003).

Rapamycin (sirolimus) inhibits the kinase mTOR (mammalian target of rapamycin), which is a key kinase in cell activation and in the $G_1 \rightarrow S$ phase transition of the cell cycle (Young and Nickerson-Nutter, 2005). This inhibitor is used extensively to maintain post-transplant immunosuppression and is currently undergoing clinical trials for use in certain cancers. Imatinib is an inhibitor of bcr-abl tyrosine kinase found in chronic myelocytic leukaemia and has been shown to have the unusual ability among cancer drugs to induce remission as a single therapeutic agent (Drucker and Lydon, 2000).

Currently high throughput screening by drug companies is rapidly replacing the more laborious and time expensive traditional screening methods. An example of this is the recent interest in CDK5 and the neurodegenerative diseases. CDK5 overexpression causes

the abnormal phosphorylation of Tau that leads to neuronal inclusions known as neurofibrillary tangles. High throughput screening has identified lead compounds that compete with ATP and will be useful for the rational design of CDK5 selective inhibitors. (Ahn *et al.*, 2005).

A major constraint in the clinical application of kinase inhibitors centres around their selectivity. The human genome codes for approximately 520 different protein kinases (Hanks, 2003). The most commonly used kinase inhibitory screening method makes use of panels of recombinant kinases that currently consist of fewer than 150 human kinases. It is therefore not possible to accurately predict the effect of inhibitors on other enzymes. It has also been shown that some “selective” kinase inhibitors also inhibit other types of enzymes involved in binding ATP and phosphotransfer. Furthermore, the structures of many of the inhibitors are based on the purine ring structure of the natural substrate ATP that may add to the non-selectivity of the inhibitors. Since a myriad of kinase inhibitors are described in the literature only the 2,6,9- trisubstituted purines used in this study will be discussed.

1.12.1 CDK selective inhibitors

The regulatory proteins of the cell cycle and cell division are highly conserved throughout evolution and therefore a variety of cellular models can be used to investigate the effect of potential inhibitors (Meijer and Raymond, 2003). One such model is the starfish oocyte model developed by Rialet and Meijer (1991) that was used for the *in vitro* testing of kinase inhibitory molecules based on the purine structure, which led to the development of the 2,6,9,-trisubstituted purines. Several compounds were found to have selective activity against CDKs.

This group includes isopentenyladenine (Rialet and Meijer 1991), olomoucine (Vesely *et al.*, 1994), roscovitine (De Azevedo *et al.*, 1997, Meijer *et al.*, 1997) and the purvalanol series of compounds (Chang *et al.*, 1999, Gray *et al.*, 1998). Olomoucine showed selective activity against mammalian CDKs but with only moderate inhibition (IC_{50} : 7 μ M).

Synthesis of derivatives of olomoucine resulted in the identification of roscovitine, which selectively inhibited a subset of CDKs when tested against 25 different kinases.

Roscovitine inhibited CDK1/cyclin B, CDK2/cyclin A, CDK2/cyclin E and CDK5/p35 with IC_{50} values of 0.65, 0.7, 0.7 and 0.2 μ M respectively, whereas the CDK4/cyclin D1

and CDK6/cyclin D2 complexes were not significantly inhibited by this molecule (De Azevedo *et al.*, 1997, Meijer *et al.*, 1997).

CDK activity is frequently deregulated in human cancers, which stimulated interest in the selective chemical inhibition of these enzymes (Monaco and Vallano 2003, Ahn *et al.*, 2005). A large number of CDK inhibitory molecules have been identified and evaluated as possible cancer therapeutic agents and some of them are currently undergoing clinical trials (Meijer and Kim 1997, Gray *et al.*, 1998, Gray *et al.*, 1999, Meijer *et al.*, 1999, Meijer and Raymond, 2003).

A number of general preclinical effects have been observed for the CDK inhibitors. CDK inhibitors inhibit neoplastic cell proliferation leading to cell cycle arrest in G₁ or G₂. CDK inhibition leads to apoptosis, and potentiates the antitumour effects of classical cytostatic agents (Grant and Roberts, 2003, Dai and Grant, 2004, Meijer and Raymond, 2003).

Flavopiridol was one of the first inhibitors thought to be selective for CDKs and is currently in an advanced state of clinical evaluation. Flavopiridol has been shown pre-clinically to inhibit cell cycle progression and to induce apoptosis (Dai and Grant, 2004). Specific effects of flavopiridol on cancer cells include the inhibition of multiple CDKs as well as inhibition of transcription due to its effect on P-TEFb (CDK9/CyclinT) (Dai and Grant, 2004). Clinical results indicate that although flavopiridol is tolerated in high doses the clinical efficacy is still questionable which is most likely related to pharmacokinetic and scheduling failure (Grant and Roberts, 2003).

Roscovitrine is currently being evaluated in clinical trials for the treatment of cancer (Dai and Grant, 2004) and holds promise as a treatment for the mantle cell lymphoma, a B-cell lymphoma with a poor prognosis. Roscovitrine was screened against four human mantle cell lymphoma lines and showed an IC₅₀ between 25 and 50 µM. Cells accumulate in the G₂-M interphase and entered apoptosis. The characteristic overexpressed cyclin D and MCL-1 proteins were downregulated and the phosphorylation status of RNA polymerase II decreased (Lacrima *et al.*, 2005).

Roscovitrine has a major advantage in that oral bioavailability is high enough to achieve plasma concentrations between 5-20µM. In preclinical studies IC₅₀ values ranged between 7.9 and 30.2 µM. At a concentration of 20 µM apoptosis was induced and the CDK mediated phosphorylation of retinoblastoma protein abrogated in mice (Meijer and

Raymond, 2003). Roscovitine has also been shown to inhibit replication of the varicella zoster virus with an IC₅₀ of 14 µM (Taylor *et al.*, 2004).

The purvalanol group of inhibitors was identified from combinatorial libraries of 2,6,9-substituted purines, and found to be more potent than roscovitine against the same subset of CDK/cyclin complexes (Gray *et al.*, 1998). Purvalanol A inhibits CDK1/cyclin B, CDK2/cyclin A, CDK2/cyclin E and CDK5/p35, with IC₅₀ values of 4, 70, 35 and 75 nM, respectively and purvalanol B inhibits the same subset with IC₅₀ values of 6, 6, 9 and 6 nM, respectively. Although purvalanol B is more potent when tested in isolated enzyme assays, it has no significant inhibitory effects on several mammalian cell lines (Gray *et al.*, 1998).

Apart from the usual isolated enzyme panel screening tests, the selectivity of the CDK inhibitors was confirmed by affinity chromatography studies. In this approach the inhibitor of interest is immobilized on agarose beads and reacted with cellular lysates from different organisms. Proteins bound to the immobilized inhibitor are then separated by electrophoresis and microsequenced to determine their identities. This approach confirmed that CDK1, CDK2, CDK5 and CDK7 are targets of the purvalanol compounds. It also showed that extracellular regulated kinase 1 (ERK1) and ERK2 are important targets of the purvalonols. Importantly it showed that casein kinase I is a major target for the purvalonols in unicellular protozoans (Knockaert *et al.*, 2000).

CDK7, CDK8 and CDK9 are not directly involved in the regulation of the cell cycle, but rather in the regulation of transcription. The development of CDK subset selective inhibitors will aid the elucidation of the roles of specific CDKs in disease processes and clarify their suitability as drug targets (Hirai *et al.*, 2005). Currently attention is focused on the development of inhibitors selective for the CDK2 and CDK1 group and for the CDK4/6 group, which are important in cell cycle control (Hirai *et al.*, 2005).

The effects of a number of kinase inhibitors on parasite invasion and survival in the red blood cell were tested by Dluzewski and Garcia in 1996. They found that tyrosine kinase inhibitors prevented parasite invasion only at concentrations above 200 µM but that Staurosporine, a Ser/Thr kinase inhibitor, prevented parasite invasion of the erythrocyte at 1 µM. This is consistent with the absence of classical tyrosine kinases, such as the insulin receptor, in the parasite. Flavopiridol and olomoucine have been shown to inhibit proliferation of the K1 *P. falciparum* isolate with IC₅₀ values of 2 and 15 µM respectively

(Greaser *et al.*, 1996b). The current availability of libraries of compounds based on CDK inhibitor lead compounds such as olomoucine provides an opportunity to screen large numbers of potential inhibitors against the pathogenic erythrocytic stage of the malaria parasite, *P. falciparum*.

1.13 Aims

The life cycle of *P. falciparum* diverges dramatically from that of the human host and offers the opportunity to identify parasite selective drug targets. Very little is known about the regulatory proteins such as kinases that control the progression of the parasite through the various stages of its life cycle. The genomic database of *P. falciparum* provides information that enables the identification of putative protein kinases with structural similarities to eukaryotic cell cycle regulators.

Bioinformatic analysis of raw sequence data from the NCBI malaria site in 2000 revealed a number of putative protein kinases that may be related to cell cycle kinases of higher eukaryotes. From this group Pfnek-3 and PfcGAK were arbitrarily chosen for cloning and expression, which constitute the first steps in evaluating a kinase as a feasible drug target. If kinases are to be useful targets in the development of new anti-malarial agents some indication of the ability of kinase inhibitors to prevent parasite proliferation is required.

Therefore the aims of this study are twofold:

- (i) Clone and express *P. falciparum* protein kinase genes possibly involved in the *P. falciparum* cell cycle with specific reference to cyclin G associated kinase (cGAK) and NIMA related kinases.
- (ii) Determine the effect of the 2,6,9,-trisubstituted purine-like cyclin dependent kinase inhibitors on *P. falciparum* survival.

Chapter 2 **Methods**

2.1 *Plasmodium falciparum* culture methods

The chloroquine resistant strain of *P. falciparum* (FCR-3) was obtained from Janet Freese of the Research Institute for Diseases in the Tropical Environment, Durban, South Africa. The *P. falciparum* 3D7 strain was a gift from Dr C. Doerig from Inserm Unit 515, Paris, France. The parasites were cultured according to the method of Jensen and Trager, 1977. All the solutions used in this study were prepared with Milli-Q water. All reagents were supplied by Sigma, USA, unless otherwise stated.

Briefly, parasites were maintained *in vitro* in culture flasks in an erythrocyte suspension (5% haematocrit) in RPMI-1640 culture medium, supplemented with 25 mM HEPES, 10 mM glucose, 0.32 mM hypoxanthine, 50 mg/l gentamicin and 10% (v/v) heat inactivated human plasma. The growth medium was replaced daily and erythrocytes were added when the cultures were in the trophozoite stage. Human plasma and whole blood were purchased from the South African National Blood Transfusion Services. Care was taken to match red blood cell groups with compatible plasma since plasma may contain antibodies to blood-group antigens and cause lysis of erythrocytes. Erythrocytes were prepared by washing them three times in an equal volume of phosphate buffered saline (PBS).

Cultures were synchronized with 5% D-sorbitol for 20 minutes at room temperature when the parasites were in the ring stage according to the method of Lambros and Vanderberg, 1979. All experiments were carried out on synchronized cultures. The percentage parasitaemia and parasite stages were assessed daily by microscopic examination of thin blood smears stained with commercial Giemsa stain (Merck, Germany).

2.2 Identification of kinase encoding open reading frames using the *Plasmodium falciparum* genomic database

Open reading frames encoding kinases were obtained prior to the completion of the *Plasmodium* genome sequence, using raw sequence data, which had been made available at the National Centre for Biotechnology Information (NCBI) (www.nlm.ncbi.gov). The data-mining approach consisted of several searches and cross checks.

Firstly, a decision was made as to the type of kinase. In this case it was a search for kinases involved in the regulation of the cell cycle. The first step in the strategy was to identify known amino-acid sequences coding for the selected kinases by doing a protein search using the ENTREZ program at NCBI. This program provides the primary amino acid sequences of known proteins from various organisms. This information was then used in the various BLAST (basic local alignment search tool) programs to identify homologous sequences.

The second step was to screen the *P. falciparum* nucleic acid database at the NCBI (www.ncbi.nlm.nih.gov/malaria/plasmodiumbl.html) with the selected amino acid sequence using the tblastn program. The amino acid sequence in FASTA format was entered into the query box. The tblastn program uses a protein query sequence and compares it to nucleotide sequence dynamically translated into six reading frames (three forward for the sense DNA strand and three reverse for the complementary strand). The contigs with nucleic acid sequences of significant similarity (low E-values) were then analysed further to obtain a complete open reading frame. Only individual high scoring alignments were examined and hits with previously assigned genes were ignored.

The *P. falciparum* contig was then submitted to the Open Reading Frame Finder program at the NCBI (www.ncbi.nlm.nih.gov/gorf.html). This program scans the nucleotide sequence for start and stop codons in six different reading frames. Once an ORF of significant length had been identified it was in turn subjected to a protein BLAST search, (blastp) of the entire GENBANK using the non-redundant database. This database contains verified gene sequence data from various organisms. The results of this BLASTP query provided information on the structure and possible function of the ORF. This method was used to identify two open reading frames with kinase signatures that were likely involved in the control of the *P. falciparum* cell cycle.

2.3 DNA isolation

Parasite cultures were centrifuged at 400 x g for 10 minutes in an appropriate centrifuge tube. The supernatant was aspirated and the erythrocyte pellets were suspended in 10 volumes of filter sterilised 0.015 % (w/v) saponin, which causes lysis of the erythrocytes (Orjih, 1994) for 5 minutes at room temperature. This was followed by centrifugation at 1750 x g for 10 minutes at 25°C after which the supernatant was discarded and the pellet

washed twice with PBS. The pellet, containing parasites, was transferred to a micro-centrifuge tube and frozen in liquid nitrogen until use.

DNA was isolated from either freshly prepared or frozen parasites based on the methods described in Sambrook *et al.*, 1989. The parasites were lysed in lysis buffer (10 mM Tris-HCl, 0.5% SDS and 100 mM EDTA, pH 8) before the addition of an equal volume of Tris-HCl buffered phenol (pH 8.0). The mixture was shaken gently for 10 minutes and centrifuged at room temperature at 10 000 x g for 15 minutes to separate the phases. High molecular weight DNA was precipitated from the upper phase by the addition of 10% v/v 3 M sodium acetate pH 5.5, and an equal volume of ice-cold isopropanol at -70°C for 30 min. This was followed by centrifugation for 20 min at 10 000 x g in a fixed angle rotor at 4°C. The supernatant was removed and the pellet dissolved in 500 µl TE buffer (10 mM Tris-HCl, 100 mM EDTA, pH 8). Residual RNA was removed by the addition of RNase A (50 mg/ml) and incubation at 37°C for 30 minutes. The RNase A was removed by a second phenol separation and the resultant supernatant was extracted with an equal volume of chloroform.

DNA was precipitated from the supernatant by the addition of 10% v/v 3M NaAc, pH 5.5 and 2.5 volumes of 100% ice cold ethanol. The mixture was left at -70°C overnight and then centrifuged at 10 000 x g, 30 min, at 4°C to collect the DNA pellet. The pellet was washed twice with 70% ethanol to remove the salts, followed by a 100% ethanol wash to hasten the drying process. Dried pellets were dissolved in sterile Milli-Q water.

A sample of DNA was diluted and scanned on a UV scanning spectrophotometer between 300 and 200 nm to determine the DNA concentration and assess the purity of the nucleic acid. The absorbance reading at 260 nm was used to calculate the DNA concentration. An A₂₆₀ of 1 is equivalent to 50 µg/ml of DNA and an absorbance ratio (A₂₆₀/A₂₈₀) of ≤1.8 is indicative of possible phenol and/or protein contamination. DNA was further analysed by electrophoresis on a 1.5% (w/v) agarose gel (section 2.5) and the bands were visualized with ethidium bromide.

2.4 The Polymerase Chain Reaction (PCR)

2.4.1 Primer design for PCR of *in silico* identified kinases

The sequence information of the selected kinase-encoding ORFs, PfcGAK and PfNek-3, was used to design primers for their amplification by PCR. The PCR products were cloned

into expression vectors once their identities had been verified. Primer regions were selected to avoid 3' complementarity and the subsequent formation of primer dimers. This resulted in the exclusion of some of the amino- and carboxyl terminal amino acids from some of the primers.

Nucleotide recognition sequences for restriction enzymes were added to the 5' ends of all the primers to ensure the directional insertion of the amplicon into the vector during ligation reactions. Primer sequences are given in Appendix B1.1. Primers were synthesised by IDT, (UK) and prepared at a 100µM concentration in sterile DNase and RNase free water, aliquoted and stored at -20°C until use. The restriction sites of the cloning vector determined the identity of the additional nucleotides. Three vectors were used in this study namely pHB6 (Roche, Germany), pQ80L (Qiagen, Germany) and pET41a (Novagen, USA). Their individual characteristics are summarized in table 2.1 and their respective vector maps are shown in Appendix A1. In all cases translation start sites as well as stop codons were coded for by vector sequence.

Table 2.1 Summary of expression vector characteristics

Vector name	<i>pHB6</i>	<i>pQ80L</i>	<i>pET41a</i>
N-terminal epitope tag	HA (haemagglutinin)	His ₆	GST
C- terminal epitope tag	His ₆	None	His ₆
Restriction sites used for insert	<i>Hind</i> III <i>Eco</i> R I	<i>Bam</i> H1 <i>Hind</i> III	<i>Bam</i> H1 <i>Hind</i> III
Promoter	P _{trp} -lac	Phage T5	Phage T7
Resistance	Ampicillin	Ampicillin	Kanamycin

2.4.2 Amplification of open reading frames

The Expand High Fidelity PCR system (Roche, Germany) was used to amplify the nucleotide sequences coding for the relevant open reading frames for cloning purposes. The polymerase in this system is a combination of *Taq* and *Pwo* DNA polymerase. *Pwo*-polymerase increases the fidelity of DNA synthesis due to its 3'-5' exonuclease proofreading activity. This enzyme combination results in a 3-fold increased fidelity of DNA synthesis with an error rate of 8.5×10^{-6} compared to that of *Taq* DNA polymerase used alone, which has an error rate of 2.6×10^{-5} .

High fidelity copying of template DNA was required since the PCR products were to be cloned into an expression vector with the aim to obtain protein for functional studies. Due to the exonuclease activity of *Pwo*, two master-mixes were prepared which separate the enzymes from the template until initiation of the polymerase reaction. The two PCR master mixes were prepared at 0°C in an ice bath to minimize non-specific binding of the primers to the DNA template. This is a common problem since *P. falciparum* DNA has abundant A and T nucleotide stretches. A modified hot start procedure was used where the samples were kept at 0°C and only transferred to the thermal cycler once the temperature of the heating block reached 90°C.

2.4.3 PCR cycling conditions

PCR was carried out in a Perkin Elmer 2400 thermal cycler (Perkin Elmer, USA) equipped with a heated lid. The PCR parameters were optimised in terms of DNA concentration, primer concentration and MgCl₂ concentration. Annealing temperatures were initially chosen based on the annealing temperatures supplied by IDT and optimised experimentally. All PCR reactions were carried out with 200µM dNTPs. Optimum annealing temperatures and MgCl₂ concentrations are given in chapter 3.2.

A typical cycle sequence for the amplification of a kinase coding sequence was as follows:

Denaturation:	2 min at 94°C		
<i>Cycle:</i>			
Denaturation:	30 sec at 94°C	}	30 cycles
Annealing:	30 sec at 66°C		
Elongation:	60 sec at 72°C		
Final extension:	7 min at 72°C		

2.4.4 Colony screening by PCR

PCR was also used to screen colonies for kinase encoding inserts. Transformed *Escherichia coli* colonies harbouring a plasmid were selected and transferred to 5 ml Luria Bertani (Appendix A2) broth containing the appropriate antibiotic (See table 2.1) and grown overnight at 37°C. A 100 µl sample from each culture was transferred to a microcentrifuge tube and the bacterial cells collected by centrifugation at 5 000 x g for 1 minute. The

supernatant was discarded, the bacterial pellet resuspended in 50 µl sterile water and boiled in a heating block for 15 minutes. The tubes were centrifuged again for 5 minutes at 10 000 x g and a sample taken for the PCR reaction. Screening reactions were carried out using the Roche Master Mix preparation (Roche, Germany).

PCR reactions were prepared as follows:

- 12 µl DNA containing sample
- 1.5 µl Sense primer (10 µM)
- 1.5 µl Anti-sense primer (10 µM)
- 15 µl PCR master mix: buffer, *Taq* polymerase, MgCl₂ and dNTPs
- 30 µl Total volume

PCR was carried out as described in section 2.4.3 and the results analysed on a 1.5% agarose gel as described in section 2.5.

2.5 Agarose gel electrophoresis

PCR products, DNA and products obtained from restriction digests of PCR products or plasmids were analysed by agarose gel electrophoresis. Agarose gels were prepared as either a 1.5 or 2 % (w/v) gel, containing 1µg/ml ethidium bromide, in a 1xTris-acetate-EDTA (TAE) buffer. The 50x buffer was prepared by mixing 242 g Tris base, 57.1 ml glacial acetic acid, 100 ml of 0.5 M EDTA, pH 8, with water to 1 litre. Gel loading buffer (1µl to 9 µl sample) was added to the sample prior to loading the samples on the gel and consisted of 0.12% bromophenol blue, 0.12% xylene cyanol FF and 30% glycerol in water.

Mini-submarine gels (Biorad, USA) were electrophoresed for 3 hours at 30V after the samples had been loaded in the sample wells. A 100 bp size marker (Roche, Germany or Promega, USA) was used to estimate the sizes of the PCR and restriction digestion products. The bands on the gel were visualised by viewing on a UV transilluminator (UVP®) (302nm).

2.6 Subcloning of PCR products

PCR products were cloned into one of three vectors; namely pHB6 (Roche, Germany), pQ80L (Qiagen, Germany) and pET41(a) (Novagen, USA). The insert containing plasmids were first established in either the DH5α (Invitrogen, USA) or JM109 (Stratagene, USA) *E.*

coli strain and then transferred to either the BL21 (Novagen, USA), SG1 (Qiagen, Germany) or MRep15 (Qiagen, Germany) *E.coli* strains after the construct sequences had been verified. The BL21 Rosetta strain (Novagen, USA) was also used and contained an additional plasmid coding for rare tRNAs.

2.6.1 Preparation of the amplicons for ligation to vector

Prior to restriction digestion the PCR reagents/enzymes were removed from the amplicons by using the High Pure PCR Product Purification Kit (Roche, Germany). DNA polymerase would abrogate the restriction digestion by adding complementary nucleotides to the digested ends. The High Pure Kit binds the DNA to silica in a mini-column followed by washing out of the proteins and unincorporated nucleotides. The amplicons were eluted in DNase/RNase free MilliQ water.

Double enzyme restriction digests were then performed on the amplicons. Amplicons were digested with either *EcoR*I and *Hind* III in buffer B from Roche, Germany (10 mM Tris.HCl, 100 mM NaCl, 5 mM MgCl₂, 1 mM β-mercaptoethanol, pH 8.0) or with *Bam*HI and *Hind*III in the same buffer. All three restriction enzymes used displayed 100% activity in buffer B. The enzymes and nucleotide fragments were removed prior to ligation by another round of silica mini-column purification.

The appropriate plasmid was also subjected to restriction digestion by the enzymes followed by a calf intestinal alkaline phosphatase (Roche, Germany) treatment to prevent re-annealing of plasmid. This step increased the number of insert containing clones. Enzymes, nucleotide fragments and buffer salts were removed by silica mini-columns.

2.6.2 Ligation of the PCR products to the expression vector

Aliquots of the amplicons and plasmid were analysed by electrophoresis on a 1.5 % agarose gel to calculate the plasmid:insert ratio. The intensity of the bands was compared to that of the 100 base pair (bp) mass ladder (Roche, Germany) for an estimation of the amounts of DNA.

Ligation reactions were carried out using T4 DNA ligase (Promega, USA) in a 10 µl reaction at one enzyme unit (U) per reaction and an insert to vector molar ratio of 3:1. The

reaction mixture contained 100 ng vector, digested amplicons, ligation buffer (30 mM Tris-HCl, 10 mM MgCl₂, 10 mM dithiothreitol (DTT), 1 mM ATP, pH 7.8) and water. Ligation reactions were allowed to proceed at 16°C overnight. A negative control reaction, to check for recirculisation of the plasmid without ligation of the insert was also performed.

In some cases the Rapid DNA Ligation Kit supplied by Roche was used. In this ligation the reactions contained 10 µl of DNA at a insert to vector molar ratio of 1:3 in 1x DNA dilution buffer, 1 x T4 DNA ligation buffer and 5 U of T4 DNA ligase in a final reaction volume of 20 µl. The ligation reactions were allowed to proceed for 30 minutes at room temperature. The compositions of the DNA dilution and ligation buffers are proprietary.

2.6.3 Preparation of competent cells

A glycerol stock of the appropriate *E. coli* strain was streaked on a Luria-Bertani (LB) agar plate (Appendix A2) and grown overnight at 37 °C. A colony from this plate was grown in 5 ml LB media overnight at 37 °C with shaking at 250 rpm. LB medium, (100 ml), was inoculated the next day with 1 ml of overnight culture and grown until the absorbance reading at 600 nm was between 0.4 and 0.5.

The culture was cooled in ice water and the bacteria collected by centrifugation at 2000 x g for 20 minutes at 4°C. Cells were resuspended in 10 ml transformation buffer (60 mM CaCl₂, 10 mM Pipes, pH 7.4 and 15 % glycerol) and left on ice for 30 min. This was followed by centrifugation as before. The supernatant was discarded and the cells gently resuspended in 3 ml of transformation buffer until a homogenous suspension was obtained while tubes were kept on ice. The competent cells were divided into 200 µl volumes into sterile polypropylene vials with screw-caps, frozen in liquid nitrogen after which they were stored at -70 °C for not more than 3 months.

2.6.4 Transformation of competent bacteria with plasmids

Competent *E. coli* cells of the appropriate strain were thawed on wet ice. Three microlitres of the ligation reactions were transferred to ice-cold microcentrifuge tubes. The thawed competent cells were mixed gently and 45 µl were transferred to the tubes containing the ligation mixture. The tubes were incubated on ice for 30 minutes followed by a heat-shock treatment for 90 seconds at 42°C, after which the cells were returned to ice. This was followed by the addition of 400 µl of warmed LB media to the tubes, which were

transferred to a shaking incubator at 37°C for 60 minutes. After this growth period, 100 and 300 µl aliquots were plated onto LB agar plates containing the appropriate antibiotic selector. Colonies were counted after overnight incubation at 37°C and analysed for insert containing plasmids. For the transformation of strains used for protein expression, between 50 and 200 ng of plasmid construct was used. Two strains were used, namely BL21 Codon Plus (Stratagene, USA) and Rosetta (Novagen, USA). A negative control without DNA was included as well as a positive control consisting of plasmid without cloned insert DNA.

2.6.5 Plasmid isolation

Colonies were screened for the presence of insert by plasmid isolation and restriction digestion as well as PCR of whole cell DNA as described in section 2.4.4. Plasmids were isolated by the alkaline lysis method of Sambrook *et al*, 1989. Colony stabs were inoculated into LB medium containing the appropriate antibiotic and grown overnight at 37°C with shaking at 200 rpm. Glycerol stock cultures (15% glycerol) were prepared and stored at -70°C.

The remainder of the bacteria were collected by centrifugation at 3 000 x g for 10 min and the growth medium removed by aspiration. The bacterial pellet was resuspended in ice-cold lysis buffer (25 mM Tris-HCl, 10mM EDTA, 50 mM glucose, pH 8.0) and incubated for 5 min at room temperature. The cells were lysed by the addition of freshly prepared 0.2 M NaOH in 1% SDS and gentle mixing. Cell walls, proteins and genomic DNA were precipitated by the addition of ice-cold 5 M potassium acetate, pH 4.8. The plasmid containing supernatant was separated from the precipitate by centrifugation at 10 000 x g for 20 min at 4°C and the supernatant was transferred to a new tube.

RNase A was added to a final concentration of 20 µg/ml and the mixture incubated at 37°C for 1 hour. Following the RNase digestion, an equal volume of Tris-HCl buffer saturated phenol, pH 7.9 was added to the mixture, which was shaken for 1 min and then centrifuged for 10 minutes at 5 000 x g. The upper phase was collected and residual phenol was removed by a chloroform extraction. The plasmid containing upper phase was kept on ice and precipitated by the addition of 1/10th of the volume of 3 M sodium-acetate, pH 5, and two and a half volumes of ethanol and storage at -70°C overnight. The plasmid was harvested by centri-fugation at 10 000 x g for 20 min at 4°C, washed once with 70% ethanol followed by a 100% ethanol wash and air-dried in a laminar flow cabinet. The

precipitated plasmid was dissolved in water and the concentration and purity determined as for DNA described in section 2.3.

2.6.6 Analysis of plasmids

Isolated plasmid DNA was subjected to restriction analysis with the appropriate enzymes. Restriction digests were carried out in 20 µl reactions at 37°C for three hours after which the reaction was stopped by incubation at 65°C for 15 minutes. The digested products were analysed on a 1.5% agarose gel. Alternatively, a small volume (20 µl) of bacteria was lysed and the supernatant subjected to PCR analysis as described in section 2.4.4 .

2.7 DNA sequencing

DNA sequencing was carried out at various stages in the study. The identity of the amplicons was confirmed by manual sequencing prior to cloning. Following cloning, manual sequencing across the insertion sites confirmed that the coding sequence was inserted in the correct reading frame. The full-length sequence of the DNA insert to check for mutations was done by automated sequencing and outsourced to Inqaba Biotech, (RSA).

Manual sequencing was carried out using the T7 Sequenase version 2.0 DNA Sequencing Kit (Amersham Life Science, Inc., UK). This kit is based on the DNA chain-termination method of Sanger *et al.*, (1977). Sequencing products were separated on a 0.4 mm thick 8% denaturing polyacrylamide gel, which was cast in a Sequi-Gen GT Nucleic Acid Electrophoresis Cell (Bio-Rad, USA) according to the instructions of the manufacturer. Prior to use, the gel assembly was transferred to the electrophoresis apparatus and the sharks tooth combs inserted after the surface of the gel was rinsed with Tris-borate-EDTA (TBE) buffer (0.09 M Tris-borate, 1 mM EDTA, pH 8.0). The gel was pre-electrophoresed for 60 minutes at 80W to increase the gel temperature to $\pm 50^{\circ}\text{C}$.

2.7.1 Internal primers for the open reading frames

Three additional primers were used to obtain the full-length sequence of PfcGAK. Their relative positions in the ORF are shown in Appendix B2. Since the ORF for PfNek-3 was only 1041 bp long internal primers were not required to determine the complete sequence of PfNek-3.

Internal PfcGAK primers:

Internal sense primer 1 : **5' TGT GAC TTC TGT TCA CAT 3'**

Internal sense primer 2 : **5' CAT ATT CCC GCA GAT ATC 3'**

Internal reverse primer: **5' TAA TTC GGG AGG TCT ATA 3'**

2.7.2 Sequencing of plasmid DNA

Sample preparation

Plasmid DNA was prepared from 1.5 ml cultures for sequencing and the plasmid pellet was dissolved in 50 µl water. A plasmid DNA sample of 20 µl was alkaline denatured for 5 minutes at room temperature by the addition of 4 µl of 2 M NaOH and 2 mM EDTA. The reaction was neutralized by adding 20 µl 3 M sodium acetate, pH 5. The DNA was precipitated with 150 µl ethanol at -70°C for at least 45 minutes, centrifuged and washed with 400 µl of 70% ethanol and air-dried overnight in a laminar flow unit and resuspended in 7µl water.

The annealing mixture was prepared which consisted of 7µl DNA, 2 µl of 5 x Sequenase reaction buffer (200 mM Tris-HCl, pH 7.5, 100 mM MgCl₂, 250 mM NaCl) and 1 µl of 10 µM sequencing primer. The mixture was heated to 65°C for 2 minutes and then slowly cooled to less than 35°C over a 15-30 minute period. The sample was then centrifuged briefly, chilled on ice and used in a sequencing reaction within 1 hour.

Labelling reaction

During the labelling reaction, α-[³²P] dATP is incorporated into newly synthesized DNA chains. The following was added to the ice-cold DNA mixture: 1 µl 0.1 M DTT, 2 µl diluted dGTP labelling mix, 0.5 µl α-[³²P]-dATP (400Ci/mmol dATP), (Amersham, UK) and 2 µl diluted Sequenase polymerase. This mixture was incubated at room temperature for 2-5 minutes.

Termination reaction

Termination reactions incorporate a specific dideoxy-nucleotide and cause termination of the synthesis of the specific chain. Reactions were terminated by transferring 3.5 µl of labelling reaction to each of the tubes containing dideoxy guanine-, adenine, thymidine or cytosine. Termination reactions were allowed to proceed for 5 minutes at 37°C. Each reaction was stopped by the addition of 4 µl stop solution (95% formamide, 20 mM EDTA, 0.05% Bromophenol Blue, 0.05% Xylene Cyanol FF).

Samples were then heated to 80°C for 2 min and 3 µl of each reaction were loaded per lane on the polyacrylamide gel. Each sample was loaded three times at three hour intervals and the gels electrophoresed for between 6 and 9 hours at 80 W. When electrophoresis was complete the plates were separated and the gel transferred onto a previously used X-ray film. The gel was covered in cling-film and exposed overnight to a multipurpose autoradiography film (Amersham, UK) at -20°C. The film was developed in the X-Ray unit of the Johannesburg General Hospital and analysed manually.

2.7.3 DNA sequence analysis

Data obtained from DNA sequencing were analysed by the DNA AssistTM programme. The programme performs standard procedures such as the transformation into complementary DNA, mRNA and translation of DNA sequence into protein sequence. It also performs predictions of protein properties such as the molecular weight, iso-electric points, amino acid analysis and hydrophobicity and hydrophilicity plots.

2.8 Expression of recombinant proteins

2.8.1 Induction of protein expression of the cloned *Plasmodium* inserts

Overnight cultures (5 ml) of bacteria containing plasmid were grown in the presence of the appropriate antibiotic in 2YT (Appendix A2) medium. The complete culture was used to seed a 100 ml culture the next morning and allowed to grow until the absorbance reached 0.7 at 600 nm. The expression of the cloned gene was then induced by the addition of 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) (Invitrogen USA).

After 4 hours of growth at 30°C the bacteria were pelleted at 3000 x g, the growth medium removed, the pellets frozen in liquid nitrogen and stored at -70°C until protein isolation. Alternatively a 100 ml culture was seeded with a 100 µl glycerol stock culture, allowed to grow to a density of 0.6 measured at 600 nm and then induced overnight with 0.5 mM IPTG at 30°C. Cells were harvested the next day and stored as described above.

2.8.2 Isolation of hexa-histidine tagged recombinant protein from bacteria

Hexa-histidine tagged proteins were isolated by immobilized metal affinity chromatography using a nickel-nitrolotriacetic acid (NTA) matrix obtained from Qiagen, Germany. Isolation procedures were carried out on ice or at 4°C. Frozen bacteria (± 1 g) were resuspended with 5 ml of lysis buffer. Lysis buffer consisted of 50 mM Tris-HCl, pH 7.4, 300 mM NaCl, 20 mM imidazole to which was added lysozyme (1 mg/ml), EDTA free protease inhibitor cocktail (Roche, Germany), 1 mM phenylmethyl-sulfonyl fluoride (Sigma, USA), and 2% Tween 20 (Calbiochem, USA). EDTA chelates the Ni^{2+} -ions from the agarose matrix. A Tris-HCl buffer was chosen since phosphate buffers interfere with downstream kinase assays. Tris-HCl at a molarity greater than 100 mM causes reduction of the Ni^{2+} ions, which are required for binding the hexa-histidine tagged proteins. The bacteria were then incubated on ice for 20 min prior to sonication on ice for 20 seconds with a microprobe.

Sonication was followed by centrifugation at 10 000 x g for 30 min at 4°C to pellet the cellular debris and nucleic acid fragments. A sample of the supernatant was taken and 5 x SDS-PAGE sample buffer (0.225 M Tris-HCl pH 6.8, 5% SDS, 50% glycerol, 0.25 M DTT and 0.05% bromophenol blue) added, and stored at -20 °C for further analysis. To the cleared supernatant 200 μl of a 50% Ni-NTA slurry (pre-equilibrated in lysis buffer) was added. The tubes were transferred to a blood mixer apparatus (Merck, Germany) and allowed to bind for 4 hours at 4°C.

The Ni^{2+} -beads were then collected by centrifugation at 1000 x g for 10 minutes and transferred to microcentrifuge tubes. The Ni^{2+} -beads were washed three times with 1000 μl wash buffer (50 mM Tris-HCl, pH 7.4, 60 mM imidazole, 300 mM NaCl and protease inhibitors) by centrifugation at 1000 x g. The hexa-histidine tagged protein was eluted from the resin by the addition of 5 x 200 μl elution buffer (50 mM Tris-HCl, pH 7.4, 300 mM NaCl, 250 mM imidazole and protease inhibitors). The eluted protein was concentrated using a disposable ultrafiltration device (Vivaspin, Sartorius, Germany) with a molecular weight cut off point of 10 000. Protein concentration was determined using the dye binding assay of Bradford *et al.*, 1976. An equal volume of glycerol was added to the samples and they were stored at -20°C. Samples from the eluates were analysed by denaturing SDS-PAGE (Laemmli, 1970) and tested for kinase activity.

2.8.3 Isolation of GST-fusion proteins

Synthesis of the recombinant protein was induced as described in 2.8.1. Cells were thawed and resuspended in Bug Buster Reagent (Novagen, USA) to which EDTA containing protease inhibitor cocktail (Roche, Germany) as well as PMSF was added. Cells were incubated for 30 minutes on ice and sonicated for 20 seconds to shear the bacterial DNA, which makes the preparation viscous. This was followed by centrifugation at 10 000 x g for 30 minutes after which the supernatant was kept and the pellet discarded. An appropriate amount of Glutathione SepharoseTM 4B beads (Amersham Biosciences, UK) was added to the supernatant and the mixture incubated at 4°C for 4 hrs on a blood tube rotator.

Following the incubation period the mixture was centrifuged at 500xg for 2 minutes to collect the GST beads and the supernatant was discarded. The GST beads were washed 4 times in 10 volumes of phosphate buffered saline supplemented with protease inhibitor cocktail and PMSF as described before. A sample of GST-beads was electrophoresed directly on SDS-PAGE gels by adding 2x electrophoresis sample buffer to a sample of beads followed by boiling for 5 minutes. Elution of the protein from the remaining beads was carried out in an equal volume of PBS containing 10 mM reduced glutathione, repeated three times. The eluates were pooled and stored at -20°C after the addition of glycerol to 50% of the volume.

2.9 Analysis of the recombinant protein

2.9.1 SDS-Polyacrylamide gel electrophoresis (SDS-PAGE)

The proteins obtained from the induced bacterial cultures were analysed by denaturing SDS-PAGE using the discontinuous system of Laemmli, 1970. Proteins were denatured by adding the appropriate amount of 5 x sample buffer. Samples were boiled for 5 min and then cooled on ice prior to loading on the gel. The stacking gel was 4% T (total acrylamide) and 2.7% C (crosslinker) in polyacrylamide and bisacrylamide respectively and buffered to a pH of 6.8 with 0.125 M Tris-HCl and containing 0.1% SDS. The separating gel was 12.5% T and 2.7% C in polyacrylamide and bisacrylamide respectively. The separating gel was buffered to pH 8.8 with 0.375 M Tris-HCl and contained 0.1 % SDS.

After loading of samples, gels were electrophoresed in a Biorad Mini Protean II system at a constant voltage of 200V for 80 minutes. After separation was complete the gels were either subjected to Western blotting or stained with Coomassie blue. Sizes of proteins on Western blots were determined with a 6xHis protein standard mixture from Qiagen, Germany and on Coomassie stained gels by using protein molecular weight standards from Biorad, USA.

2.9.2 Western blotting and detection of expressed proteins

Polyvinylidene fluoride (PVDF) (Microsep, USA) membranes were cut according to gel sizes and pre-wetted in methanol for 2 min. The membranes were washed once (5 min) in water and equilibrated in transfer buffer (48 mM Tris-HCl, 39mM glycine, 20% methanol, pH 9.2). Gels were equilibrated for 30 minutes in transfer buffer with two buffer exchanges to remove the SDS. The gel sandwich was assembled and transferred to the blotting apparatus. The proteins were transferred electrophoretically to the PVDF membrane by using a Biorad Mini Trans-Blot® Electrophoretic Transfer Cell. Transfer was carried out at 100V, 130 mA for 60 minutes at 4°C.

Following transfer, the PVDF membrane was removed and incubated in blocking solution for 60 minutes at room temperature. Blocking solution consisted of 5% bovine serum albumin (Sigma, USA) in Tris-HCl-buffered saline-Tween 20 solution, (TBST), (50 mM Tris-HCl, 150 mM NaCl, pH 7.5 with 0.1 % v/v Tween 20). The albumin binds to the PVDF membrane where protein is not bound to prevent non-specific binding of antibodies to the membrane. This was followed by incubation of the membrane with the primary anti-HA (Roche, Germany) or anti-Histidine (Novagen, USA) mouse antibodies, overnight at 4°C. Thereafter the membranes were washed six times for 10 minutes with 30 ml TBST, incubated with the second antibody (conjugated to horse-radish peroxidase) for 1 hour at room temperature and washed again six times for 10 minutes with TBST. The TBST was dried off with filter paper and the damp membranes incubated for 60 seconds with Western Lightning™ Chemiluminescence Reagent (Perkin Elmer, USA) which was sensitive to 1 pg of protein. The chemiluminescent light signal was visualized by exposing the membranes for 1, 3 and 7 minutes to Hyperfilm™ -multipurpose X ray film (Amersham, UK).

The film was developed by processing through an automated X-ray processor at the Johannesburg General hospital. The molecular weights were estimated by including a

histidine tagged molecular weight marker (Qiagen, Germany) on the gel or alternatively a standard protein marker was electrophoresed which was stained with Coomassie blue dye as described below. The molecular weights of the recombinant proteins were calculated from the construction of a calibration curve of relative migration distance (R_m) against the log protein molecular weight. R_m is calculated by dividing the migration distance of a protein band from the origin by the total migration distance of the indicator dye.

2.9.3 Coomassie brilliant blue staining of PVDF membrane

The protein molecular weight markers were visualized by cutting the respective lane from the rest of the membrane and staining it with Coomassie brilliant blue R-250 (0.1% w/v in 50% methanol and 10% acetic acid) for 15 minutes. The membrane strips were destained with a mixture consisting of 50% methanol and 10% acetic acid. This staining method was also used for the staining of protein gels.

2.9.4 Stripping and re-probing of PVDF membranes

For clones inserted into the pHB6 vector the membranes were probed with both anti-hexahistidine and anti-HA antibodies to determine if both tags were present on the recombinant protein. The absence of one of the tags indicates a faulty transcription or translation process. The first antibody was removed by incubating the membrane in stripping buffer (65.5 mM Tris-HCl, pH 6.8, 2% SDS, 100 mM β -mercaptoethanol) for 1 hour at 60°C. The stripped membranes were washed 3 times for 10 minutes with TBST before proceeding to incubation with the second primary antibody.

2.9.6 Determination of protein concentration

The concentration of purified recombinant protein was estimated by using the Bio-Rad Protein Assay that is based upon the dye-binding method of Bradford *et al*, 1976. This method was chosen because it is fast, sensitive to 1 μ g of protein and has little interference from various chemicals like the glycerol used for storage of the isolated protein. A disadvantage of the method is the construction of a calibration curve for every assay. Calibration curves were constructed by using a range of standard BSA protein concentrations from 1-20 or 1–100 μ g.

2.10 Kinase assays and densitometric analysis

2.10.1 Kinase assays

Recombinant proteins were tested for kinase activity by performing a standard kinase assay using histone protein 1, myelin basic protein and bovine casein as exogenous substrates (Wang and Roach, 1993). NaF and β -glycerophosphate were added to inhibit phosphatases. Kinase assays were carried out in a final volume of 30 μ l, which contained the following:

Recombinant protein (0.5-3 μ g)
20 mM Tris-HCl, pH 7.4
20 mM MgCl_2
2 mM MnCl_2
10 mM NaF
10 mM β -glycerophosphate
10 μ M ATP
0.5 μ l γ -[^{32}P] ATP 3000 Ci/mmol (Amersham, UK)
Substrate (2 μ g casein, 5 μ g myelin basic protein, 10 μ g histone H1 protein)

Reaction mixes were prepared containing all the relevant reagents and the appropriate volume was then transferred to microcentrifuge tubes containing the recombinant enzyme. Kinase reactions were carried out at 30°C for either 30 or 45 minutes. Reactions were terminated by the addition of 5 x SDS sample buffer, followed by mixing and boiling for 5 minutes. A sample of 20 μ l was loaded onto a SDS-12% polyacrylamide gel and electrophoresed as described before. Following electrophoresis the gel was stained in Coomassie blue stain and the destained gel was dried in a Biorad Sr. Gel Dryer (BioRad, USA). The dried gel was transferred to a X- ray cassette and exposed to HyperfilmTM (Amersham, UK) for 12 hours unless indicated otherwise. The film was developed as described before.

2.10.2 Densitometric analysis

Densitometry was performed on selected Coomassie blue stained SDS-PAGE gels and their corresponding autoradiographs using a Hoefer model GS300 transmittance /reflectance scanning densitometer. The data were integrated using the Hoefer Electrophoresis data reduction system Software GS365W version 3.01 and were presented as casein peak areas.

For each sample the casein peak area of the autoradiograph was divided by the Coomassie stained gel casein peak area, to obtain a ratio. Since the casein control showed incorporation of a small amount of radioactivity this ratio was arbitrarily assigned a value of 1. All the other ratios were compared to this baseline level.

2.11 Hypoxanthine incorporation assay

The IC₅₀ values of the purine derived CDK inhibitors on whole parasites were determined by the [³H]-hypoxanthine incorporation assay as described by Desjardins *et al*, 1979. Parasites were exposed to the agents for 76 hours, a period which permits near completion of two cycles of DNA synthesis. The longer exposure time (as opposed to 48 hours) also permits evaluation of the merozoite reinvasion of red blood cells.

Briefly, parasites synchronized in the ring stage were transferred to a 96-well plate at a 0.5% parasitaemia and 1% haematocrit in the presence of various inhibitor concentrations (0.01-100µM). In each experiment individual concentrations were repeated in quadruplicate. The purines were dissolved in DMSO at stock concentrations of 1 or 10 mM. Labelled [³H]-hypoxanthine (0.5µCi per well, 27.0 Ci/ mmol) (Amersham Biosciences, UK) was added after 24 hours and the cells were harvested on a GFB-filtermat (Wallac, UK) with a Titertek[®] (Flow Laboratories, USA) cell harvester after 76 hours of exposure.

The filtermats were dried, transferred to plastic bags, which were filled with scintillation cocktail (10 ml) (Betaplate Scint , Wallac , UK) and sealed. Incorporation of [³H]-hypoxanthine into DNA was measured in a 1205 Betaplate liquid scintillation counter (Wallac, UK). The IC₅₀ values are reported as the mean ± standard deviation of at least four experiments. IC₅₀ values were calculated by the Enzfitter[®] program which uses the Marquard algorithm to perform nonlinear regression analysis on non-transformed data. The Prism[®] software package was used to draw the sigmoidal log dose response curves.

2.12 Stage experiments

Parasites in the ring stage were plated at a 2% parasitaemia and a 5% haematocrit in a final volume of 2 ml in a 6 well plate. The test compound was added to the ring, trophozoite and early schizont stages at a concentration equivalent to an IC₉₀. Culture medium, containing the appropriate concentration of compound to be tested was replenished daily. Slides were

made at 12 hour intervals and parasites were visualized with Giemsa stain. Ten fields of each slide were counted and the percentage of parasites calculated. The data were analysed on a spreadsheet.

2.13 Protein motif search

The PROSITE database (<http://www.expasy.org/prosite>) of the Swiss Institute of Bioinformatics, was used to identify conserved motifs and domains in the protein sequences. The database identifies biologically meaningful amino acid signatures that are described as patterns or profiles. Conserved residues in proteins are good predictors of protein function. Patterns are more rigid in nature and are usually located in highly conserved regions such as enzyme catalytic sites. Profiles are less rigid and are based on similarity scores of alignments between a profile and a query sequence. An advantage of this method is the identification of poorly conserved domains. PROSITE makes use of the hidden Markov model and DALI algorithms. PROSITE signatures are linked to annotation documents where information on the signature such as origin, taxonomic occurrence, domain architecture, function, 3D structure, main characteristics and references are found. (Hulo *et al.*, 2004). The presence of signal sequences were predicted by the SignalP 3.0 Server which makes use of artificial neural networks and the hidden Markov model algorithms as described by Bendtsen *et al.*, 2004 (www.cbc.dtu.dk/services/SignalP/).

Multiple alignments of protein sequences were performed using the ClustalW programme at the European Institute for Bioinformatics (www.ebi.ac.uk). Blast analysis was carried out using their interface as well as the interface at the NCBI.

Chapter 3 Results

3.1 Identification of kinase-encoding open reading frames

Several putative kinase encoding open reading frames were identified in 2000 using cell cycle related protein kinase sequences with the BLAST programs on raw sequence data. At the time of the searches the *P. falciparum* genome project was approximately 80% completed. Some of the identified ORFs coded for protein kinases already described in the literature like PfPK5 and Pfckr1. New putative ORFs included Pfnek-1, PfPKB, PfPI₃K, a leucine zipper kinase, PfcGAK and Pfnek-3. In consultation with Prof. C. Doerig, PfcGAK and PfNek3 were arbitrarily chosen for this study.

3.1.1 *P. falciparum* cyclin G associated kinase

A *P. falciparum* 1614 bp open reading frame, coding for a putative 538 amino acid homologue of cyclin G associated kinase (PfcGAK) was identified in unfinished sequence data of chromosome 12 located at the NCBI malaria site (www.ncbi.nlm.gov/malaria) in 2000 and is shown in Figure 3.1.

```
157 atgttgaaatttcgaagtatttgtactactttaatcgggtggcaag
    M L K F R S I C T T L I G G K
202 gtttacaatattaatgggaagacgataagagaagagaaattaatt
    V Y N I N G K T I R E E K L I
247 tctgagggagcgtattcatttggtttatatggcgaaggatttaaac
    S E G A Y S F V Y M A K D L N
292 acgaatcgagcttataactataaagaagaccatatgtcaagataaa
    T N R A Y T I K K T I C Q D K
337 gagaaaatagatatggcgaataaagaaatcaacattccttaaagt
    E K I D M A N K E I N I L K S
382 ttacctccacataaaaatatagtacaatattatggttccactatt
    L P P H K N I V Q Y Y G S T I
427 atttctgagaataattataaaaattggttattatggtgatggagtat
    I S E N N Y K I V I M L M E Y
472 tgtgagagaggaaatttattaaatatttttcaaaaaataaagac
    C E R G N L L N I F Q K N K D
517 aagattaaagaaattcacattgtaaataatttttaaagatatatt
    K I K E I H I V N I L K D I I
562 accggtttatgttttttacataatcaagaaatccctattatacat
    T G L C F L H N Q E I P I I H
607 cgtgatatttaaattagaaaatattttatgtgataaaaatggtgta
    R D I K L E N I L C D K N G V
652 tataaaatttgtgacttctgttcacattctatatcaaattccttc
    Y K I C D F C S H S I S N S F
```

```

697 tttccaaatgatttaagtaaatacagaattatctaatttgaaagaa
   F P N D L S K S E L S N L K E
742 gaaattgagagagataaccacattaatatatagacctcccgaatta
   E I E R D T T L I Y R P P E L
787 attgatttatattctaattttgaaatatcttggaaaattgatata
   I D L Y S N F E I S W K I D I
832 tggatgggttggttgatattttatatctacttctttttaaaatccat
   W M V G C I L Y L L L F K I H
877 ccatttcaaaataataattttttatccattattaatgctagtttt
   P F Q N N N F L S I I N A S F
922 actattccatattgtacgaaatattccaaaagaattatttctatt
   T I P Y C T K Y S K R I I S I
967 cttctgatgactcttaataaaaaatcctcaaaaaagaattgatacc
   L L M T L N K N P Q K R I D T
1012 tctactcttttattttatccttgaaaattatgcagacctcaaaaaa
   S T L L F I L E N Y A D L K K
1057 tggtttgtgcatattcccgcagatatcaaaaagagggttaaccaa
   W F V H I P A D I K K R V N Q
1102 atttttgaaaaaatgaatgaacttaacatgaataataaacacact
   I F E K M N E L N M N N K H T
1147 gatccttatcgaagtttagtcatgaaaagctgatcgatgttaaaata
   D L I E V S H E K L I D V K I
1192 tcttcacaaaaatatgaaaccaaccttccgaagagtaataataat
   S S Q K Y E T N L P K S N N N
1237 ttcctcaagttttttacaaccagtaatttgagtgggtgatattttt
   F L K F F T T S N L S G D I F
1282 aagaaaaaaatcgaaaaggaggaaactaaaaaagaaatcaagcaa
   K K K I E K E E T K K E I K Q
1327 gtaggcaacaagcaagtgagcgaaacagaaaaggaaccaaaaggat
   V G N K Q V S E T E K E P K D
1372 aaattggaaaataatccagatggtaataaattacaaaatgataaa
   K L E N N P D G N K L Q N D K
1417 gatagtggtaaattacaaaatgataaagatagtggtaaattacaa
   D S G K L Q N D K D S G K L Q
1462 aatgaaaaagatagtggtaaattacaaaatgataaagatagtgggt
   N E K D S G K L Q N D K D S G
1507 aaattacaaaatgataaagatagcggttaaattacaaaatgataaa
   K L Q N D K D S G K L Q N D K
1552 gatagtggtaaattacaaaatgataaagttagtggtaaattacaa
   D S G K L Q N D K V S G K L Q
1597 aatgataaagatagtggtaaattacaaaatgataaagttagtgggt
   N D K D S G K L Q N D K V S G
1642 aaattacaaaatgatcgagatggtaataaattacaaaataatcat
   K L Q N D R D G N K L Q N N H
1687 attattgattatacagataataaaatggataccgataaaaagaat
   I I D Y T D N K M D T D K K N
1732 aaatgtgatgaaatatcagaaaatatcaagttcaggtataa 1774
      K C D E I S E N I Q V Q V*

```

Figure 3.1 The ORF of PfcGAK. The ORF found at the NCBI site during 2000 that had significant similarity to cyclin G associated kinase. The underlined red sequences were used for the design of sense and antisense strand primers.

This open reading frame was coded for by a single exon which is a common occurrence in *P. falciparum* since forty six percent of *P. falciparum* genes have no introns (Gardner *et al.*, 2002). The sequences indicated in red were used to design primers to amplify the nucleotide sequence for cloning into an expression vector.

The current full length 909 amino acid sequence of the putative PfcGAK (PFL2280w) protein in PlasmoDB version 4.4 (www.plasmoDB.org) protein is shown in Figure 3.2. The complete protein is also coded for by a single exon. The portion of the protein expressed in this study is indicated in red and blue and includes the catalytic domain at the amino terminus and is followed by approximately 240 amino acids. Although the nucleotides coding for the sequence QVQV at the 3' end were used in the primer design, this sequence is not present in the final predicted PfcGAK sequence in PlasmoDB. The nucleotide coding sequence for the first three amino acids Met-Leu-Lys were excluded from the primer design to reduce complementarity between the sense and antisense primers.

```

MLKFRSICTT LIGGKVYNIN GKTIREEKLI SEGAYSFVYM AKDLNTNRAY TIKKTICQDK 60
EKIDMANKEI NILKSLPPHK NIVQYVGSTI ISENNYKIVI MLMEYCERGN LLNIFQKNKD 120
KIKEIHIVNI LKDIITGLCF LHNQEIPPIH RDIKLENILC DKNGVYKICD FCSHSISNSF 180
FPNDLSKSEL SNLKEEIERD TTLIYRPPPEL IDLYSNFEIS WKIDIWMVGC ILYLLLFKIH 240
PFQNNNFLSI INASFTIPYC TKYSKRIISI LLMTLNKNPQ KRIDTSTLLF ILENYADLKK 300
WVFIHIPADIK KRVNQIFEKM NELNMNNKHT DLIEVSHEKL IDVKISSQKY ETNLPKSNNN 360
FLKFFTTSNL SGDIFKKKIE KEETKKEIKQ VGNKQVSETE KEPKDKLENN PDGNKLQNDK 420
DSGKLQNDKD SGKLQNEKDS GKLQNDKDSG KLQNDKDSGK LQNDKDSGKL QNDKVSGLQ 480
NDKDSGKLQN DKVSGKLQND RDGNKLQNNH IIDYTDNKMD NDKKNKSDEI SENITSSGDK 540
QNSVPDQSEN QSKSHSYNLI KIQSKDTLII DHVDRDFPSD KTNTDFSFD TLCEFEKKCM 600
NNSSNYSNDK KVAGDIKKDD CSVNIQVNN NNDVCHNDMK DVNECTNLLN LDVFEYENET 660
NDSIKKINDG DFIKQSGCDF IKQNGGDFIK PNDSNCIFAD INFNNDINLN NDINLNNDIN 720
FNNDINLNND IDFFDVIQNH DLKKKEDVLN INRNNEHDKN AITCSNEYYN LSLFYDPWNN 780
NTKDNHFFGD DDYNVNKSSF IFNNDLFDNY ELCTNNTEK STHHVKNNCM DINNNIFSNK 840
NYDEHVDKKE KNISYDYNVK SHPIYSSYKN EKRCNLNDIK QNNSSKYIMS DKNDKFSSELL 900
DDFQKCTIK

```

Figure 3.2 Amino acid sequence of the putative PfcGAK kinase (PFL2280w) in PlasmoDB version 4.4. The portion of the PfcGAK open reading frame expressed in this study includes the catalytic domain (shown in red) and the following sequence shown in blue. Black sequence denotes the uncloned portion of the protein.

3.1.2 *P. falciparum* NIMA related kinase

A 1041 base pair ORF (Figure 3.3) coding for a *P. falciparum* putative NIMA-related kinase was identified in unfinished sequence data of chromosome 12 in 2000 at the NCBI malaria web site (www.ncbi.nlm.gov/malaria). This ORF corresponds to the putative

```

326 atggtttgcatttacttgttttgttttttattattccatttgttt
   M  V C I Y L F C F L L F H L F
371 gttgttaatactcgtatgaggatattgtgtaatttgttgatagtt
   V V N T R M R I L C N L L I V
416 atttgtgagtatatgttccctttttactcttttatattgtttaaa
   I C E Y M F P F Y S F I L F K
461 tttgttaagatgttatctttattgttaccgtcaatattatcgtgt
   F V K M L S L L L P S I L S C
506 gagaagaaataaccaggtgtataagaataatggatataaatttgag
   E K K Y Q V Y K N N G Y K F E
551 accgtgttagatttcatgacatcagatagtgaaattcatttgata
   T V L D F M T S D S E I H L I
596 agatctatagaaagtgatgagatatatatatctaaggatatatgat
   R S I E S D E I Y I S K V Y D
641 atatatggtataaatgaagatgatttgaataaatatatgaatgaa
   I Y G I N E D D L N K Y M N E
686 ttatatataatgaataagttaagaaattgtgagaatattgtaaat
   L Y I M N K L R N C E N I V N
731 ataatagatttatattaagaaaaatgatacattatcttttattctt
   I I D Y I K E N D T L S F I L
776 gagttttgtaatcaaggagatttacattcagatatcttaagaaag
   E F C N Q G D L H S D I L R K
821 aaattaaataatgaaatatatacagaaagtgaatatattaatata
   K L N N E I Y T E S E I F N I
866 ttacatcaaataattaatggtttaaatataattcatcagaatgggt
   L H Q I L N G L N I I H Q N G
911 ataatacatggagattttaaaaagtactaatatatttattaaagat
   I I H G D L K S T N I F I K D
956 aataaaataaagattgggtgatttcggtatatcatcagaacaaagt
   N K I K I G D F G I S S E Q S
1001 tcaaataataattaggaacattaaattgtttaagttatgaatca
   S N N N L G T L N C L S Y E S
1046 attaaatttaagaaaacaaataaattaagtgatttatttcaagtt
   I K F K K T N K L S D L F Q V
1091 ggatgtattttatatgaacttgccactttgtcatcacccttttgt
   G C I L Y E L A T L S S P F C
1136 gctacaaatatcaatgatatgatatctctttttgaggataaaaat
   A T N I N D M I S L F E D K N
1181 tataaaagctatattatcaaaaatatatcttctatatattcacaa
   Y K S Y I I K N I S S I Y S Q
1226 aaattagtaaatgttattttctaaactattgtcttttaatacatta
   K L V N V I S K L L S L N T L
1271 gaaagattggaagtaattactaattataacttagaaacgacacaa
   E R L E V I T N Y N L E T T Q
1316 cacgtacatttagaaaatgcaaataaaaataatatcatgtataacg
   H V H L E N A N K I I S C I T
1361 gttcaataa 1369
      V Q*

```

Figure 3.3 The ORF of Pfnek-3 (PFL0080c) in PlasmDB version 4.4. The same ORF found at the NCBI site during 2000 had significant similarity to NIMA kinases. The underlined red sequences were used to design sense and antisense strand primers.

PfNek-1 gene (PFL0080c), described in PlasmoDB version 4.4, which is encoded by a single exon (www.plasmoDB.org.) Another NIMA-related kinase (PFL1370w) characterised by Dorin *et al*, 2001 has also been called Pfnek-1. In a discussion with Prof C. Doerig it was decided to call the protein in this study, (PFL0080c), Pfnek-3 to prevent confusion. The entire nucleic acid sequence, with the exception of the ATG start site, was amplified for cloning into an expression vector.

3.2 Amplification of the PfcGAK and Pfnek-3 open reading frames

PCR was carried out on *P. falciparum* DNA using *Taq* polymerase and the primers described in Appendix B1.1. Initial PCR reactions were performed to determine optimal annealing temperatures, primer and DNA concentration. These parameters were then applied to the Expand Hi-Fidelity enzyme system of Roche, which was used to amplify DNA for all cloning procedures. The PCR reactions were optimised for MgCl₂ concentrations with the Expand High Fidelity system.

3.2.1 Annealing temperature

The optimal annealing temperatures for the pQ80L and pET41a primer sets were 66°C for PfcGAK and 64°C for Pfnek-3 respectively as shown in Appendix B2. In contrast the optimal annealing temperature for the pHB6 primers sets were 64°C for PfcGAK and 66°C for Pfnek-3 respectively as shown in Appendix B1.2.1. The difference in optimal annealing temperatures of nearly identical primer sets (except for the added restriction site nucleotides) are likely due to differences in the individual PCR runs.

3.2.2 MgCl₂ concentration

Results of MgCl₂ concentration optimisation experiments are shown in Appendix B1.2.2. The High Fidelity PCR reaction with PfcGAK primers amplified a single clean band at 2.5 mM MgCl₂ for both primer sets (Panel A and C). At higher MgCl₂ concentrations (Panel A) a band of ± 150 bp, which increases in intensity with higher MgCl₂ concentrations, was present. However the concentration of this band was negligible compared to the increase in yield obtained with higher concentrations of MgCl₂ (3-4 mM). Although the yield of PfcGAK at 2.5 mM MgCl₂ was poor in comparison to the higher concentrations there were no apparent non-specific bands (300 - 700 bp) present (Panel A and C). Some primer dimers have formed at the lower MgCl₂ concentrations (Panel C) and a band of ± 100 bp was observed with MgCl₂ concentrations higher than 2.5 mM. It was therefore decided to rather

use the conservative 2.5 mM MgCl₂ concentration for the PfcGAK primer set with the option of increasing the concentration if the yield of amplicon was not sufficient for cloning. The Pfnk-3 primer set for pHB6 was optimal at 2mM MgCl₂ (Panel B) and at 3.5 mM for the pQ80L/pET41a (Panel D) primer set. At a concentration of 2.5 mM for the pHB6 primer set, (Panel B), a distinct second band of $\pm$ 900 bp was present which supported the decision to use the lower concentration of MgCl₂. At the higher MgCl₂ concentrations non-specific smears were more prominent.

Table 3.1 Final PCR parameters

	Annealing temperature	MgCl ₂ concentration	Primer concentration	DNA concentration
PfcGAK (pHB6)	64° C	2.5 mM	300 nM	300 ng
PfcGAK (pQ80L; pET41a)	66° C	2.5 mM	300 nM	300 ng
Pfnk-3 (pHB6)	66° C	2.0 mM	300 nM	300 ng
Pfnk-3 (pQ80L;pET41a)	64° C	3.5 mM	300 nM	300 ng

3.2.3 PCR for cloning

Since the *P. falciparum* genome is very AT-rich, PCR control reactions to check for non-specific annealing of primers were performed using the sense or antisense primers individually. These results are shown in Figure 3.4.

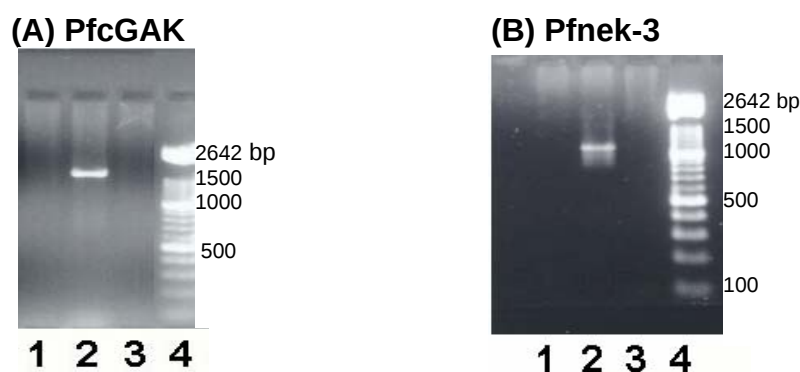


Figure 3.4 PCR products obtained from *P. falciparum* DNA using PfcGAK and Pfnk-3 primers separated on a 1.5% agarose gel. 5 μ l samples from each PCR reaction were electrophoresed.

Panel (A)

- 1= PfcGAK sense primer
- 2= PfcGAK sense and antisense primer
- 3= PfcGAK antisense primer
- 4= 100 bp DNA size marker

Panel (B)

- 1= Pfnk-3 sense primer
- 2= Pfnk-3 sense and antisense primer
- 3= Pfnk-3 antisense primer
- 4= 100 bp DNA size marker

When used individually, none of the primers amplified DNA fragments. The pHB6 primer sets produced products corresponding to the predicted sizes, namely 1629 bp for PfcGAK and 1062 bp for Pfnek-3.

Plasmodium DNA preparations obtained from *in vitro* culture may contain small amounts of human DNA. Control PCR reactions were carried out using human DNA as a template with the *Plasmodium* kinase primers. Human α -spectrin primers for exon 1 and 2 (obtained from Prof. T.L. Coetzer, Red Blood Cell Membrane Unit, NHLS, RSA) served as a positive control. No amplicons were formed with the PfcGAK and Pfnek-3 primers and human DNA as shown in Figure 3.5.

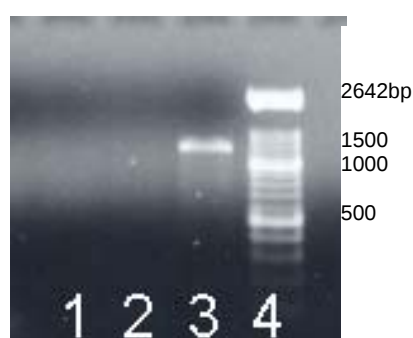


Figure 3.5 PCR products of Pfnek-3 and PfcGAK using human DNA and a spectrin positive control reaction separated on a 1.5% agarose gel. 10 μ l samples were analysed.

1 = PfcGAK 2 = Pfnek-3 3 = spectrin 4 = 100 bp DNA size marker

3.3 Cloning of PfcGAK and Pfnek-3 into expression vectors

Prior to cloning, the amplicons were partially verified to be coding for the relevant *Plasmodium* sequence by manual sequencing and BLASTN analysis (results not shown). Figure 3.6 shows an agarose gel used to estimate the concentrations of the PCR products used in the ligation reaction with pHB6. Only PfcGAK was cloned successfully into this vector despite several attempts to clone Pfnek-3. Pfnek-3 was cloned successfully into the pQ80L vector.

The PfcGAK and Pfnek-3 amplicons were cloned in a directional manner into either of the pHB6, pQ80L or the pET41a vector. The pET41a vector was used initially for cloning but

the inserts of each ORF was found to be of variable sizes in this vector, indicating possible recombination events. Therefore, further cloning using this vector was abandoned and the pHB6 or pQ80L vectors were used instead.

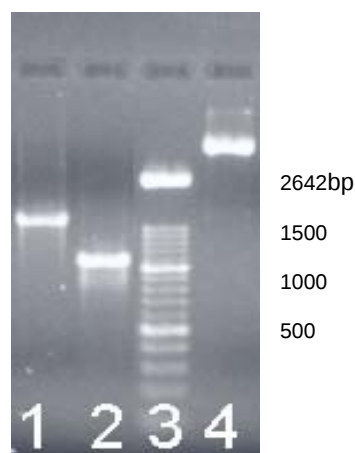


Figure 3.6 Agarose gel electrophoresis (1.5%) used to estimate the amplicon concentration used in a typical ligation reaction. 1= PfcGAK; 2 = Pfnck-3; 3 = 100 bp DNA mass ladder (5 μ l); 4 = digested pHB6. (5 μ l of the PfcGAK and Pfnck-3 PCR reactions were loaded).

3.4 Analysis of positive clones

Colonies growing on the specific antibiotic selector plates were screened for inserts either by plasmid isolation and restriction digestion, or by PCR of bacterial lysate obtained from the individual colonies.

3.4.1 Analysis of PfcGAK clones by plasmid preparation and restriction digestion

Plasmids isolated from positive PfcGAK clones in the pHB6 vector showed a band of the correct size of ± 1629 bp after restriction digestion with *Hind* III and *Eco*R1, and separation on a 2% agarose gel. These results are shown in Figure 3.7.

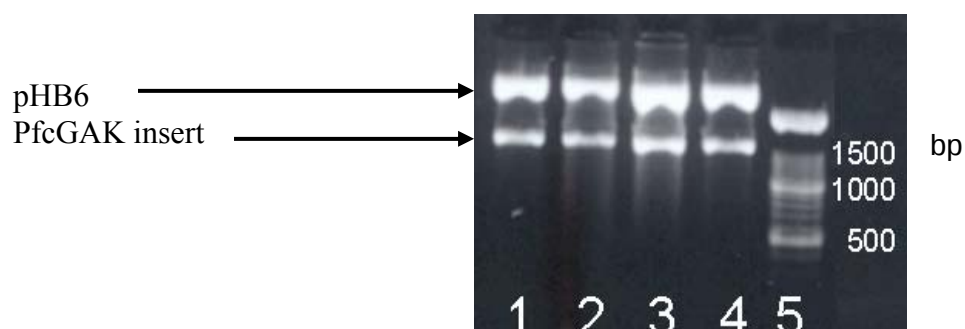


Figure 3.7 Agarose gel electrophoresis (2%) to verify the cloning of PfcGAK inserts into pHB6. Restriction digestion of plasmid isolated from PfcGAK clones with *EcoR*I and *Hind*III. Lanes 1-4 = PfcGAK clones; lane 5 = 100 bp DNA size marker. 5 μ l samples from a 20 μ l plasmid preparation, prepared from a 5 ml culture were loaded.

3.4.2. Analysis of PfcGAK clones by PCR amplification

Plasmid containing pHB6 clones were analysed for the presence of insert using the PfcGAK primers and PCR. All the clones screened showed the presence of the correct size insert as shown in Figure 3.8. A negative control without DNA was included and no bands were detected (data not shown).



Figure 3.8 Agarose gel electrophoresis (2%) of PfcGAK PCR products using pHB6 plasmid DNA as template for PfcGAK primers. 5 μ l of the PCR reaction was loaded per lane. Lanes 1-3 : PfcGAK clones, Lane 4 = 100 bp DNA size marker.

3.4.3 Analysis of Pfnek-3 clones by plasmid preparation and restriction digestion

Plasmid isolated from Pfnek-3 pQ80L colonies were digested with *Hind*III and *Bam*H1. All the colonies had inserts of the correct size as seen in Figure 3.9. Pfnek-3 clones were also screened by PCR with the same results (data not shown).

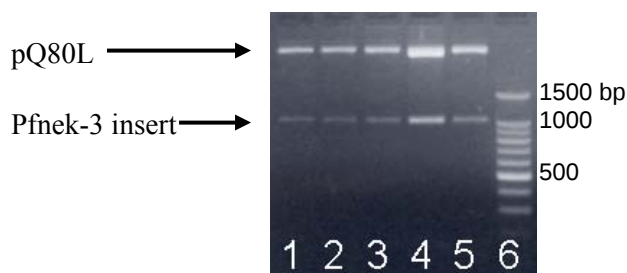


Figure 3.9 Agarose gel electrophoresis (2%) to verify the cloning of Pfnek-3 inserts into pQ80L. Restriction digestion of plasmid isolated from Pfnek-3 clones with *Hind*III and *Bam*H1. Lanes 1-5 = PfNek1 clones; Lane 6 = 100 bp DNA size marker. 5 μ l samples from a 20 μ l plasmid preparation, prepared from a 5 ml culture were loaded.

3.5 Expression of recombinant PfcGAK and Pfnek-3

It is a commonly accepted practice to verify the complete insert sequence prior to attempting expression. In this study the insert sequences were partially verified by manual sequencing and a number of clones were then screened for expression. Promising clones were then submitted to Inqaba Biotech (South-Africa) for complete sequencing of inserts. A number of expression protocols recommend screening of colonies for expression to identify clones that express protein efficiently. Due to the nature of the results, expression data are thus presented before sequencing data.

3.5.1. PfcGAK recombinant protein expression

Several pHB6 vector PfcGAK clones, with an insert of the correct size and partially verified sequence, were screened for expression of the recombinant protein (Figure 3.10). All the clones produced $\pm$ 1g (wet weight) of bacterial cells. Recombinant PfcGAK was isolated from the soluble fraction using nickel affinity chromatography. Eluates (5 μ l of a 200 μ l eluate) were separated on a SDS-12% polyacrylamide gel, transferred to PVDF membrane

and probed with an anti-histidine antibody. Thereafter the blots were stripped and reprobed with anti-HA antibody. A full-length recombinant PfcGAK protein has a haemagglutinin (HA) tag on its amino terminal and a hexa-histidine tag on its carboxyl terminal with an expected size of 65 kDa.

With the exception of clone 31 and the vector control (VC), probing the immunoblot with the anti-His antibody detected bands of the expected size as well as a number of additional bands. (Figure 3.10 panel A). When the blot was stripped and reprobed with anti-HA antibody, clones 12, 14, and 26 showed a single band of the correct size ($\pm$ 65 kDa) with the notable exception of clone 28. Compared to the other clones, clone 24 and 33 had a different banding pattern and the protein yield of clone 24 was approximately 10 times higher. (Figure 3.10 clone 24).

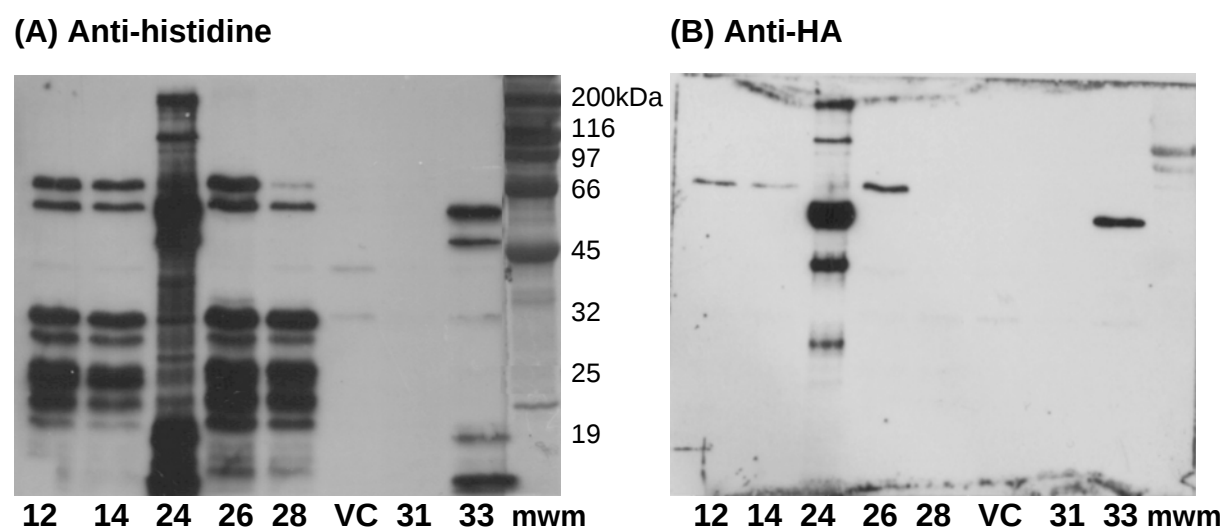


Figure 3.10 Western blots of pHb6 PfcGAK recombinant protein purified by Ni^{2+} -affinity chromatography from a number of clones, probed with anti-histidine (panel A) followed by anti-HA (panel B) antibodies. The clone numbers are indicated underneath the Western blots. VC = vector control; mwm = molecular weight marker. Amount of protein loaded: Clone 12 = 0.41 μg , Clone 14 = 0.45 μg , Clone 2.4 = 2.7 μg , Clone 26 = 0.67 μg , Clone 28 = 0.4 μg , VC = 0.54 μg , Clone 31 = 0.4 μg , Clone 33 = 0.54 μg .

Clones 12, 14, 26, and 28 all had a similar banding pattern with several hexa-histidine positive bands that ranged between 19 and 32 kDa, which were presumably degradation products (Figure 3.10). All the abovementioned clones showed the presence of a second, large hexa-histidine tagged protein of approximately 58 kDa. This protein did not react with anti-HA when the blots were stripped and reprobed. This 58 kDa protein may have been a

translation product formed from an internal initiation site. It has been reported that *E. coli* can use internal methionine residues as initiation sites in addition to the primary initiation site in expression vectors (Preibisch *et al*, 1988). If translation was initiated at Met₅₁ or Met₇₆ the resultant protein would be 5-7.5 kDa shorter (which corresponds to the size of the 58 kDa protein observed for clones 12, 14, 26 and 28) and would lack the N-terminal HA tag.

Clone 24, on the Anti-HA Western blot (Figure 3.10 panel B), showed the presence of several protein bands with sizes of $\pm$ 200-, 89-, 58-, 45.5- and 29 kDa. The 200, 89 and the most prominent 58 kDa bands can also be seen on the anti-His blot as well as a second prominent $\pm$ 51 kDa protein product, which was not present on the anti-HA blot. The two larger proteins of 89.5 and 200 kDa may have been translational read-through products since they have both hexa-histidine and HA tags present. It is possible the $\pm$ 51 kDa protein of clone 24 on the anti-His blot was the equivalent of the smaller 58 kDa protein produced by clones 12, 14, 26 and 28. The two smaller bands of approximately 40 and 27 kDa on the anti-HA blot were probably either degradation products or truncated translation products since they did not react with the anti-histidine probe.

The anti-His blot also showed the presence of protein products smaller than 20 kDa which may have been degradation products although a protease inhibitor cocktail was present throughout the purification procedure. It is possible that these products may have formed during the induction of expression in *E.coli*. (Baneyx, 1999). It has also been shown that *P.falciparum* protein expression can be initiated at internal initiation ATG sites of the cloned ORF when expressed in *E.coli* (Mehlin *et al.*, 2006). These truncated proteins will lack the amino terminal HA tag but would still have the carboxyl terminal hexa-histidine tag. With the exception of clone 24, there was no smaller protein products present on the anti-HA blot. (See table 4.7 in the discussion).

The stringency of the isolation procedure was increased by increasing the concentration of imidazole in the washing buffer to 60 mM. This did not make a difference to the banding patterns or number of bands observed for the different clones. The difference in protein yield of PfcGAKclone24 and the rest of the clones was intriguing since it is known that many *Plasmodium* proteins do not express at high levels in *E. coli*. Recombinant PfcGAK bound to 50 μ l Ni²⁺-agarose beads was initially eluted with three 200 μ l volumes of a 250

mM imidazole/ Tris buffer, pH 7.5. This was not sufficient to completely remove the protein from the beads and the concentration of imidazole was therefore increased to 500 mM.

3.5.2 Determination of the nucleic acid sequences of the inserts of pHB6

PfcGAK clone 14 and 24

The discrepancy between the sizes of the proteins which reacted positively with both anti-HA and anti-His antibody prompted the full length automated sequencing of PfcGAK clones 14 and 24 using internal as well as vector primers. The full length open reading frame of the cloned PfcGAK construct showing the various primer sites is shown in Appendix B2. Sequence chromatograms of both clones are shown in Appendix E1.

The nucleic acid sequences of clone PfcGAK14 and clone PfcGAK24 were aligned with the PfcGAK open reading frame in PlasmODB using the DNA Assist programme. The alignment showed a deletion of 243 bp from clone 24. This deletion had a 27 bp sequence (AAAGATAGTGGTAAATTACAAAATGAT/A) that was repeated nine times, hence the deletion maintained the reading frame. The sequence chromatograms indicating the sites where the deletion occurred are shown in Figure 3.11(a) and 3.11(b).

Apart from the in frame deletion of the repeat region, PfcGAK clone 24 had one silent mutation at position 1071 where a thymidine was replaced by a cytosine. One silent mutation was also observed for PfcGAK clone 14 at position 949, where thymidine was again replaced by cytosine. Since these mutations were silent they caused no change in the amino acid sequence. These mutation sites are shown in Figure 3.12.

The DNA Assist program was used to analyse the nucleic acid sequence and align the various sequences. Appendix B5 shows the nucleic acid alignment of PfcGAK clone 14, 24 and the predicted PfcGAK sequence in PlasmODB version 4.4. All the mutation sites are indicated in Appendix B5 and identical nucleic acids are shown in shaded text. The sequences of the clones were then translated into amino acid sequences that were compared to that of the predicted PfcGAK amino acid sequence in PlasmODB. The alignment is shown in Figure 3.13. The amino acid sequence KDA/VSGKLQND/E is repeated nine times in the predicted open reading frame. PfcGAK clone 24 lacked this sequence whereas it was present in PfcGAK clone 14. The absence of this repeat region appears to have enhanced the expression of the protein since clone PfcGAK24 generated expressed protein to a ± 10 times higher level than the clones containing the full sequence.

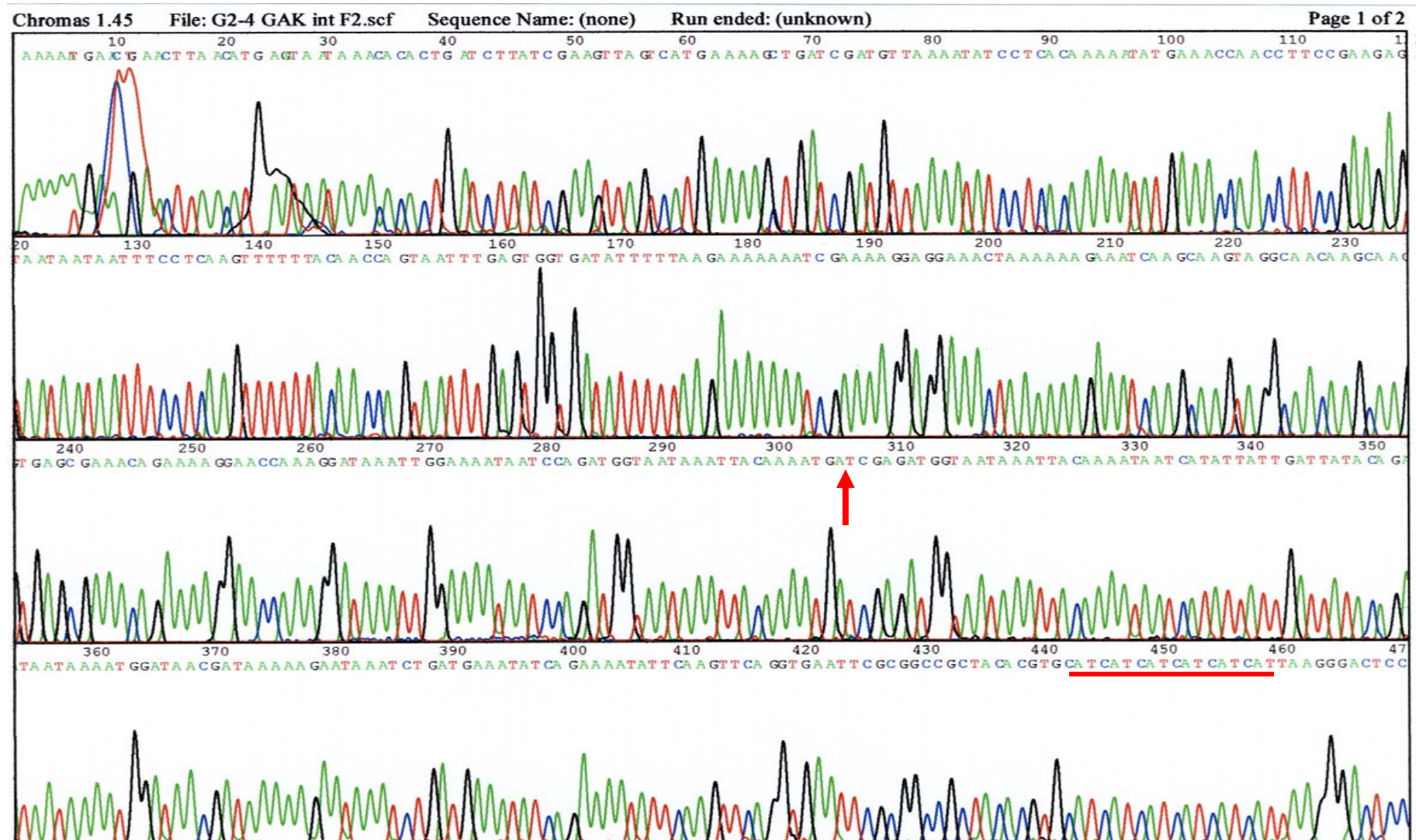


Figure 3.11(a) Partial sequence chromatogram of clone PfcGAK24 indicating the deletion site. Red underlined sequence = hexa-histidine tag, red arrow = deletion site

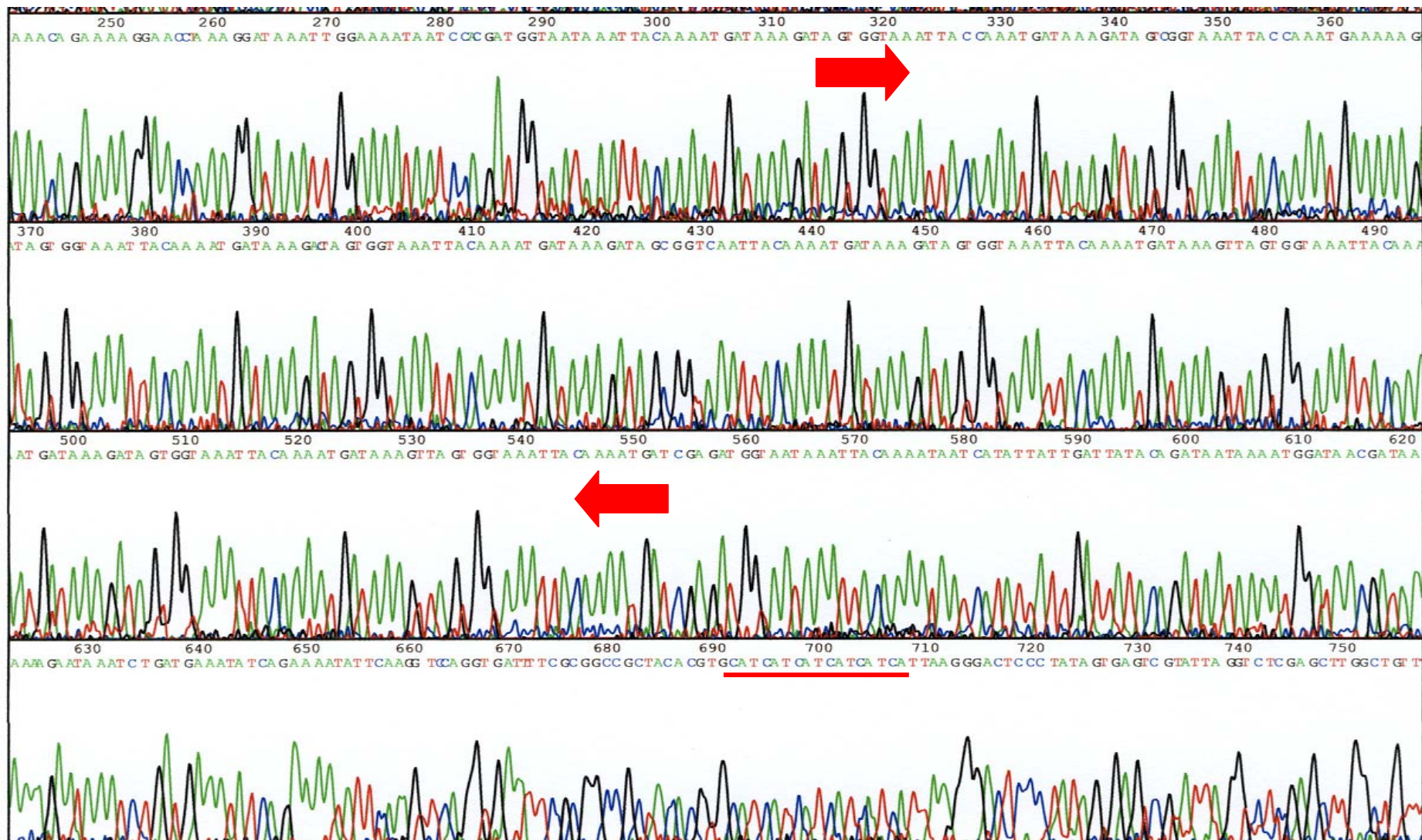


Figure 3.11(b) Partial sequence chromatogram of PfcGAK clone 14 showing the normal sequence. Arrows indicate the deleted sequence in clone 24 and the underlined sequence codes for the hexa-histidine tag.

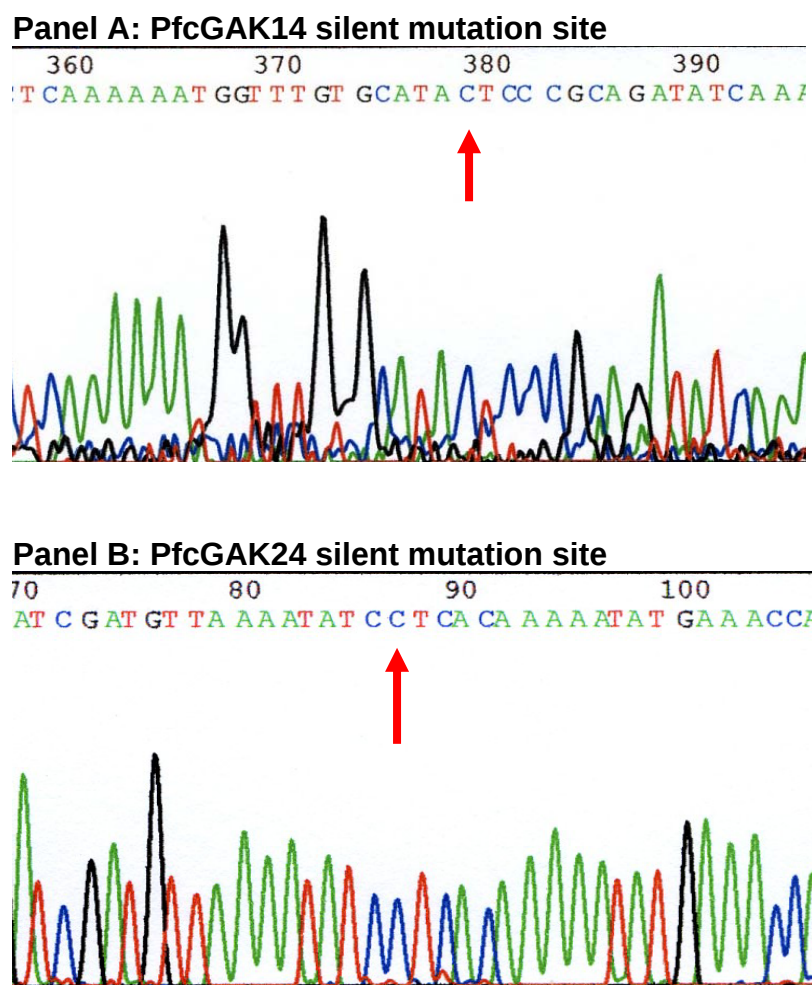


Figure 3.12 Portions of the sequence chromatograms showing the silent T→C mutations of PfcGAK14 and PfcGAK24. The mutated residues are indicated by the red arrows.

Sequencing of the inserts of the two PfcGAK clones showed that they were in the correct reading frame in the plasmid, which correlated with the presence of both N and C terminal tags on the expressed proteins. Clone PfcGAK14 had the normal nucleic acid sequence and the size of the protein correlated with the predicted size, 65 kDa (see Table 3.2). Protein produced by PfcGAK24 was smaller than expected ($\pm$ 51kDa) due to the in-frame deletion of 81 amino acids.

Table 3.2 provides a summary of some of the predicted physical properties of the putative PfcGAK protein compared to the recombinant PfcGAK proteins. The molecular weights and iso-electric points were predicted using the DNA-Assist program. The information for the “PfcGAK cloned fragment” excludes the additional amino acids acquired by the cloning vector sequence and the hexa-histidine tag.

Table 3.2 Predicted properties of PfcGAK recombinant proteins compared to the predicted properties of the native protein.

	PfcGAK (full sequence)	PfcGAK cloned fragment*	PfcGAK clone 24	PfcGAK clone 14**
Number of amino acids	909	531	418	562
Molecular weight (kDa)	106	61	56	65
Iso-electric point	5.7	8.8	8.6	8.7

* Refers to the predicted sequence in PlasmoDB version 4.4 without any additional amino acids.

** Refers to the sequence with additional amino acids formed by the extra nucleotides required for cloning.

PfcGAKclone24	1	MGYPYDVPDYAGSL	FRSICTTLIGGKVYNINGKTIREEKLISEGAYSFVYMAKDLNTN	58
PfcGAKclone14	1	MGYPYDVPDYAGSL	FRSICTTLIGGKVYNINGKTIREEKLISEGAYSFVYMAKDLNTN	58
PfcGAK	1	MLK	FRSICTTLIGGKVYNINGKTIREEKLISEGAYSFVYMAKDLNTN	47
PfcGAKclone24	59	RAYTIKKTICQDKEKIDMANKEINILKSLPPHKNIVQYYGSTIISENNYKIVIMLMEY		116
PfcGAKclone14	59	RAYTIKKTICQDKEKIDMANKEINILKSLPPHKNIVQYYGSTIISENNYKIVIMLMEY		116
PfcGAK	48	RAYTIKKTICQDKEKIDMANKEINILKSLPPHKNIVQYYGSTIISENNYKIVIMLMEY		105
PfcGAKclone24	117	CERGNLLNIFQKNKDKIKEIHIVNILKDIITGLCFLHNQEIPIIHRDIKLENILCDKN		174
PfcGAKclone14	117	CERGNLLNIFQKNKDKIKEIHIVNILKDIITGLCFLHNQEIPIIHRDIKLENILCDKN		174
PfcGAK	106	CERGNLLNIFQKNKDKIKEIHIVNILKDIITGLCFLHNQEIPIIHRDIKLENILCDKN		163
PfcGAKclone24	175	GVYKICDFCSHSISNSFFPNDSLKSELNLKEEIERDTTLIRPPELIDLYSNFEISW		232
PfcGAKclone14	175	GVYKICDFCSHSISNSFFPNDSLKSELNLKEEIERDTTLIRPPELIDLYSNFEISW		232
PfcGAK	164	GVYKICDFCSHSISNSFFPNDSLKSELNLKEEIERDTTLIRPPELIDLYSNFEISW		221
PfcGAKclone24	233	KIDIWMVGCILYLLLFKIHFPQNNNFLSIINASFTIPYCTKYSKRIISILLMTLNKNP		290
PfcGAKclone14	233	KIDIWMVGCILYLLLFKIHFPQNNNFLSIINASFTIPYCTKYSKRIISILLMTLNKNP		290
PfcGAK	222	KIDIWMVGCILYLLLFKIHFPQNNNFLSIINASFTIPYCTKYSKRIISILLMTLNKNP		279
PfcGAKclone24	291	QKRIDTSTLLFILENYADLKKWFVHIPADIKKRVNQIFEKMNELNMNNKHTDLIEVSH		348
PfcGAKclone14	291	QKRIDTSTLLFILENYADLKKWFVHIPADIKKRVNQIFEKMNELNMNNKHTDLIEVSH		348
PfcGAK	280	QKRIDTSTLLFILENYADLKKWFVHIPADIKKRVNQIFEKMNELNMNNKHTDLIEVSH		337
PfcGAKclone24	349	EKLIDVKISSQKYETNLPKSNNNFLKFFTTSNLSGDIFKKKIEKEETKKEIKQVGNKQ		406
PfcGAKclone14	349	EKLIDVKISSQKYETNLPKSNNNFLKFFTTSNLSGDIFKKKIEKEETKKEIKQVGNKQ		406
PfcGAK	338	EKLIDVKISSQKYETNLPKSNNNFLKFFTTSNLSGDIFKKKIEKEETKKEIKQVGNKQ		395
PfcGAKclone24	407	VSETEKEPKDKLENNPDGNKLQND	-----	430
PfcGAKclone14	407	VSETEKEPKDKLENNPDGNKLQND	KDSGKLQNDKDSGKLQNEKDSGKLQNDKDSGKLQ	464
PfcGAK	396	VSETEKEPKDKLENNPDGNKLQND	KDSGKLQNDKDSGKLQNEKDSGKLQNDKDSGKLQ	453
PfcGAKclone24	431	-----	RDGNKLQNNHI	441
PfcGAKclone14	465	NDKDSGKLQNDKDSGKLQNDKVSGLQNDKDSGKLQNDKVSGLQND	RDGNKLQNNHI	522
PfcGAK	454	NDKDSGKLQNDKDSGKLQNDKVSGLQNDKDSGKLQNDKVSGLQND	RDGNKLQNNHI	511
PfcGAKclone24	442	IDYTDNKMNDKKNKSDEISENI	QVQVNSRPLVHHHHHH	481
PfcGAKclone14	523	IDYTDNKMNDKKNKSDEISENI	QVQVNSRPLVHHHHHH	562
PfcGAK	512	IDYTDNKMNDKKNKSDEISENI	TSSGDKQNSVPDQSENQ	551

Figure 3.13 Amino acid sequence alignment of PfcGAK expressed from clone 24 and clone 14 with the PlasmoDB version 4.4 PfcGAK amino acid sequence. Blue sequence = vector sequence, shaded sequence = identical aligned sequence; Red sequence = deleted nanomeric repetitive region in clone 24.

3.5.3 Pfnek-3 recombinant protein expression

Hexa-histidine tagged Pfnek-3 proteins were isolated from the soluble fraction of positive clones and analysed by immunoblotting. Unlike pHB6, which has N- and C-terminal tags the pQ80L vector only adds a hexa-histidine tag to the N-terminus of the recombinant protein and can therefore only be probed with anti-histidine antibodies. The expected size of the recombinant protein was 42 kDa.

A number of clones were screened for expression as shown in Figure 3.14. The Ni^{2+} -purified Pfnek-3 protein samples of clones obtained from PCR reaction 1 (Lane 1 and 2) showed no expressed protein, while clones 1 to 4 obtained from PCR reaction 2 showed the presence of a protein of the correct size of approximately 42 kDa. PCR reaction 1 and 2 were two different PCR reactions carried out at the same time. Lanes 4-6 in Figure 3.14 show the presence of additional faint bands, substantially larger than 50 kDa. This may have been caused by a failure of the transcription process to terminate at the stop codon or may have been caused by non-specific binding of the anti-histidine antibody.

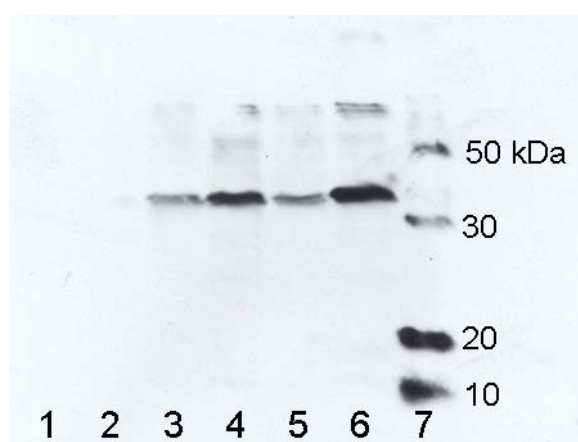


Figure 3.14 Western blot of pQ80L Pfnek-3 recombinant protein, purified by Ni^{2+} affinity chromatography from six clones, probed for expression by anti-histidine antibody. Lane 1 and 2 = clones from PCR 1; Lanes 3, 4, 5, 6 = clone 1-4 from PCR 2 ; Lane 7 = Histidine-tagged molecular weight marker proteins. Amount of protein loaded: Lane 1 = 0.35 μg , Lane 2 = 0.27 μg , Lane 3 = 0.22 μg , Lane 4 = 0.3 μg , Lane 5 = 0.25 μg , Lane 6 = 0.33 μg .

3.5.4 Determination of the DNA sequence of the PfNek-3 insert

The sequences of the two Pfnek-3 clones from PCR 2 (2 and 4) were determined by automated sequencing (Inqaba Biotech, RSA) using vector primers. Appendix B6 shows the alignment of the nucleotide sequence of Pfnek-3 with that of the cloned inserts of Pfnek-3(2) (lane 4 in Figure 3.14) and Pfnek-3(4) (lane 6 in Figure 3.14). The ORF was inserted in the correct reading frame in relation to the ATG start site, and the hexa-histidine tag coding sequence. Sequence chromatograms are shown in appendix E2.

No mutations were observed in Pfnek-3 clone 2 (Pfnek-3(2)) and this recombinant protein demonstrated kinase activity (Fig 3.20 and Table 3.5). One mutation was present in Pfnek-3 (4), guanosine replaced the cytosine at position 240 and is shown in the sequence chromatogram of Figure 3.15. This mutation caused an amino acid substitution in which the Phe₈₁ residue was replaced with a leucine residue. Phenylalanine and leucine are both non-polar amino acids. Since this mutation occurs in subdomain I of the kinase catalytic site in the region of the glycine triad it may have implications for substrate binding. (See Figure 3.30). Despite several attempts no kinase activity could be demonstrated for Pfnek3 clone 4, confirming the importance of subdomain I for activity (See Figure 3.20 and table 3.5).

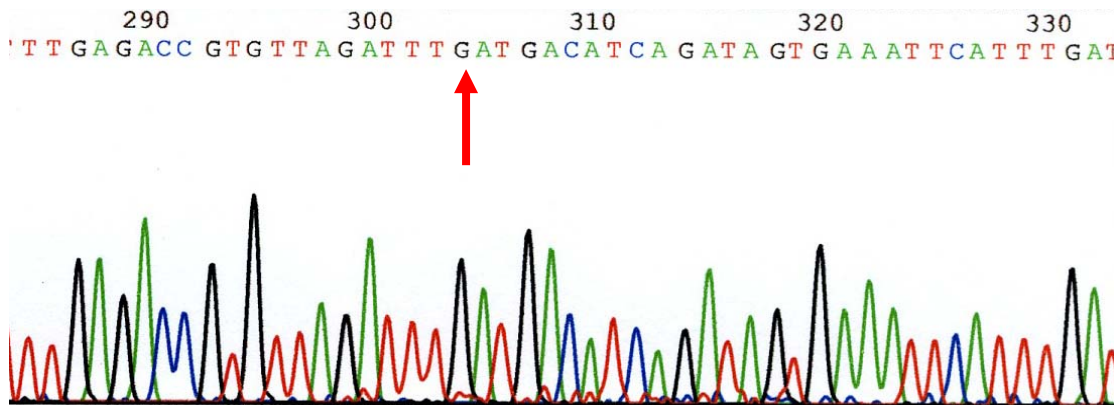


Figure 3.15 A portion of the sequence chromatogram of Pfnek-3(4) showing the cytosine to guanosine mutation. The mutated residue is indicated by a red arrow.

An alignment of the translated amino acid sequences of Pfnek-3 with that of clone 2 and 4 is shown below in Figure 3.16. Table 3.3 shows the predicted molecular weight and the isoelectric point of the putative Pfnek-3 protein in PlasmoDB version 4.4 and that of the cloned

protein with the additional vector derived hexa-histidine residues. The additional histidine residues caused an increase in the isoelectric point from 5.8 to 6.4.

Pfnek-3	1	MVCIYLCFLLFHLFVNTRMRILCNLLIVICEYMFPFYSFILFK	45
Pfnek-3(2)	1	HHHHHHGSMVCIYLCFLLFHLFVNTRMRILCNLLIVICEYMFPFYSFILFK	52
Pfnek-3(4)	1	HHHHHHGSMVCIYLCFLLFHLFVNTRMRILCNLLIVICEYMFPFYSFILFK	52
Pfnek-3	46	FVKMLSLLPSILSCEKKYQVYKNNGYKFETVLD	97
Pfnek-3(2)	53	FVKMLSLLPSILSCEKKYQVYKNNGYKFETVLD	104
Pfnek-3(4)	53	FVKMLSLLPSILSCEKKYQVYKNNGYKFETVLD	104
Pfnek-3	98	IYISKVYDIYGINEDDLNKYMNELYIMNKLRCENIVNIIDYIKENDTLSFI	149
Pfnek-3(2)	105	IYISKVYDIYGINEDDLNKYMNELYIMNKLRCENIVNIIDYIKENDTLSFI	156
Pfnek-3(4)	105	IYISKVYDIYGINEDDLNKYMNELYIMNKLRCENIVNIIDYIKENDTLSFI	156
Pfnek-3	150	LEFCNQGDHSDILRKKLNNEIYTESEIFNILHQILNGLNIIHQNGIIHGDL	201
Pfnek-3(2)	157	LEFCNQGDHSDILRKKLNNEIYTESEIFNILHQILNGLNIIHQNGIIHGDL	208
Pfnek-3(4)	157	LEFCNQGDHSDILRKKLNNEIYTESEIFNILHQILNGLNIIHQNGIIHGDL	208
Pfnek-3	202	KSTNIFIKDNKIKIGDFGISSEQSSNNNLGTLNCLSYESIKFKKTNKLSDLF	253
Pfnek-3(2)	209	KSTNIFIKDNKIKIGDFGISSEQSSNNNLGTLNCLSYESIKFKKTNKLSDLF	260
Pfnek-3(4)	209	KSTNIFIKDNKIKIGDFGISSEQSSNNNLGTLNCLSYESIKFKKTNKLSDLF	260
Pfnek-3	254	QVGCILYELATLSSPFCATNINDMISLFEDKNYKSYIKNISSIYSQKLVNV	305
Pfnek-3(2)	261	QVGCILYELATLSSPFCATNINDMISLFEDKNYKSYIKNISSIYSQKLVNV	312
Pfnek-3(4)	261	QVGCILYELATLSSPFCATNINDMISLFEDKNYKSYIKNISSIYSQKLVNV	312
Pfnek-3	306	ISKLLSLNTLERLEVITNYNLETTQHVHLENANKIISCITVQ	347
Pfnek-3(2)	313	ISKLLSLNTLERLEVITNYNLETTQHVHLENANKIISCITVKL	355
Pfnek-3(4)	313	ISKLLSLNTLERLEVITNYNLETTQHVHLENANKIISCITVKL	355

Figure 3.16 The amino acid sequence alignment of Pfnek-3 with clone 2 and 4.

Blue sequence = indicates the hexa-histidine tag and two additional amino acids supplied by the cloning vector, shaded sequence = identical aligned sequence, white text (shaded) = mutation site (Phe→Leu)

Table 3.3 Predicted properties of Pfnek-3 recombinant protein compared to the predicted properties of the native protein

	Putative Pfnek-3	Recombinant Pfnek-3
Number of amino acids	347	359
Molecular weight (kDa)	40	42
Iso-electric point	5.8	6.4

3.6 Kinase assays

3.6.1 Kinase activity of PfcGAK

Kinase assays were carried out using casein, myelin basic protein (MBP) and histone (H1) protein as substrates. MBP and H1 were combined in the same assay tubes since their molecular weights have a substantial difference. Commercial casein preparations have some contaminants that may cause phosphorylation, therefore a casein control reaction without recombinant enzyme was included.

Casein was phosphorylated by recombinant PfcGAK as seen in Figure 3.17. There was no phosphorylation of MBP or H1 protein. With the exception of lanes 2 and 7, $\pm 0.5 \mu\text{g}$ of Ni^{2+} -purified recombinant protein was used in kinase assays. Recombinant protein was omitted from the control reactions as seen in lane 1 and lane 6 for the respective substrates. The autoradiograph shows phosphorylation of higher molecular weight species especially in lane 3, which may have been phosphorylation of diffuse casein aggregates. The recombinant PfcGAK used in lane 2 and 7 was a freshly prepared Ni^{2+} -affinity chromatography isolate from clone PfcGAK24, of unknown concentration. The Coomassie stained gel shows a number of additional bands present in the freshly prepared PfcGAK24 that were not present in the other preparations. These proteins may have caused the inhibition of kinase activity since no casein phosphorylation was observed in this reaction (lane 2).

The other PfcGAK samples used in this assay (Figure 3.17) had been stored at 4°C in 30% glycerol for 21 days prior to the kinase assay. It is possible that the proteins folded into the correct tertiary structure during this period. *E. coli* is known to cause misfolding of heterologously expressed eukaryotic proteins (Baneyx, 1999). In this assay, absence of the nanomeric repeat region (lane 4, clone 24) did not cause a complete loss of kinase activity.

To correct for loading discrepancies and background incorporation of radioactivity by casein, densitometric analysis of the Coomassie stained gel and corresponding autoradiograph was performed. The data in table 3.4 indicate that the recombinant protein obtained from PfcGAK14 and PfcGAK 24 had kinase activity, whereas recombinant protein obtained from PfcGAK26 was not able to phosphorylate casein.

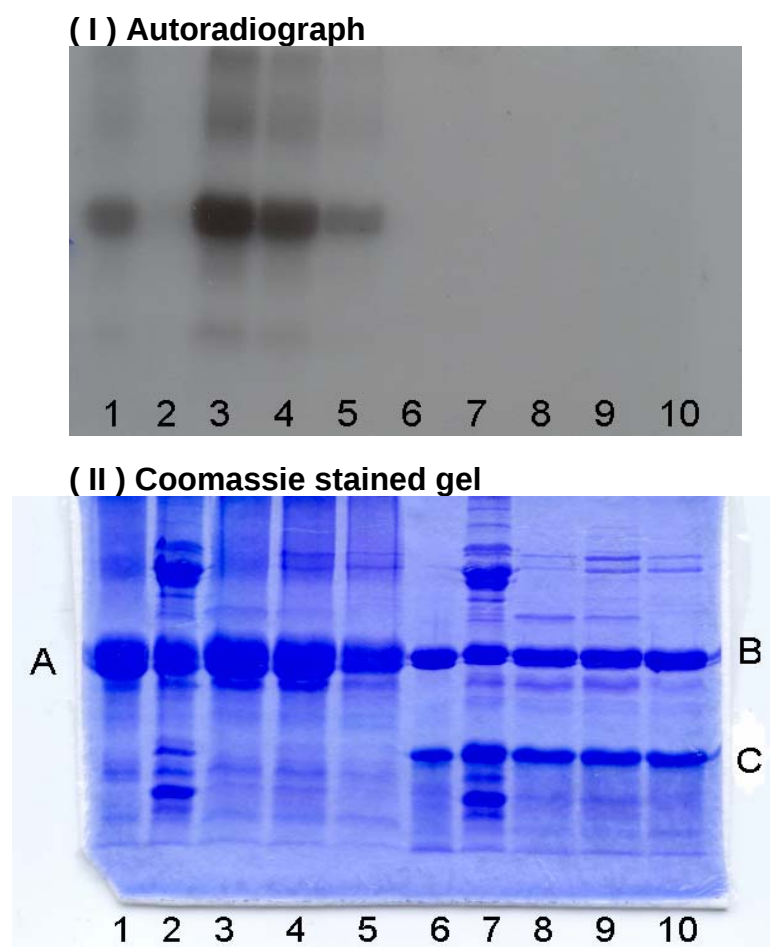


Figure 3.17 Kinase assay of PfcGAK using different substrates (A) casein (lane 1-5) (B) myelin basic protein and (C) histone protein (lane 6-10) with different isolates of PfcGAK. Panel (I): autoradiograph of the Coomassie stained gel (12% SDS-PAGE) (II) of the kinase reaction mixtures. Lane 1 = casein control; Lanes 2,7 = PfcGAK clone 24*; Lanes 3,8= 0.48 μ g PfcGAK clone 14; Lanes 4, 9 = 0.52 μ g PfcGAK clone 24; Lanes 5,10 = 0.49 μ g PfcGAK clone 26 ; Lane 6 = MBP and H1 control. * = freshly prepared protein. All preparations except lanes 2,7 were stored in 30% glycerol for 21 days.

Table 3.4 Densitometric analysis of Coomassie stained gel and autoradiograph from Figure 3.17

Sample name	No	Coomassie stained gel (Peak area)	Autoradiograph (Peak area)	Autoradiograph/ Coomassie gel	Relative intensity	Kinase activity
Casein control	1	101 812	34795	0.342	1.0	No
Fresh PfcGAK24	2	82 811	No peak	Not determined	ND	No
PfcGAK14	3	93 868	95085	1.013	2.96	Yes
PfcGAK24	4	90 058	59 675	0.663	1.94	Yes
PfcGAK26	5	53 403	21641	0.394	1.15	No

A small possibility exists that the casein phosphorylation seen in lanes 3 and 4 may have been caused by PfcGAK stimulating the phosphorylating activity of the casein contaminants.

PfcGAK recombinant proteins have inconsistent kinase activity as seen in Figure 3.18. Recombinant PfcGAK enzymes from two different preparations that had been stored in 30% glycerol were used in the kinase assay in Figure 3.18 and are referred to as “old isolate 1” and “old isolate 2”. Despite the fact that $\pm 7 \mu\text{g}$ of PfcGAK24 protein were used in the kinase assay (lane 3), minimal kinase activity was observed. The activity of the Pfnek-1 and Pfnek-4 clones used in Figure 3.18, will be discussed in section 3.6.2.1.

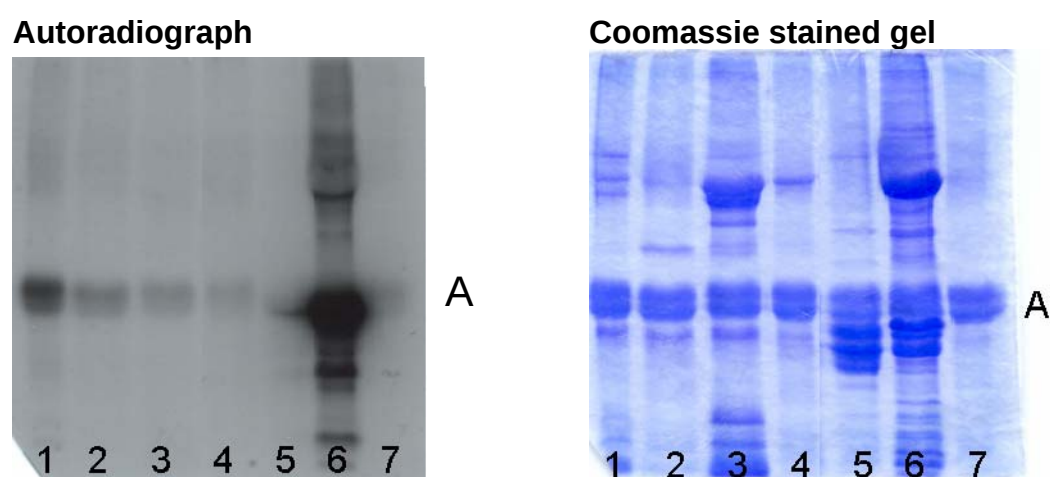


Figure 3.18 Casein kinase activity of different recombinant PfcGAK isolates and the GST-fusion proteins Pfnek-1 and Pfnek-4. Lane 1 = $0.58 \mu\text{g}$ PfcGAK14, (old isolate 1); Lane 2 = $0.47 \mu\text{g}$ PfcGAK14, (old isolate 2); Lane 3 = $6.85 \mu\text{g}$ PfcGAK24 (old isolate 1); Lane 4 = $0.9 \mu\text{g}$ PfcGAK24, (old isolate 2); Lane 5 = Pfnek-1, $1.9 \mu\text{g}$; Lane 6 = Pfnek-4, $6.2 \mu\text{g}$; Lane 7 = Casein control, A = casein bands; old isolate 1 & 2 are two different preparations of PfcGAK, 14 days old. The kinase assay mixtures were separated by 12% SDS-PAGE.

In an attempt to optimise the conditions, kinase assays were incubated for a longer time period. When the PfcGAK protein used in the above gel (Figure 3.18 lane 1) was incubated for 45 minutes instead of the normal 30 minutes, it displayed autophosphorylation activity that had not been previously observed (Figure 3.19). However, this did not improve the casein kinase activity, which was still minimal after 45 minutes. The reaction was carried out in triplicate to determine the reproducibility of the reactions.

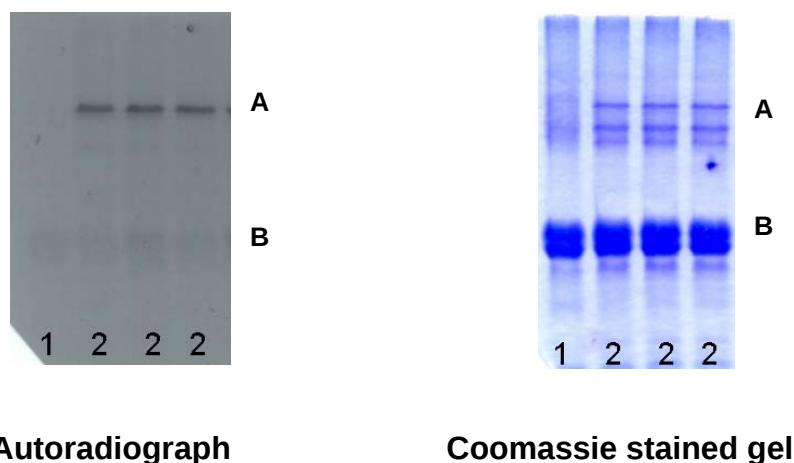


Figure 3.19 Prolonged incubation of kinase reaction mixture caused auto-phosphorylation of PfcGAK14. Legend: 1 = casein control reaction ; 2 = PfcGAK14 (fresh isolate, 0.8 μ g) ; A = phosphorylated PfcGAK ; B = casein. The kinase reaction mixtures were separated on 12% SDS-PAGE.

3.6.2 Kinase activity of recombinant Pfnek enzymes

3.6.2.1 Kinase activity of Pfnek-1 and Pfnek-4

Clones of two other Pfnek homologues, Pfnek-1 and Pfnek-4, both GST-fusion proteins, were a gift from Prof. C. Doerig, INSERM U 609, Wellcome Laboratory for Molecular Parasitology, University of Glasgow. Pfnek-4 has excellent kinase activity against casein as seen in Figure 3.18 and was included as a kinase positive control reaction. Although more Pfnek-4 protein (6.2 μ g) than Pfnek-1 protein (1.9 μ g) was used in the kinase assay, Pfnek-1 showed no kinase activity. Pfnek-1 is reported to be active against casein (Dorin *et al*, 2001) but this specific protein isolate did not show any activity towards casein.

3.6.2.2 Kinase activity of Pfnek-3

Pfnek-3 recombinant protein from several clones was screened for kinase activity. Figure 3.20 shows the casein kinase activity of recombinant Pfnek3. Kinase activity was only observed for Pfnek3(2), while the other reactions did not differ from the background phosphorylation of casein. This was confirmed by densitometric analysis which is presented

in Table 3.5. The clones were also tested for kinase activity using myelin basic and histone H1 protein as substrates, but no kinase activity was observed. (Results not shown).

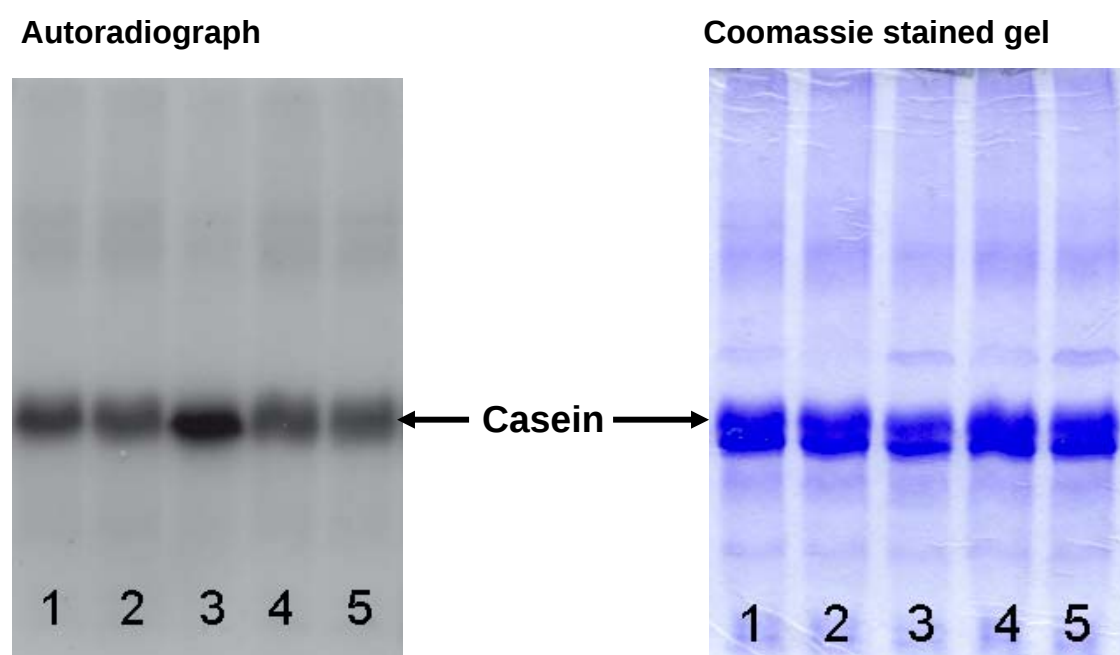


Figure 3.20 Casein kinase activity of different recombinant Pfnek-3 isolates.

Lane 1 = 0.39 μ g Pfnek3 clone 1; Lane 2 = 0.34 μ g Pfnek3 clone 3; Lane 3 = 0.31 μ g Pfnek3 clone 2; Lane 4 = 0.25 μ g Pfnek3 clone 4; Lane 5 = vector/casein control. Kinase reaction mixtures were separated by 12% SDS-PAGE

Table 3.5 Densitometric analysis of Coomassie stained gel and autoradiograph from Figure 3.20

Sample	No	Coomassie stained gel (Peak area)	Autoradiograph (Peak area)	Autoradiograph / Coomassie stain	Relative intensity	Kinase activity
Pfnek3(1)	1	34978	11 142	0.318	1.2	No
Pfnek3(3)	2	34701	10 565	0.304	1.14	No
Pfnek3(2)	3	32579	18 842	0.578	2.18	Yes
Pfnek3(4)	4	38822	11 107	0.286	1.07	No
Casein	5	36133	9 610	0.265	1	No

Subsequent protein isolates of Pfnek-3 clones failed to display any kinase activity. The lack of consistent kinase activity for both PfcGAK and Pfnek-3 recombinant protein currently makes it impossible to use them in screening assays for kinase inhibitors. In summary: the

two putative *Plasmodium* kinases PfcGAK and Pfnck-3 were expressed as recombinant proteins and their sequences verified. Both recombinant proteins displayed kinase activity towards casein. The proteins require a comprehensive study to determine the optimum purification conditions as well as cofactor requirements, pH and temperature optima and kinetic parameters, of the enzymes.

3.7 Bio-informatic analysis of PfcGAK

Bio-informatic analysis of predicted gene sequences and their corresponding proteins supply clues regarding the possible function of the putative protein as well the possible cellular location, protein interaction sites, and regulatory sites. This analysis is based on structural similarities between the putative proteins and characterised proteins with known functions. It is important to note that information obtained from bioinformatic analysis ultimately has to be confirmed experimentally.

3.7.1 Blastp analysis of PfcGAK

The putative PfcGAK protein in PlasmoDB version 4.4 was analysed using the WU-2 Blastp program at the European Bioinformatics Institute (EBI) (www.ebi.ac.uk). A summary of the first 30 hits is presented in table 3.4. The complete summary table of the PfcGAK Blastp is shown in appendix B.3. Blastn analysis is not included since nucleic acid searches of the non-redundant database only tend to identify similar *Plasmodium* proteins as a result of the AT-richness of the *P. falciparum* genome.

The Blastp analysis of PfcGAK showed similarity to numerous putative and hypothetical serine / threonine protein kinases. More importantly, there was significant similarity to the cyclin G associated kinase (cGAK) of human, rat and mouse origin, human and mouse BIKE kinases as well as human adaptor associated kinase 1 (AAK1). BIKE is an acronym for Bone morphogenic protein-2 inducible protein kinase. BIKE kinases are thought to be involved in osteoblast differentiation (Kearns *et al.*, 2001).

AAK1 proteins are responsible for the phosphorylation of the μ 2 subunit of adaptor protein 2 (AP2) that is involved in endocytosis. AP2 is also phosphorylated by GAK (Conner & Schmid, 2003). The percentage identical amino acids for the above proteins ranged between 32 and 34, percentage positives between 55 and 58 and the E values between 6.5e-35 and 7.0e-37.

Table 3.6 Summary of PfcGAK Blastp results using WU-2 at EBI

DB:ID	Source	Length	%ID	%Pos	E ()
UNIPROT:Q8I4V7_PLAF7	Cyclin g-associated kinase, putative.	909	100	100	0.
UNIPROT:Q86HW6_DICDI	Hypothetical protein.	798	25	44	5.9e-57
UNIPROT:Q9VMY8_DROME	CG1107-PA, isoform A (Cg1107-pb, isoform b).	1165	34	55	6.5e-39
UNIPROT:Q5RJ35_BRARE	Novel protein similar to vertebrate cyclin G associated kinase (GAK) (Fragment).	845	34	58	7.3e-39
UNIPROT:Q7QG29_ANOGA	ENSANGP00000005528 (Fragment).	1158	32	57	7.9e-38
UNIPROT:Q8VYC0_ARATH	Putative serine/threonine protein kinase.	650	31	55	1.0e-37
UNIPROT:O50071_ARATH	Putative serine/threonine protein kinase (Putative serine threonine protein kinase) (Putative serine/threonine kinase).	650	31	55	1.3e-37
UNIPROT:Q8H0Z0_ARATH	Putative serine/threonine protein kinase.	670	31	55	2.2e-37
UNIPROT:Q6H5H4_ORYSA	Putative AAK1 protein.	703	31	52	4.3e-37
UNIPROT:Q5RDC0_PONPY	Hypothetical protein DKFZp469F1313.	1311	34	57	5.5e-37
UNIPROT:GAK_HUMAN	Cyclin G-associated kinase (EC 2.7.1.-).	1311	34	57	7.0e-37
UNIPROT:Q69L75_ORYSA	Putative AAK1 protein.	308	32	53	1.4e-36
UNIPROT:09BUD9_HUMAN	AAK1 protein.	474	32	58	1.8e-36
UNIPROT:Q6P1I8_MOUSE	Cyclin G associated kinase.	1305	34	57	2.4e-36
UNIPROT:GAK_RAT	Cyclin G-associated kinase (EC 2.7.1.-).	1305	34	57	3.2e-36
UNIPROT:Q6P9S5_MOUSE	Cyclin G associated kinase.	1305	34	57	6.7e-36
UNIPROT:BMP2K_HUMAN	BMP-2 inducible protein kinase (EC 2.7.1.37) (BIKE) (HRIHFB2017).	1161	33	55	9.4e-36
UNIPROT:Q69L76_ORYSA	Cyclin G-associated kinase-like protein.	667	32	53	1.2e-35
UNIPROT:Q9NSY1-2	Splice isoform 2 of Q9NSY1	662	33	55	2.3e-35

DB:ID	Source	Length	%ID	%Pos	E ()
UNIPROT:Q9NSY1-3	Splice isoform 3 of Q9NSY1	677	33	55	2.7e-35
UNIPROT:BMP2K_MOUSE	BMP-2 inducible protein kinase (EC 2.7.1.37) (BIKE).	1138	33	55	6.5e-35
UNIPROT:AKL1_YEAST	Serine/threonine-protein kinase AKL1 (EC 2.7.1.37).	1108	27	49	7.0e-35
UNIPROT:Q91Z96-2	Splice isoform 2 of Q91Z96	649	33	55	9.8e-35
UNIPROT:Q5PRC8_BRARE	Zgc:101641.	646	33	58	1.6e-34
UNIPROT:Q6BMC6_DEBHA	Debaryomyces hansenii chromosome F of strain CBS767 of Debaryomyces hansenii.	1003	24	44	4.8e-34
UNIPROT:Q9UPV4_HUMAN	KIAA1048 protein (Fragment).	897	32	58	6.1e-34
UNIPROT:Q8MSQ7_DROME	SD07306p.	698	34	56	3.3e-33
UNIPROT:Q7KT38_DROME	CG10637-PC, isoform C.	698	34	56	3.3e-33
UNIPROT:Q6FPJ1_CANGA	Candida glabrata strain CBS138 chromosome J complete sequence.	622	24	44	4.9e-33
UNIPROT:Q6FRA3_CANGA	Candida glabrata strain CBS138 chromosome H complete sequence.	1027	31	53	1.3e-32

The term “% identity” refers to the number of amino acids that are identical in the alignment and the term “% positive” refers to the number of amino acid residues that belong to the same class. Abbreviations used in Table 3.6: ANOGA = *Anopheles gambiae*, ARATH = *Arabidopsis thaliana*, BRARE = *Danio rerio*, CANGA = *Candida glabrata*, DICDI = *Dictyostelium discoideum*, DROME = *Drosophila melanogaster*, DEBHA = *Debaryomyces hansenii*, MOUSE = *Mus musculus*, ORYSA = *Oryza sativa*, PONPY = *Pongo pygmaeus*, RAT = *Rattus norvegicus*, YEAST = *Saccharomyces cerevisiae*

3.7.2 Comparative analysis of the catalytic domain of PfcGAK with its homologues

Multiple sequence alignments were carried out using the DBClustal program at EBI (www.ebi.ac.uk). Several alignments were carried out to determine the relationship of the amino acid sequence of PfcGAK with its closest homologues. Figure 3.21 shows the alignment and analysis of the catalytic domain of the predicted PfcGAK amino acid sequence with cGAK sequences of other organisms. As expected, significant similarity was observed for the kinase catalytic domain (Figure 3.21). In this analysis the kinase consensus sequence is shown above the alignment and cGAK conserved sequences are shown below the alignment. The complete alignment is shown in Appendix C.1 and was also used for the sequence analysis of the catalytic domain shown in Figure 3.21. This analysis included six characterized and putative cGAKs that are most closely related to PfcGAK.

The kinase catalytic domain was found to start close to the N-terminus with the boundary for subdomain I starting three residues from the N-terminus at amino acid Phe4. The catalytic domain ends at the C-terminal amino acid of subdomain XI, Phe303. Analysis indicated a few conserved amino acid residues in the central region of PfcGAK (between amino acid 308 and 756) but there was poor conservation at the C-terminal region (Appendix C1). This is in contrast to cGAKs of other proteins where there are considerable conservation in the C-terminal region (Appendix C5). PfcGAK is shorter than its closest homologues and gaps were introduced in the alignment of the C-terminal domain. The kinase catalytic domain of PfcGAK is well conserved. All the invariant residues required for kinase catalytic activity are present in the catalytic domain of PfcGAK.

Most protein kinases have a glycine (G) triad (gxGxxg) in subdomain I. The capital G indicates the glycine residue that is thought to be essential for kinase activity. Subdomain I start at a hydrophobic residue about seven residues upstream from the glycine triad. Although all the glycine triad residues are not required for catalytic activity it is a commonly conserved motif in serine/threonine kinases (Hanks and Hunter, 1995). The absence of the glycine triad is not unusual in *Plasmodium* kinases (Ward *et al.*, 2004). In subdomain I the typical glycine triad (gxGxxg) was conserved in the cGAK sequences but not in the PfcGAK and the sequence of *Drosophila melanogaster* (Q9VMY8 DROME). The glycine triad forms a hairpin structure enclosing part of the ATP molecule and acts as a flexible clamp that anchors the non-transferred phosphates of the ATP (Hanks and Hunter, 1995). This hairpin loop is found between β -sheet **1** and **2** that comprises subdomain I in the small lobe of the catalytic domain (Figure 3.22).

In PfcGAK four glycine residues were found in subdomain I with one of them located in the glycine triad. It is likely the invariant glycine residue that forms part of the kinase consensus sequence. However an alternative alignment places this residue outside the typical region of the glycine triad. All the cGAK sequences lacked the highly conserved valine (V) residue that is found with the glycine triad. With the exception of PfcGAK all the cGAK sequences had a conserved valine residue 10 amino acids downstream of the third glycine. A conserved motif, including a third glycine residue, (ooxEGxuxFVYxAxDo (see legend to Figure 3.21)) was found at the C-terminus of this subdomain.

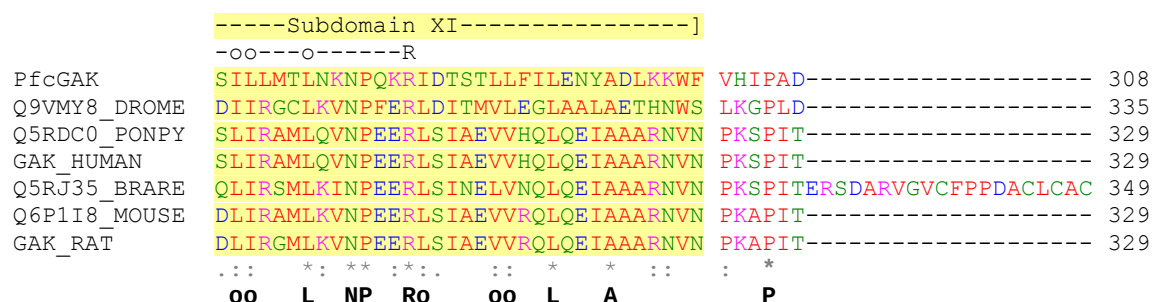


Figure 3.21 Analysis of amino acid sequence alignment of the catalytic domain of PfcGAK. Alignments were obtained from the comparison of the amino acid sequences of cGAK from various organisms using the ClustalDB programme at www.ebi.ac.uk. The eleven subdomains of the kinase catalytic site are shaded. The code for the consensus line above the alignment is as follows: uppercase letters = invariant residues; lowercase letters = nearly invariant residues; o = conserved non-polar residues; * = conserved polar residues; + conserved residues with neutral polarity. The consensus sequences of cGAK are indicated underneath the alignment. The code for conserved sequences is as follows: o = non-polar amino acid, b = basic residues, z = acidic residues, u = aromatic residues x = any amino acid. The alignment is shown up to amino acid 308 of PfcGAK. Colour code for amino acids: **Red** = small hydrophobic amino acid residues (including aromatic residues); **Blue** = acidic amino acid residues; **Magenta** = basic amino acid residues; **Green** = hydroxyl, amine and glutamine residues, [*] = identical amino acids; [:] = conserved substitutions; [.] = semi-conserved substitutions. The aligned sequences are PfcGAK = putative PfcGAK sequence, Drome = *Drosophila melanogaster*, PONPY = *Pongo pygmaeus*, BRARE = *Danio rerio*, MOUSE = *Mus muscules*, RAT = *Rattus norvegicus*.

Subdomain II, is found in the small lobe of the catalytic domain and is formed by β -strand **3** and α -helix **B** (Figure 3.22) In subdomain II the consensus lysine (K) residue was conserved in the PfcGAK sequence as well as the other cGAK sequences. This lysine is essential for maximal enzyme activity and makes contact with the α - and β -phosphate of ATP to anchor and orientate the molecule. This subdomain starts of with a conserved serine or threonine residue at the amino terminus followed by a conserved tyrosine three residues N-terminal of the consensus lysine residue. The nearly invariant alanine (A) residue on the oaoK motif found in the other cGAK sequences of this alignment was replaced by a threonine in PfcGAK. Threonine is more reactive than alanine and may undergo phosphorylation. An acidic residue (glutamic acid (E) or aspartic acid (D)) is conserved close to the C-terminus of this subdomain.

Subdomain III forms the large α -helix **C** in the small lobe of the kinase domain (Figure 3.22). The consensus glutamic acid residue as well as two non-polar residues close to the C-

terminal, common to serine-threonine kinases, was conserved in Subdomain III of PfcGAK. The glutamic acid residue of subdomain III forms a salt-bridge with the conserved lysine residue of subdomain II. In this alignment the consensus glutamic acid residue was followed by a conserved non-polar residue. At the C-terminus a lysine and a leucine residue were conserved.

Subdomain IV is formed by the β strand **4** in the small lobe of the catalytic domain and is characterised by a nearly invariant histidine (H) residue close to the N-terminus and four conserved non-polar residues. This subdomain forms a hydrophobic β -strand in the small lobe of the kinase domain. In this alignment the histidine residue was conserved as well as an asparagine residue found before the first conserved non-polar residue clusters. Only two of the four non-polar residues were conserved. The first non-polar residue was either valine or isoleucine (I), and the second non-polar residue was valine for all the cGAKs. The third non-polar residue was replaced by an aromatic residue, either phenylalanine (F) or tyrosine (Y).

Subdomain V connects the small lobe to the large lobe of the catalytic domain and consists of the hydrophobic β -strand **5** in the small lobe and the small helix **D** in the large lobe as well as an extended connecting chain (Figure 3.22). In subdomain V of the serine-threonine kinases a cluster of seven non-polar residues interrupted by a polar residue is found near the N-terminus. In this alignment the first five residues of PfcGAK were non-polar and likely to be involved in the formation of the hydrophobic β -strand. The acidic glutamic acid residue was conserved for all the sequences in this alignment and served as the interrupting polar residue. Of interest was a conserved cysteine (C) residue that may be involved in the formation of a disulphide bridge. In the central region of subdomain V the conserved motif, Loz, is found in the aligned cGAKs. The conserved leucine (L) is followed by a nonpolar valine or leucine residue and an acidic residue.

According to Hanks and Hunter (1995) subdomain VIA forms the large hydrophobic α -helix **E** that extends through the large lobe of the catalytic domain (Figure 3.22). It does not interact directly with the ATP molecule and this helix is generally thought to be a support structure (Hanks and Hunter, 1995; Schenk & Snaar-Jagalska, 1999). Conserved amino acids of this domain include a number of non-polar residues as well as a polar residue and a small uncharged residue. In the alignment of the cGAKs protein this domain starts with a conserved non-polar residue. Six residues downstream of the N-terminus of this domain a

cluster of non-polar residues are found that includes a fully conserved isoleucine. This is different from the consensus sequence indicated for kinases. The polar (*) and the small uncharged residues (+) indicated in the consensus line were conserved. Close to the C-terminus a nearly invariant histidine residue is common. This residue as well as a nearby glutamine (Q) residue is fully conserved in this cGAK alignment. The two C-terminal residues of this domain consist of a double or single proline (P) residue, which is likely to function as a helix breaker.

Subdomain VIB is essential for kinase activity and is formed by the two small hydrophobic β -strands **6** and **7** which are linked by the intervening catalytic loop (shown in red in Figure 3.22). The sequence oohrDoKo+Nooo was fully conserved in the cGAK alignment with the consensus motif oohrDoKoENoL present. Aspartic acid, lysine and asparagine residues are critical for the phosphotransfer reaction (Hanks and Hunter 1995).

Subdomain VII is formed by β - strands **8** and **9** with the aspartic acid-phenylalanine-glycine (DFG) triplet in the connecting loop (Figure 3.22). The DFG triplet is a common feature of Subdomain VII. Only the aspartic acid residue of this motif is essential for kinase activity. It chelates the primary activating Mg^{2+} -ions that bridge the β and γ phosphates of the ATP, thereby contributing to the correct orientation of the γ -phosphate during transfer (Hanks and Hunter 1995). The DFG motif that was conserved in the other cGAK homologues was replaced by a DFC motif in PfcGAK. The other cGAK proteins in this alignment had the DFG triplet present in the conserved region IKLCDFGS, whereas the sequence YKICDFCS was present in PfcGAK. In summary, the sequence KoCDFxS was conserved in this subdomain of all the cGAK sequences. Closer to the C-terminus of this domain another serine residue was conserved in all the cGAK sequences.

Subdomain VIII consists of a loop that faces the catalytic cleft. As described in Hanks and Hunter (1995), subdomain VIII of typical serine-threonine kinases is characterised by a highly conserved APE motif of which the glutamic acid residue is invariant and preceded by a nearly invariant proline residue. The invariant glutamic acid residue is followed by two non-polar residues. The cGAK kinases had the conserved motif TPE, with the exception of the *Drosophila* cGAK which had the sequence SPE. In PfcGAK the sequence PPE was found. This may have a major impact on the orientation of the loop formed by this subdomain. The eleven amino acids upstream from the TPE motif in cGAK were well conserved and the motif TtxooYRxPEooD was identified in this alignment. Close to the C-

terminus the motif uSNxxI was conserved. Subdomain VIII plays a major role in substrate recognition and binding and is also a target for phosphorylation and subsequent change in activation status (Hanks and Hunter 1995). A potential phosphorylation site, namely SNFE was found in this subdomain and a second phosphorylation site SWK spans the border between subdomain VIII and IX in PfcGAK.

Subdomain IX forms a large α -helix **F** in the large lobe of the protein kinase domain. The invariant aspartic acid close to the N-terminus stabilizes the catalytic loop and is followed by a cluster of non-polar residues with a nearly invariant glycine residue in the middle and a Pro residue close to the C-terminus (Hanks and Hunter 1995). In the alignment all these residues were conserved in PfcGAK. In eukaryotic serine-threonine kinases the sequence DxWxxG is frequently conserved (Schenk and Snaar-Jagalska, 1999). In this alignment the amino acid sequence KxDIWooGCILYoL was conserved for all the cGAKs. Closer to the C-terminus a HPu sequence was conserved.

Subdomain X is generally a poorly conserved domain and is formed by the small α -helix **G** at the base of the catalytic loop with no defined function (Hanks and Hunter 1995). The cGAK proteins showed some identical residues leucine, isoleucine, asparagine (N) forming LxIoN close to the centre of the domain. A conserved proline residue is found closer to the C terminus of this domain.

Subdomain XI forms α -helix **H** that extends to the COOH terminus of the catalytic loop. The kinase consensus sequence of Subdomain XI has four conserved non-polar sequences and a conserved arginine (R) residue (Hanks and Hunter 1995). The four non-polar consensus sequences as well as the invariant arginine residue were conserved in this subdomain. The cGAKs showed conservation of additional amino acids and polar amino acids between the invariant arginine and the C-terminus.

The ribbon diagram in Figure 3.22 shows the typical features of the kinase catalytic domain. Subdomain I is found at the NH₂ terminus and folds into a β -strand - turn- β -strand structure consisting of β - strands 1 and 2. Subdomain II, containing the invariant Lys forms β - strand 3 and α - helix B in the small lobe of the catalytic domain. Subdomain III forms the large α -helix C in the small lobe. Subdomain IV forms β -strand 4 in the small lobe. Subdomain V connects the small lobe to the large lobe and forms the hydrophobic β -strand

5 in the small lobe and the small helix D in the large lobe as well as an extended connecting chain. (Hanks and Hunter 1995).

Subdomain VIA folds into the large hydrophobic α -helix E that extends through the large lobe and mainly acts as a support structure. Subdomain VIB forms the two β - strands 6 and 7 with an intervening catalytic loop. Subdomain VII forms β - strands 8 and 9 with the DFG triplet in the connecting loop. Subdomain VIII forms a chain that faces the catalytic cleft.

Subdomain IX forms a the large α -helix F of the large lobe and helps to stabilise the catalytic loop. Subdomain X is poorly conserved in kinases and forms the small α -helix G at the base of the catalytic loop. Subdomain XI forms α -helix H that extends to the COOH terminus of the catalytic loop. Many protein kinases have an additional α -helix (I) that follows helix H that lies outside the conserved catalytic core at the NH₂ – terminus. (Hanks and Hunter, 1995).

Figure 3.23 is a graphical representation and comparison of the kinase catalytic region of PfcGAK with all the major consensus sequences indicated. It shows that all the consensus sequences of the catalytic domain were conserved in PfcGAK with the exception of the glycine triad in subdomain I. This motif was conserved in the other cGAK homologues. Although the DFG motif in subdomain VII is common to many kinases only the Asp residue is required for activity and considered to be an invariant residue and a DFC motif is found in PfcGAK.

A number of alternative motifs were identified for the cGAK homologues as shown in Figure 3.23. In subdomain IV the HxNoV motif is found, the KoCDFxS motif in subdomain VII, the TTxoYRxPEooD motif in subdomain VIII and the KxDIWooGCILYoL motif in subdomain IX. Current information does not make it possible to predict the function of these motifs.

An alignment of the complete PfcGAK protein with cGAKs and BIKE kinases was performed and is shown in Appendix C2. As expected there was good conservation in the catalytic region with all kinase consensus residues conserved. However, the conservation in the carboxyl region was poor. A separate alignment of the catalytic domain of PfcGAK with that of characterised human, mouse and rat cGAK showed 91 identical amino acids

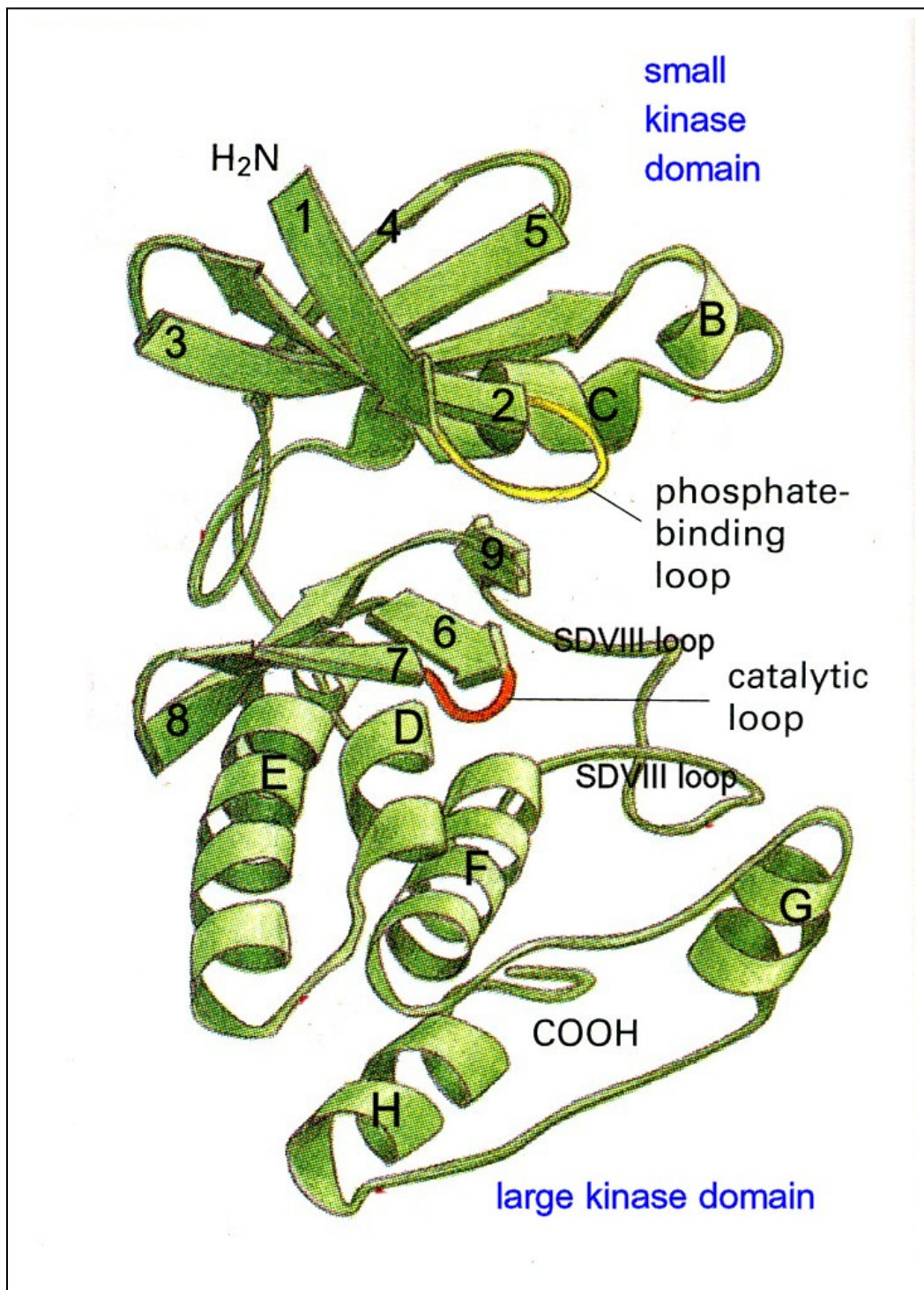


Figure 3.22 Ribbon diagram of a typical serine threonine kinase catalytic region.

β -strands are indicated by numbers and α -helices are indicated by alphabetical letters and SD indicates subdomain. SD I = 1 & 2, SD II = 3 & B, SD III = C, SD IV = 4, SD V = 5 & D, SD VIA = E, SD VIB = 6 & 7, SD VII = 8 & 9, SD VIII = chain, SD IX = F, SD X = G, SD XI = H (Alberts *et al.*, 2002; Hanks and Hunter, 1995)

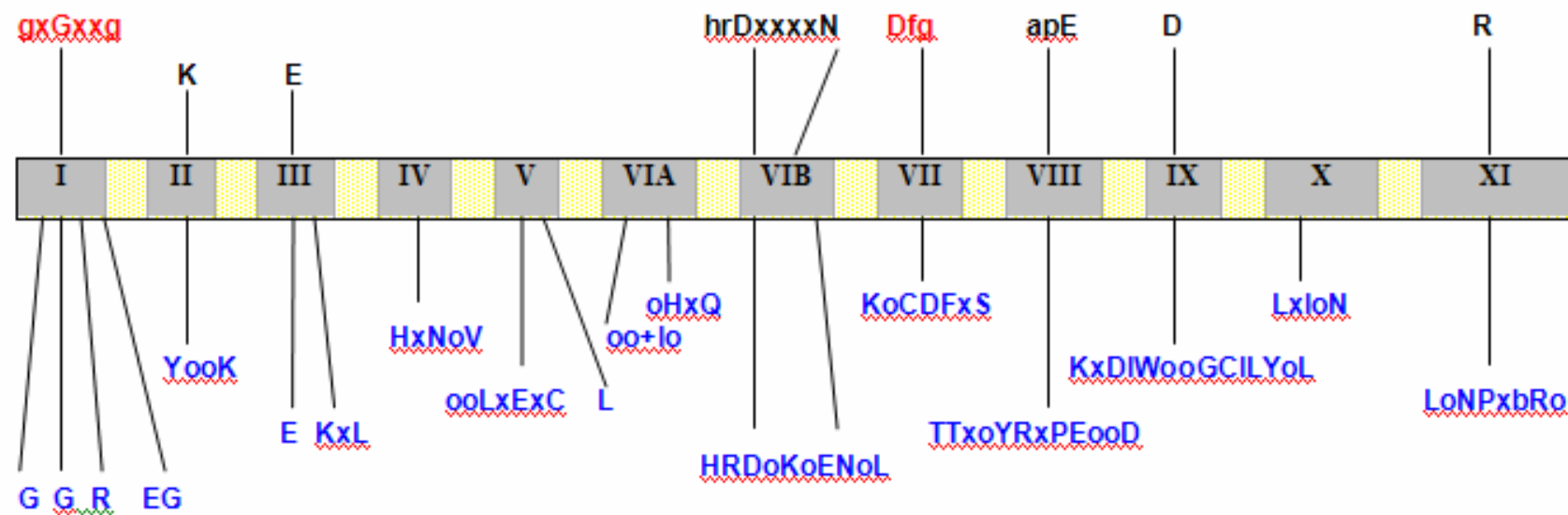


Figure 3.23 Comparison of the major consensus amino acid motifs in the catalytic domain of Ser / Thr protein kinases with PfcGAK and cGAK homologous proteins. The twelve subdomains of the protein kinase catalytic domain are indicated by the central bar and consensus residues for protein kinases are indicated above the central bar. The consensus motifs in red are conserved in the cGAK homologues but not in PfcGAK. The corresponding conserved residues and motifs for PfcGAK and other cGAK residues are indicated beneath the central bar. Conserved amino acids are shown in capital letters, o refers to semi-conserved non-polar residues, b refers to semi-conserved basic residues and z to semi-conserved acidic residues, u to aromatic amino acid residues, x indicates any amino acid.

and a high percentage of conservation of 59%, 60% and 65% in subdomains VIB, VIII and IX respectively (Appendix C3). This was compared to the catalytic region alignment of PfcGAK with the BIKE kinases (shown in Figure 3.25), which also showed 91 identical amino acids in the catalytic domain. However, compared to the cGAK alignment the extensive conservation in subdomain VIII and IX was not observed.

3.7.3 Comparative analysis of the carboxyl terminal region of PfcGAK with mouse, human and rat cGAKs

The carboxyl region of cGAK proteins (excluding PfcGAK) shows considerable homology. At the C terminal domain human cGAK also shares homology to auxilin I, a protein found exclusively in brain tissue. Auxilin I lacks the N-terminal kinase domain found in cGAK proteins. It contains a J-domain, responsible for interaction with heat shock cognate 70 protein, clathrin binding motifs and a region of homology to the phosphatidylinositol 3-phosphatase PTEN. (Korolchuck and Banting, 2002)

Figure 3.24 below is an alignment, excluding the catalytic domain, of PfcGAK with the characterised human, rat and mouse cGAKs to identify conserved regions or motifs falling outside the catalytic domain. This alignment of PfcGAK with the cGAK members of higher organisms (human, mouse and rat) showed good conservation in the catalytic site (shown in Appendix C3) and some conservation in the carboxyl terminal domain (Figure 3.24). Due to the insertion of gaps in the alignment a number of amino acids were conserved but the overall identity was only 12%. No distinct J domain or clathrin binding site was identified, and their absence in PfcGAK was confirmed by a Prosite motif search.

PfcGAK	-----VHIPADIKKRVNQIFKMNELNMNNKHTDLIEVSHE	338
GAK_HUMAN	-----PKSPITELLEQNGGYGSATLSRGPPFPVGPAGSGYS	359
Q6P1I8_MOUSE	-----PKAPITELLEQNGGYGNSGSPRAQPPCGGTVNS--S	357
GAK_RAT	-----PKAPITELLEQNGGYGNSGSPRAQPPSGGPVNS--S	357
	: * . * :	
	P N	
PfcGAK	KLIDVKISSQKYETNLPKSNNNFLKFFT-TSNLSGDIFFKKKIEKEETKKEIKQVGN----	393
GAK_HUMAN	GGLALAEYDQPYGGFLDILRGGTERLFTNLKDTSSKVIQSVANYAKGDLDISYITSRIAV	419
Q6P1I8_MOUSE	GVLALAEYDQPYGGFLDILRGGTERLFTNLKDTSSKVIQSVANYAKGDLDISYITSRIAV	417
GAK_RAT	GVLALAEYDQPYGGFLDILRGGTERLFTNLKDTSSKVIQSVANYAKGDLDISYITSRIAV	417
	: : . * * * . . . : : * : : . : * . : .	
	Q Y L oFT S oo zI o	
PfcGAK	-----KQVSETEKEPKDKLENNPDGNKLQNDKDSGKLQNDKDSGKLQNE----KDS	440
GAK_HUMAN	MSFPAEGVESALKNNIEDVRLFLDSKHPGHYAVYNLSPRTYRPSRFHNRVSECGWAARRA	479
Q6P1I8_MOUSE	MSFPAEGVESAIKNNIEDVRLFLDAKHPGHYAVYNLSPRIYRASKFHNRVTECGWAVVRA	477
GAK_RAT	MSFPAEGVESAIKNNIEDVRLFLDAKHPGHYAVYNLSPRIYRASKFHNRVTECGWAVVRA	477
	 : : : * : : * : : . . : . : . : : : : : :	
	L G o	

PfcGAK GKLQNDKDSGK-----LQNDKDSGKQLQNDKDSGKQLQNDKVSGLQNDKDSGKQLQNDKVSGL 495
 GAK_HUMAN PHLHLYNICRNMHAWLRQDHKNVCVHMDGRAASAVAVCSFLCFCRLFFSTAEAAVYMF 539
 Q6P1I8_MOUSE PHLHSLYTLCRSMHAWLRQDHKNVCVHMDGRAASAVAVCAFLCFCRLFFSTAEAAVYMF 537
 GAK_RAT PHLHSLYTLCRSMHAWLRQDHKNVCVHMDGRAASAVAVCAFLCFCRLFFSTAEAAVYMF 537
 *:.. : *::*: . : : * . *.. * : .. :
 L L D o D V

PfcGAK KLQNDRDGNKLQNNHIIIDYTDNKMNDNKKNKSD----- 529
 GAK_HUMAN SMKRCPPGIWPSHKRYIEYVCDMVAEEPITPHSKPILVRAVMTVPVPLFSKQSGCRPFC 599
 Q6P1I8_MOUSE SMKRCPPGIWPSHKRYIEYVCDMVAEEPITPHSKPMLVKSVMTPVPVPLFSKQNGCRPFC 597
 GAK_RAT SMKRCPPGIWPSHKRYIEYVCDMVAEEPITPHSKPMLVKSVMTPVPVPLFSKQNGCRPFC 597
 * * * * * * * * * *
 G I Y

PfcGAK ----ISENITSSGDKQNSVPDQSENQSKSHSYNLIKIQSKDTLIIDHVDKDFPSDKTNTD 585
 GAK_HUMAN EVYVGDERVASTSQEYDKMRDFKIEDGKAVIPLGVTVQGDVLIVIIYHARSTLGGRLQAKM 659
 Q6P1I8_MOUSE EVYVGEEVTTTSQEYDMKEFKIEDGKAVIPLGVTVQGDVLIIIIYHARATLGGRLQAKM 657
 GAK_RAT EVYVGEEVTTTSQEYDMKEFKIEDGKAVIPLGITVQGDVLIIIIYHARSTLGGRLQAKM 657
 * * * * * * * * * *
 R K Q I H

PfcGAK FSFQDTLCEFEKKCMNN---SSNYSNDKKVAGDIKKD-----DCSVNIQVNNNNNDVCHN 637
 GAK_HUMAN ASMKMFQIQFHTGFVPRNATTVKFAKYDLDACDIQEKYPDLFQVNLEVEVEPRDRPSREA 719
 Q6P1I8_MOUSE ASMKMFQIQFHTGFVPRNATTVKFAKYDLDACDIQEKYPDLFQVNLEVEVEPRDRPSREA 717
 GAK_RAT ASMKMFQIQFHTGFVPRNATTVKFAKYDLDACDIQEKYPDLFQVNLEVEVEPRDRPSRDV 717
 * * * * * * * * * *
 S F A DI V

PfcGAK DMKDVNECTNLLNLDVFYEYENETNDSIKKINDGDFIKQSG--CDFIKQNGGDFIKPN--- 692
 GAK_HUMAN PPWENSSMRGLNPKILFSSREEQQDILSKFGKPELPRQPGSTAQYDAGAGSPEAEPTDSD 779
 Q6P1I8_MOUSE PPWENTSLRGLNPKILFSNREEQQDILSKFGKPELPRQPGSTAQYDAEAGSPEAEITESD 777
 GAK_RAT PPWENTSLRGLNPKILFSNREEQQDILSKFGKPELPRQPGSTAQYDAEAGSPEAEITESD 777
 : .. * * * * * * * * * *
 L F E D K Q G G

PfcGAK -----DSNCIFADINFNNNDINLNNNDIN----- 714
 GAK_HUMAN SPPSSSADASRFLHTLDWQEEKEAETGAENASSKESESALMEDRDESEVSDEGGSPISSE 839
 Q6P1I8_MOUSE SPQSSSTDTNHFLHTLDWQEEKEPETGLDNTSPKESQSVLIADGDGSEVSDEEEASFPSE 837
 GAK_RAT SPQSSSTDTNHFLHTLDWQEEKDPETGVDNTSPKESQSNLIADGDGSEVSDEEEASCPSE 837
 * * * * * * * * * *
 D

PfcGAK -----LNNNDINFNN-----DINLNNNDIDFFDVIQNHDLK----- 743
 GAK_HUMAN GQEPRADEPPGLAAGLVQQLDLFVEVETPAVLPEPVPQEDGVDLLGLHSEVGAGPAVPPQ 899
 Q6P1I8_MOUSE ERKPGAGEDTPRLAAGTKQQLDIFDVGMLAAPQEPVQPEEGVDLLGLHSEGLDLRPAAPLQ 897
 GAK_RAT ERKPGAGEDTPRLAAGTKQQLDIFDVGMLAAPQEPVQPEEGVDLLGLHSEGLDLRPAAPLQ 897
 ::: * * * * * * * * * *
 Do F Doo o

PfcGAK -----ACKAPSSNTDLLSCLLPPEAASQGGPEDLLS-EDPLLLASAPPLSVQSTPFGGPPAAA 958
 GAK_HUMAN ACGVSSNTDLLSCLLEPSDAAQVGPPGDLGGEAPLLLASPVSPGLQNNLQGKVPDVT 957
 Q6P1I8_MOUSE ACGVSSNTDLLSCLLEPSDAAQVGPPGDLGGEAPLLLASPVSPGLQNNLQGKVPDVT 957
 GAK_RAT ASGVQSSNTDLLSCLLEPSDAAQVGPPGDLGGETPLLLASPVSLLGVSQNLQGKVPDVT 957

PfcGAK -----KKEDVLNINRNNE-----HDKNAITCSNEYNNLSLF- 774
 GAK_HUMAN DPFGPLLPSSGNSQPCSNPDFGFEFLNSDSVTVPSPFSAHSAPPPSCSADFLHLGLDLP 1018
 Q6P1I8_MOUSE DPFDQFLLSSNSDTQPCSKPDFGFEFLNSDSVASSTAFPSTHSAPPPSCSTAFHLHLGLDLP 1017
 GAK_RAT DPFDQFLLPSSSDTQPCSKPDFGFEFLNSDSVASSTAFPSTHSAPPPSCSTAFHLHLGLDLP 1017
 : * * * * * * * * * *
 D N H CS L

PfcGAK -----YDPWNN-----NTK 783
 GAK_HUMAN GEPKMTASSNPDLGGAAWTETAASAVAPTATEGLFSPGGQPAPCGSQASWTKSQ 1078
 Q6P1I8_MOUSE AEPKVIASSHPDLLGWDTWADTATPGPASIPVPEGLTFSSAGHPAPPGFNPSTKSQ 1077
 GAK_RAT AEPKVIASSHPDLLGWDTWADTALPGPASMPVPEGLTFSSAGHPAPPGFNPSTKSQ 1077
 : . * * * * * * * * * *
 W

```

PfcGAK      DNHFFGD----- 790
GAK_HUMAN   NPDPPADLDLSSGLQGSPAGFPFGGFIPKTATTPKGSSSWQTSRPPAQGASWPPQAKPP 1138
Q6P1I8_MOUSE NLDPPADLDLSSSLQGLPAGLPAGGFVGAAPTQKSNSPWQANRPTAPGTSWTPQAKPA 1137
GAK_RAT      NPDPPADLDLSSSLQGLPAGLPAGSFVGTSAATTHKSNSSWQTTTRPTAPGTSWPPQAKPA 1137
          : . * . *
          F D

PfcGAK      -----DDYNVNKSSFIFN-----NDLFDN--YELCTNNNTE 819
GAK_HUMAN   PKACTQPRPNYASNFVIGAREERGVRAPSFQKPKVSENFEDLLSNQGFSSRSDDKKGP 1198
Q6P1I8_MOUSE PRASEQLR----SHFSVIGAREERGVRVPSFAQKPKVSENFEDLLPNQGFSS-KSDKKGP 1192
GAK_RAT      PRASEQLR----SHFSVIGAREERGVRAPSFQKPKVSENFEDLLPNQGFSS-KSDKKGP 1192
                      :: . * . * * : : * : * : : :
                      V SF DL N

PfcGAK      KSTHHVKNNCMDINNNIFSNKNYDEHVDKKEKNISYDYNVKSHPIYSSYKNEKRCNLNDI 879
GAK_HUMAN   KTIAEMRKQDLAKDTPPLKLK-LLDWIEGKERNIRALLSTLHTVLWDGESRWTPVGMADL 1257
Q6P1I8_MOUSE KTMAEMRKQELARDTDPPLKLK-LLDWIEGKERNIRALLSTLHTVLWDGESRWTPVSMADL 1251
GAK_RAT      KTMAEMRKQELARDTDPFCLK-LLDWIEGKERNIRALLSTLHTVLWDGESRWTPVSMADL 1251
          * : : : : : : : : : : * : : : : : : : : : : * :
          K K KE NI D

PfcGAK      KQ-----NNSSKYIMSDKNDKFSELLDDFQKCTIK 909
GAK_HUMAN   VAPEQVKKHYRRAVLAVHPDKAAGQPYEQHAKMIFMELNDAWSEFENQGSRLPLF- 1311
Q6P1I8_MOUSE VTPEQVKKQYRRAVLVHHPDKATGQPYEQYAKMIFMELNDAWSEFENQGSRLPLF- 1305
GAK_RAT      VTPEQVKKQYRRAVLVHHPDKATGQPYEQSAKMIFMELNDAWSEFENQGSRLPLF- 1305
                      : : * * : : * * : * : : : : :
                      K K ND SEo

```

Figure 3.24 An alignment of the carboxyl region of PfcGAK with human mouse and rat cGAK proteins. Conserved sequences are indicated underneath the alignment. Colour code for amino acids: **Red** = small hydrophobic amino acid residues (including aromatic residues); **Blue** = acidic amino acid residues; **Magenta** = basic amino acid residues; **Green** = hydroxyl, amine and glutamine residues, [*****] = identical amino acids; [**:**] = conserved substitutions; [**.**] = semi-conserved substitutions.

3.7.4 Comparative analysis of PfcGAK with BIKE kinases

Blastp analysis of PfcGAK also showed homology to BIKE kinases (Table 3.4). The analysis in Figure 3.25 shows the alignment of PfcGAK with two BIKE kinases. Despite the fact that the BIKE kinases were glycine rich in subdomain I they shared with PfcGAK the absence of the common glycine triad motif (GxGxxG). Subdomains VIII and IX showed a lower percentage of conserved residues (50% and 48% respectively) when compared to the cGAK alignment. All the amino acid residues required for catalytic activity were conserved in the catalytic domain. In the carboxyl terminal region 81 residues (13%) were conserved in comparison to the 75 residues (12%) of the PfcGAK alignment with cGAK as shown in Figure 3.24. Although more identical residues were identified in this alignment no clearly conserved motifs were identified. When a Clustal alignment

combining the cGAK and BIKE kinases were performed, minimal conservation of the C terminus (only 7 residues) was observed. However the catalytic domain had good conservation with 63 residues conserved. This alignment is shown in appendix C.2.

		Subdomain I											
PfcGAK	MLK	FRSICTT-----LIGGKVYNINGKTIREEKLI	SEG	33									
BMP2K_HUMAN	MKK	FSRMPKSEGGSGGGAAGGGAGGAGAGCGSGGSSVGVRVFAVGRHQVTLEESLA	EG	60									
BMP2K_MOUSE	MKK	FSRMPKSEG-SGGGAAAGGAAGGGLGGGFAS--SSMGVRVFAVGRYQVTLEESLA	EG	57									
	* * *	: .:						: *	: *:	: .:	:	*: .: *	
	M K F							G	V		E	EG	
		Subdomain II					Subdomain III					Subdomain IV	
PfcGAK	AYSFVYMAKDLN	TNRAYTIKKTICQDKE	KIDMANKEINILKSL	PPHKNIVQYYGSTI	ISE	93							
BMP2K_HUMAN	GFSTVFLVRTHG	G-IRCALKRMVYNNMP	DLNVCKREITIMKEL	SGHKNIVGYLDCAVNSI		119							
BMP2K_MOUSE	GFSTVFLVRTHS	G-IRCALKRMVYNNTP	DLNICKREITIMKEL	SGHKNIVGYLDCAVNSI		116							
	.: * * : .:	: *:	: :	.: .: .: .: *				*: *: *: *		*: .: .: *			
	S V		K					EI I K L		HKNIV Y		o S	
		Subdomain V					Subdomain VIA						
PfcGAK	-NNYKI	VIMLMEYCE	RGNLLNIFQKNKD	-	IKEIHIVNILKDIITGLCFLHNQEIP	IIHR	151						
BMP2K_HUMAN	SDNVWE	VLILMEYCRAGQVVNQMNKKLQTG	FTEPEVLQIFCDTCEAVARLHQCKTP	IIHR	179								
BMP2K_MOUSE	SDNVWE	VLILMEYCRAGQVVNQMNKKLQTG	FTESEVLQIFCDTCEAVARLHQCKTP	IIHR	176								
	: *	: *: .: .: .: *	: .: .: *	: : *				: .: .: .: *		: .: .: *		: *	
	N	V LMEYC	G o o N	K				E	I D		LH	P IIHR	
		Subdomain VIB					Subdomain VII					Subdomain VIII	
PfcGAK	DIKLENILCDKN	GVYKICDFCSHSISNSFFPNDLSKSELSNL	KEEIERD	TTLIYRPEELI	211								
BMP2K_HUMAN	DLKVENILLNDG	GNYVLCDFG--SATNKFLN--PQKDG VNVV	EEEIKKYTTLSYRAPEMI		235								
BMP2K_MOUSE	DLKVENILLNDA	GNYVLCDFG--SATNKFLN--PQKDG VNVV	EEEIKKYTTLSYRAPEMI		232								
	: *: .: .: *	: *	: .: .: *	: *: .: *				: *	: .: .: *	: *: .: *	: *: .: *	: *	
	DoKoENIL	G G CDF	S N F	K				EEI	TTL YR	PE I			
		Subdomain IX					Subdomain X						
PfcGAK	DLYSNFEISW	KIDIWMVGCILYLLLFKIHPFQN	NNFLSIINASFTIPYCTKYSKR	IISIL	271								
BMP2K_HUMAN	NLYGGKPITT	KADIWALGCLLYKLCFFTLPFGE	S-QVAICDGNFTIPDNSRYSRN	IHCLI	294								
BMP2K_MOUSE	NLYGGKPITT	KADIWALGCLLYKLCFFTLPFGE	S-QVAICDGSFTIPDNSRYSRN	VHCLI	291								
	: *: .: .: *	: *	: .: .: *	: *: .: *				: *	: .: .: *	: *: .: *	: *: .: *	: *	
	LY	I	K DIW oGCoLY L F	PF				I	FYIP	YS			
		Subdomain XI											
PfcGAK	LMTLNKNPQKRIDTSTLLFILENYADLKWF	VHI-----PADIKKRVNQIFEKMNELNMN		326									
BMP2K_HUMAN	RFMLEPDPPEHRPDIFQVSYFAFKFAKKDCPV	SNINNSSIPALPEPMTASEAAARKSQIK		354									
BMP2K_MOUSE	RFMLEPDPPECRPDIFQVSYFAFKFAKKDCPV	SNINNSFLPSTLPEPMTATEAAARKSQMK		351									
	: *: .: .: *	: *	: .: .: *	: *: .: *				: *	: .: .: *	: *: .: *	: *: .: *	: *	
	L	P R D		A				I	P				
PfcGAK	NKHTDLIEVSHEKLIDVKISSQKYETNLP-----KSNNNFKLFFTTSNLSGDIFFKKIE			380									
BMP2K_HUMAN	ARITDTIGPTETSIAPRQRPKANSAATATPSVLTIQSSATPVKVLAPGEFGNHRPKGALR			414									
BMP2K_MOUSE	ARITDTIGPTETSIAPRQRPKAN-STAAATSSVLTIQSSATPVKVPAPGEFSNHKPKGALR			410									
	: *: .: .: *	: *	: .: .: *	: *: .: *				: *	: .: .: *	: *: .: *	: *: .: *	: *	
	TDxI		T	S				oK		K			
PfcGAK	KEETKKEIKQVGNKQVSETEKEPKDKLEN-----NPDGNKLQNDKDSGKLQN			427									
BMP2K_HUMAN	PGNGPEILLGQGPQQPPQQHRVLQQIQQGDWRLQQIHLQHHPHQQQQQQQQQQQQQQQ			474									
BMP2K_MOUSE	PGNGSEVLMVQGPQQPPQQHRVLQQIQQGDWRLQQIHL-HRHPHHHHQQQQQQQQQQQQQQQ			469									
	: *: .: .: *	: *	: .: .: *	: *: .: *				: *	: .: .: *	: *: .: *	: *: .: *	: *	
		G Q		L				P	Q		Q		

[illegible]

```

PfcGAK      -INNIFSNKNYDEHVDKK-EKNISYDYN-----VKSHPIYSSYKNEKRCNLNDIKQNN 883
BMP2K_HUMAN PFGAKPFFHSPDLSWHPPHQGLSDIRADHNTVLPGRPRQNSLHGSFHSADVLKMDDFGAVP 1127
BMP2K_MOUSE PFGAKPFFHPPDL-WHQPHQGLSDICVDHTTILPGRPRQNSVHGSFHSAETLRMDDFGAVP 1107
          :. : * : * :. : * * :. :. :. : * :. : . :. : * :
          F      H      I  D      S      D

PfcGAK      SSKYIMSDKND-----KFSELLDDFQKCTIK--- 909
BMP2K_HUMAN FTELVVQSITPHQSQQSQPVELDPFGAAPFPSKQ 1161
BMP2K_MOUSE FTELVVQSVTP---QQSQPVELDPFGAAPFPSKQ 1138
          :. :. :. : . . . * * * . :. :
          LDxF

```

Figure 3.25 An alignment of PfcGAK with human and mouse BIKE kinases.

Conserved residues are indicated underneath the alignment. Colour code for amino acids: **Red** = small hydrophobic amino acid residues (including aromatic residues); **Blue** = acidic amino acid residues; **Magenta** = basic amino acid residues; **Green** = hydroxyl, amine and glutamine residues, **[*]** = identical amino acids; **[:]** = conserved substitutions; **[.]** = semi-conserved substitutions.

3.7.5 Repetitive sequences

Amino acid analysis of PfcGAK revealed some unusual features such as the presence of two repetitive sequences, as shown in Figure 3.26. The smaller asparagine rich repetitive region was found between residues 700 and 730 and consisted of the residues DINF/LNN repeated five times with phenylalanine occurring twice and leucine three times. The hydrophobicity plot (Figure 3.28) also shows that this region is hydrophilic in nature. Blastp analysis showed similarity to numerous *Plasmodium* asparagine rich repetitive regions.

The larger nanomeric repeat region is found after the catalytic site, between amino acid residue 415 and 496 and consists of a lysine and aspartic acid rich nanomeric peptide, KLQND/EKD/VSG, that repeats itself nine times. The glutamic acid residue (E) occurs once in repeat 3 and the valine (V) residue occurs in the 7th and 9th repeat.

Nanomeric repeat region	Asparagine rich repeat region
KLQNDKDSG 424	
KLQNDKDSG 433	DINFNN 706
KLQNEKDSG 442	DINLNN 712
KLQNDKDSG 451	DINLNN 718
KLQNDKDSG 460	DINFNN 724
KLQNDKVSG 469	DINLNN 730
KLQNDKDSG 478	
KLQNDKVSG 487	

Figure 3.26 The two repetitive sequences of PfcGAK.

The sequence KLQN was also present in the following two nanomeric peptides although the last four amino acids were not conserved. The nanomeric repeat region is also shown in Figure 3.29 in italic print. The hydrophobicity plot of PfcGAK (Figure 3.28) shows that this region is hydrophilic. The nucleic acid sequence coding for this region and the amino-acid sequence were subjected to a Blastn and Blastp analysis respectively.

Blastn analysis of the nanomeric repeat region nucleic acid sequence showed similarity to a number of different *P. falciparum*, mouse and zebrafish sequences as indicated in appendix D1. The 100% identity hit in the first line refers to the identical sequence in the PfcGAK ORF on chromosome 12 of PlasmoDB. The aim of the search was to identify identical nucleic acid sequence in either *P. falciparum* or other organisms. The pairwise Blastn alignments of the first 15 hits is shown in appendix D1 and did not show any meaningful similarities.

Blastp analysis of the of the nanomeric repeat sequence, KLQND/EKD/VSG, showed significant similarity to a repeat region in the protease inhibitor enzyme, tissue factor pathway inhibitor (TFPI), found in *Canis familiaris*. This alignment is shown in Figure 3.27. TFPI has the nanomeric peptide sequence (GGLQHDSSES/GD) that repeats itself six times and its function is unknown (Girard *et al.*, 1994).

```

gi|50979034|ref|NP_001003243.1|  tissue factor pathway inhibitor [Canis familiaris]
gi|1083171|pir|S53325 tissue factor pathway inhibitor - rabbit
gi|833924|gb|AAB32443.1|  tissue factor pathway inhibitor; TFPI [Canis familiaris]
Length = 396

Score = 69.3 bits (168), Expect = 3e-10
Identities = 46/120 (38%), Positives = 65/120 (54%), Gaps = 6/120 (5%)

Query: 106 VGNKQVSETEKEPKDKLENNPD---GNKLQNDKDSGKLQNDK--DSGKLQNEKDSG-KLQ 159
      +GN  ET ++ K+  E + D      + N    G + N    +SG      DSG  +
Sbjct: 159 LGNLNNFETLEQCKNTCEGSIDLMDETVNNNTGSPGSMNNTSLFNSGDSLPLPADSGDSMP 218

Query: 160 NDKDSGKLQNDKDSGKLQNDKDSGKLQNDKVSGLQNDKDSGKLQNDKVSGLQNDKDRDGN 219
      D + G LQ+D +SG LQ+D +SG LQ+D  SG LQ+D +SG LQ+D  SG LQ+D  N
Sbjct: 219 PDSEIGGLQHDSSESGGLQHDSSESGGLQHDSSESGGLQHDSSESGGLQHDSSESGGLQHDSGDN 278

```

Figure 3.27 Blastp analysis of the nanomeric repeat region of PfcGAK

When the individual repeat KLQND/EKD/VSG was used in a Blastp search of the protein database it was found that several proteins contain part of the repeat sequence. The Blastp result of the hits with E values less than 1000 is shown in Appendix D.2. The sequence KLQNDKD occurs in a number of putative cell surface associated proteins of *Staphylococcus aureus* as well as *Trypanosoma cruzi*.

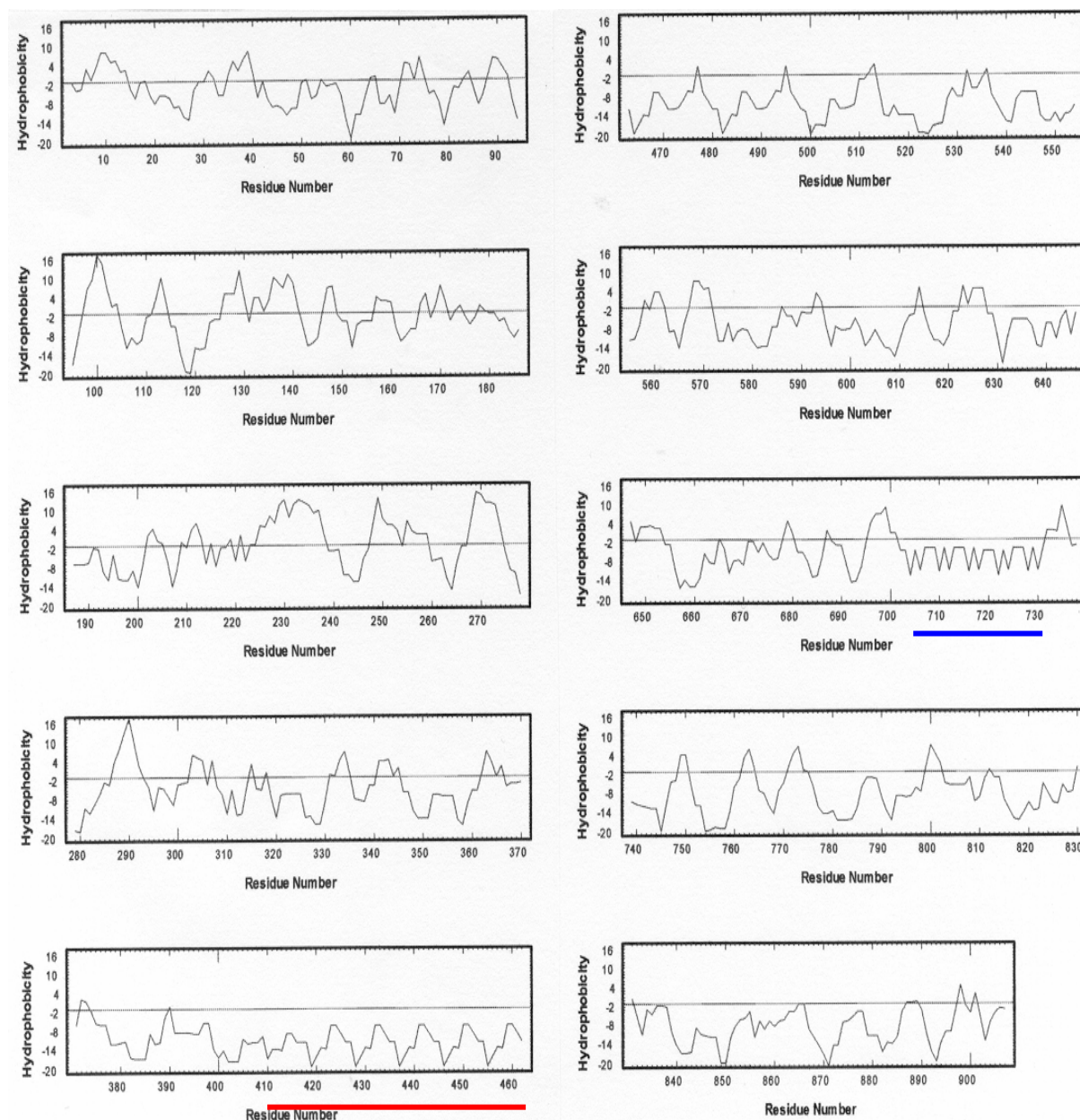


Figure 3.28 The hydrophobicity plot of PfcGAK. The amino acid residue numbers are plotted on the X axis and Y values indicate their hydrophobicity index. Negative Y values indicate hydrophilic residues and positive values indicate hydrophobic residues. The red underlined region shows the nanomeric repeat region and the blue underlined region shows the asparagine rich repetitive region.

3.7.6 Prosite motif search of PfcGAK

Several motifs were identified in the putative PfcGAK protein using the Prosite motif search programme at the EBI (Hulo *et al.*, 2004). They included N-glycosylation sites, protein kinase C and casein kinase phosphorylation sites, a phosphorylation site for tyrosine

kinase, a N-myristoylation site, a prenylation site and a kinase active site, all which are shown in Figure 3.29. Apart from the obvious kinase catalytic domain, three other domains were identified, namely a lysine rich domain, an asparagine rich domain and a domain involved in Golgi dynamics (GOLD). These domains are indicated by shaded text in Figure 3.29. Features of the catalytic domain have already been discussed in section 3.8.1.2.

The kinase catalytic domain starts three amino acids away from the N-terminus of the protein. The lysine rich domain encompasses the nanomeric repeat, which was located in the central region of the protein. The asparagine rich region overlaps with the GOLD domain, which is located close to the C-terminus of PfcGAK.

Proteins with GOLD domains are normally involved in lipid trafficking and associated with the Golgi apparatus. Proteins where the GOLD domain occurs close to extreme termini, usually have an additional lipid interaction domain. GOLD domains are usually of low complexity and between 90 and 150 amino acids long and are thought to be involved in protein-protein interactions. (Anantharaman and Aravind 2002). The PfcGAK GOLD domain is of low complexity and there is a potential prenylation site on the C-terminus of the protein, which may function to anchor the protein in a lipid membrane.

A striking feature was the presence of 18 potential asparagine glycosylation sites with the consensus sequence Asn-x-Ser/Thr (x is any amino acid except proline). In two instances two potential sites are adjacent to each other (amino acid 601-608 and 657-664). It is not known if these are true glycosylation sites or if they are accidental motifs caused by the asparagine rich nature of PfcGAK. *P. falciparum* proteins are thought not to undergo N- and O-glycosylation (Davidson and Gowda, 2001). Folding of the protein plays an important role in glycosylation and it is known that not all potential sites are glycosylated. (Miletich and Broze, 1990). A study of the native PfcGAK protein is required to determine the extent and significance of glycosylation.

The Prosite program identified many potential phosphorylation sites for both protein kinase C (S/T-X-R/K) and casein kinase II (S/T-XX-D/E) in PfcGAK. Some of these sites overlap as seen in Figure 3.29. Nine of the protein kinase C phosphorylation sites (S-x-K) were found in the nanomeric repeat region, which is indicated in bold italics in the lavender shaded lysine rich domain. More recently the motif RxxS/TxR has been shown to be the



Figure 3.29 The amino acid sequence of PfcGAK indicating the various motifs identified by the Prosite program. **Blue** = N-glycosylation sites, **red** = protein kinase C phosphorylation sites, **orange** = casein kinase II phosphorylation site, **pink** = myristoylation site, **purple** = prenylation site, **green** = tyrosine kinase phosphorylation site, **brown** = protein kinase site, **RxxS/TxR** = stringent protein kinase C phosphorylation sites, tyrosine sulphation sites = ▼, **blue shading** = kinase catalytic domain; **lavender shading** = lysine rich domain, **light green shading** = GOLD domain; **light yellow shading** = involucrin site; italics = asparagine rich region. Red asterisks indicate tetra- and tripeptides. Overlapping motifs are indicated above and below by repeating the relevant amino acid sequence.

optimal motif for protein kinase C where the R is a basic amino acid such as arginine, glutamine, asparagine, lysine or histidine. Six potential phosphorylation sites were identified when this consensus sequence was applied and is indicated in red above the amino acid sequence line (Figure 3.29). Kinome analysis showed that *P. falciparum* does not have an obvious homologue to protein kinase C. The phosphorylation sites are of low complexity and most likely appear randomly. On the other hand though, the possibility that phosphorylation on these sites can be effected by other kinases cannot be excluded.

The casein kinase II consensus site SxxE is variable. Serine is preferred to threonine as the phosphorylation site. An acidic amino acid residue is required three residues upstream from the C-terminus of the phosphate acceptor. This is true for only one of the predicted casein kinase sites of PfcGAK. Aspartic acid is preferred to glutamic acid as a proton donor. Additional acidic residues surrounding the phosphate acceptor increase the phosphorylation rate (Pinna *et al.*, 1990).

Of interest is the potential tyrosine kinase phosphorylation site close to the amino terminus. Tyrosine kinase substrates are characterized by a lysine or arginine residue seven residues to the N- terminus of the potential phosphorylation target tyrosine residue. An acidic residue (Asp or Glu) is found 3 or 4 residues N - terminus of the target tyrosine residue. The sequence EEKLISEGAY fulfils these requirements.

Although there is an internal myristoylation site present in PfcGAK it is unlikely that this is a true site since it occurs in the kinase catalytic domain. Proteins undergo cleavage before myristoylation and insertion into membranes take place.

Two potential tyrosine sulphation sites were also identified and are indicated by a (▼) . The first site occurs between amino acid Leu649 and Ser663 with Tyr656 as the target and the second site between amino acid His786 and Phe800 with Tyr793 as the target. No specific consensus site for tyrosine sulphation exists but several of the amino acid residues surrounding the target tyrosine residue are acidic. Several acidic amino acid residues are found close to the tyrosine sulphation residues of PfcGAK. Tyrosine sulphation occurs in the Golgi apparatus and is therefore a common occurrence in secreted and membrane proteins. Tyrosine sulphation is thought to modulate protein-protein interactions. (Kehoe and Bertozzi, 2000).

Some eukaryotic proteins are modified by the attachment of a farnesyl or geranyl-geranyl group to a cysteine residue. This modification takes place on a cysteine residue located three residues upstream from the C-terminus. The two residues between the cysteine and the C-terminus are generally aliphatic. This pattern has been called a Ca₁a₂X-box or prenyl binding site. Some kinases, G-protein subunits, ras proteins, phosphodiesterases and DNA J like proteins have been shown to undergo this modification (Moores *et al.*, 1991). The first aliphatic amino acid (a₁) refers to any aliphatic amino acid, the second (a₂) refers to valine, leucine or isoleucine, and X usually refers to methionine or serine (Alberts *et al.*; 2002). PfcGAK has a potential Ca₁a₂X box at the carboxyl terminus as shown in Figure 3.29. Ca₁a₂X boxes are involved in anchoring proteins in membranes through farnesyl or geranyl isoprenoid units and are frequently found in proteins anchored in the nuclear membrane (Pollard and Earnshaw, 2002, p265). However, the N-terminal lysine residue of PfcGAK is not the optimal substrate for farnesyl- or geranyl-geranyl transferase (Moores *et al.*, 1991, Roskoski, 2003). When the protein was subjected to the Prenylation Prediction Suite program (PrePS) the analysis showed that the motif is not likely to be a true Ca₁a₂X - box (Maurer-Stroh and Eisenhaber, 2005). The 20 amino acids preceding the Ca₁a₂X - box play an important role in substrate recognition by farnesyl transferase and geranyl-geranyl transferase. The PrePS program however, was not trained specifically on *Plasmodium* proteins and it may be that the *Plasmodium* homologous enzymes have different substrate requirements.

A possible involucrin repeat region was identified in PfcGAK between amino acid residue 545 and 554. Involucrin repeats are rich in glutamine, glutamic acid, leucine, proline, glycine, lysine and histidine. Involucrin protein is found in the cytoplasm of terminally differentiated squamous epithelium where it forms part of a cross-linked layer of proteins. Glutamine residues function as substrates for transglutaminase in cross linking reactions (Dijan *et al.* 1993). Human cGAK also has an involucrin-like region between residues 1260 and 1269. The significance of this region in cGAK and PfcGAK is not known.

PfcGAK also displays several tripeptide repeats consisting of either, asparagine (5), lysine (2) or aspartic acid (1). With the exception of three, all the tripeptides are located in the carboxyl domain of the protein. Asparagine as a tetrapeptide also occurred once. It is possible that some of the asparagine or lysine tripeptides are nuclear signal patches since they are rich in basic amino acids (Alberts *et al.*, 2002, p671). The second lysine tripeptide

(DLKKKEDV) occurs close to the C- terminus but does not display all the surrounding conserved residues (PKKKRKV) normally required for a nuclear localization signal. It is also possible that the poly-lysine and asparagine sequences are merely a reflection of the AT richness of the genome. Lysine is coded for by AAA and AAG while asparagine is coded for by AAC and AAU.

Human cGAK shares with PfcGAK the following motifs; a protein kinase catalytic domain, several N-glycosylation sites, casein kinase and protein kinase C phosphorylation sites, myristoylation sites and an involucrin repeat. In addition human cGAK has a proline rich domain, a DNA J domain and clathrin binding sites.

3.8. Bio-informatic analysis of Pfnek-3

3.8.1 Blastp analysis of Pfnek-3

The putative Pfnek-3 protein in PlasmoDB was analysed using the WU-2 blastp program at the European Bioinformatics Institute (www.ebi.ac.uk). A summary of the first 25 hits is presented in table 3.7. The catalytic site comprises the entire protein with the exception of the first 70 amino acids.

Table 3.7 Summary of Pfnek-3 Blastp results using WU-2 at EBI

DB:ID	Source	Length	% Id	% Pos	E ()
UNIPROT:q81629-PLAF7	Serine/threonine-protein kinase Nek1, putative.	347	100	100	6.0e-184
UNIPROT:NEK9_XENLA	Serine/threonine-protein kinase Nek9 (EC 2.7.1.37) (NimA-related protein kinase 9) (XNek9) (Nercc1 kinase).	944	30	58	2.2e-25
UNIPROT:Q6DKD3_XENLA	LOC398600 protein.	944	30	58	2.2e-25
UNIPROT:NEK9_HUMAN	Serine/threonine-protein kinase Nek9 (EC 2.7.1.37) (NimA-related protein kinase 9) (Nercc1 kinase) (NIMA-related kinase 8) (Nek8).	979	33	57	3.8e-25
UNIPROT:NEK9_MOUSE	Serine/threonine-protein kinase Nek9 (EC 2.7.1.37) (NimA-related protein kinase 9).	984	30	55	3.8e-25
UNIPROT:Q69Z43_MOUSE	MKIAA1995 protein (Fragment).	1005	30	55	4.0e-25
UNIPROT:Q7R1M1_GIALA	GLP_28_62487_61384.	367	29	52	7.0e-23
UNIPROT:Q8BSB6_MOUSE	Mus musculus 12 days embryo embryonic body between diaphragm region and neck cDNA, RIKEN full-length enriched library, clone:9430031M01 product:NIMA (never in mitosis gene a)-related expressed kinase 1, full insert sequence.	302	29	52	1.5e-22
UNIPROT:Q7QUR6_GIALA	GLP_231_43409_44626.	405	29	53	2.4e-22

DB:ID	Source	Length	% Id	% Pos	E ()
UNIPROT:Q8CD72_MOUSE	Mus musculus 13 days embryo male testis cDNA, RIKEN full-length enriched library, clone:6030407P11 product:NIMA (never in mitosis gene a)-related expressed kinase 1, full insert sequence.	424	29	52	2.6e-22
UNIPROT:Q86I06_DICDI	Similar to Mus musculus (Mouse). 13 days embryo male testis cDNA, RIKEN full-length enriched library, clone:6030407P11 product:NIMA (Never in mitosis gene a)-related expressed kinase 1, full insert sequence.	1123	32	52	4.6e-22
UNIPROT:Q7R6C5_GIALA	GLP_574_51801_52931.	376	28	51	6.3e-22
UNIPROT:NEK1_MOUSE	Serine/threonine-protein kinase Nek1 (EC 2.7.1.37) (NimA-related protein kinase 1).	774	29	52	1.8e-21
UNIPROT:Q69Z74_MOUSE	MKIAA1901 protein (Fragment).	1234	29	52	3.8e-21
UNIPROT:Q5RFZ0_BRARE	Novel protein (Zgc:56141) (Fragment).	344	29	51	7.2e-21
UNIPROT:Q9N9C3_LEIMA	NEK-related serine/threonine-protein kinase nek1.	555	28	55	1.3e-20
UNIPROT:Q6P3R8_HUMAN	MGC75495 protein (OTTHUMP00000040896).	708	30	53	1.4e-20
UNIPROT:Q96PY6-2	Splice isoform 2 of Q96PY6	1214	28	52	2.7e-20
UNIPROT:Q19530_CAEEL	Hypothetical protein F19H6.1.	302	31	51	2.8e-20
UNIPROT:NEK1_HUMAN	Serine/threonine-protein kinase Nek1 (EC 2.7.1.37) (NimA-related protein kinase 1) (NY-REN-55 antigen).	1258	28	52	2.8e-20
UNIPROT:Q5H9T1_HUMAN	Hypothetical protein DKFZp686D06121.	1287	28	52	2.9e-20
UNIPROT:Q659I9_SPHGR	Nek protein.	893	28	49	3.5e-20
UNIPROT:Q7QQZ8_GIALA	GLP_442_9173_11869.	898	29	51	5.8e-20
UNIPROT:Q7ZUN2_BRARE	NIMA (Never in mitosis gene a)-related kinase 2 (Zgc:55602).	440	31	53	6.6e-20

Abbreviations of organisms found in Table 3.7 PLAF7 = *Plasmodium falciparum*; XENLA = *Xenopus laevis*; GIALA = *Giardia lamblia*; LEIMA = *Leishmania major*; CAEEL = *Caenorhabditis elegans*; SPHGR = *Sphaerechinus granularis*; BRARE = *Danio rerio*, CANGA = *Candida glabrata*, DICDI = *Dictyostelium discoideum*, YEAST = *Saccharomyces cerevisiae*

The Blastp result of the first 50 hits is shown in appendix B.4. The Blastp analysis showed that PfNek-3 has similarity to many serine/threonine protein kinases of the Nek 9 and Nek 1 families. Nek is an alternative name for the NIMA kinases, given to NIMA related kinases of eukaryotic organisms.

3.8.2. Comparative amino acid sequence analysis of Pfnek-3

Figure 3.30 shows the alignment of Pfnek-3 with various Nek related proteins. The kinase consensus line is shown above the amino acid sequence and the conserved amino acids in this alignment are shown underneath the amino acid sequence. Pfnek-3 was most similar to

the Nek9 and Nek1 families of protein kinases. The first 70 amino acids of Pfnek-3 showed no similarity to the Neks used in this alignment. In the kinase catalytic domain all the invariant residues required for kinase catalytic activity were conserved with the exception of the glycine residue in subdomain I. This residue is replaced by a methionine residue in this alignment, which although non-polar, is structurally very different from glycine.

Pfnek-3

NEK9_XENLA
NEK9_HUMAN
NEK9_MOUSE
NEK1_MOUSE
Q5RFZ0_BRARE
NEK1_HUMAN

MVCIIYLCFLFHFLFVVNTMRILCNLLIVICMYMFPFYSLFKFVKMLSLLLPSILSC 60
-----MSALGRYDRHCDSINSDFGSDSVRSCG----- 26
-----MSVLGEYERHCDSINSDFGSESGGCGDSSPGPSASQGPPRA 40
-----MSVLGEYERHCDSINSDFGSESGGGGDSPGPSAVPGRPRA 40

-----Subdomain I----- --Subdomain II---- --Subdo-
o ogxGxxg-v oaoK-o E
F → L mutation

Pfnek-3
NEK9_XENLA
NEK9_HUMAN
NEK9_MOUSE
NEK1_MOUSE
Q5RFZ0_BRARE
NEK1_HUMAN

EKKYQVYKNNG YKFETVLDFTMSDSEIHLIRISIE SDEIYISKVVDIYGINED DLNKYMNE 120
----PEQEELH YIPIRVLGHGAYG-EATLYRRTE DDSLVVWKEVGRLARLSEK ERRDALNE 81
GGGAEEQEELH YIPIRVLGKGAFG-EATLYRRTE DDSLVVWKEVDLTRLSEK ERRDALNE 99
GGGAEEQEELH YIPIRVLGKGAFG-EATLYRRTE DDSLVVWKEVDLTRLSEK ERRDALNE 99
-----MEK YVRLQKIGEGSFG-KAVLVKSTE DGRHYVIKEINISRMSDK ERQESRRE 51
-----MDK YERLKKIGEGSFG-KAILVKSTRT DGRQYVIKEIGISRMSNK ERQESRKE 51
-----MEK YVRLQKIGEGSFG-KAILVKSTE DGRQYVIKEINISMSSK EREESRRE 51

* : * : .. : * : .. : * : .. : * : .. : *

Y o L b o K o z E

main III -Subdomain IV- -----Subdomain V----- -----Sub-
--oo h--oo-o---o--- ooooo*oo---o---o----- o-----

Pfnek-3
NEK9_XENLA
NEK9_HUMAN
NEK9_MOUSE
NEK1_MOUSE
Q5RFZ0_BRARE
NEK1_HUMAN

LYIMNKL NCEINIVNIIDYIKEN DTLSFILEFCNQGLHSDILRKLLNNE IYTESEIFNI 180
IVILSLQ -HDNIIAAYNHFLDS NTLLIELEYCNGGNLFDPKIVHQKA--Q LFQEETVLWY 138
IVILALLQ -HDNIIAAYYNHFMDN TTLLIELEYCNGGNLYDKILQKD--K LFEEEMVVWY 156
IVILALLQ -HDNIIAAYYNHFMDN TTLLIELEYCNGGNLYDKILQKD--K LFEEEMVVWY 156
VAVLANMK -HPNIVQYKESFEEN GS LYIVMDYCEGGDLFKKRINAQKG--A LFQEDQILDW 108
VAVLANMK -HPNIVQYKESFEES GC LYIVMDYCEGGDLFKKRINNQRG--S LFFPEEQILDW 108
VAVLANMK -HPNIVQYRESFEEN GS LYIVMDYCEGGDLFKKRINAQKG--V LFQEDQILDW 108

: : : ** : : : * : : : * : : : * : : : *

o o o NI o z L o ozuC G L I bb ou E o

Pfnek-3 does not share the conserved glutamic acid and arginine sequences found on the N-terminal region of subdomain III with the other Nek proteins in this alignment. The central invariant glutamic acid residue as well as one of the two non-polar residues close to the C-terminus, common to serine/threonine kinases, were conserved in this subdomain.

Subdomain IV is characterised by a nearly invariant histidine residue close to the N-terminus and four conserved non-polar residues. This histidine residue was not conserved for Pfnek-3 but three of the four non-polar residues were conserved. In this alignment an asparagine and an isoleucine residue were fully conserved. There was conservation of an acidic residue adjacent to the C-terminal residue.

In subdomain V of the serine/threonine kinases a cluster of seven non-polar residues interrupted by a polar residue is found near the N-terminus. Three of the seven non-polar residues and a polar residue were conserved which included the fully conserved leucine residue. Pfnek-3 has four non-polar residues in this cluster. The two non-polar consensus residues closer to the C-terminus were conserved with the identical residues, leucine and isoleucine. This domain of the Neks included a conserved cysteine, which formed part of the non-polar cluster, and a conserved glycine residue.

All the usual features of subdomain VIA were conserved. This domain has a number of non-polar residues as well as a polar residue and a small uncharged residue. Close to the C-terminus a nearly invariant His residue is common (Hanks and Hunter 1995). This residue was conserved for Pfnek-3 and the other Nek kinases. The polar and small neutral residue sites were replaced by non-polar residues for Pfnek-3 and the other Nek kinases.

Subdomain VIB is critical for kinase activity (Hanks and Hunter 1995). The critical consensus sequence Dok+xNooo was fully conserved in Pfnek-3 and the comparative Nek sequences. A single gap was introduced close to the C terminus of Pfnek-3. Pfnek-3 had the unusual catalytic site sequence of HGD~~L~~KSTN whereas the other Nek proteins had the more common HRDIKTLN or HRDIKSQN sequences. The common conserved catalytic motif for this PfNek-3 alignment was IoHxDok+xNIF.

The sequence aspartic acid-phenylalanine-glycine (DFG) is a common feature of Subdomain VII. Only the aspartic acid residue of this motif is invariant and essential for kinase activity (Hanks and Hunter 1995). In Pfnek-3 the sequence GDFG is found which

correlates with the conserved sequence of the Nek1 group of kinases. The Nek9 group of kinases has the conserved sequence GDYG. On the N terminus of this subdomain two polar residues flank a nearly invariant lysine residue. Despite the high level of conservation in this subdomain there is a gap at the carboxy terminus, which may have an effect on the local folding of the subdomain.

In typical serine/threonine kinases subdomain VIII is characterised by a nearly invariant glycine residue close to the N-terminus and an invariant glutamic acid residue preceded by a nearly invariant proline residue. The invariant glutamic acid residue is followed by two non-polar residues. Alanine-proline-glutamic acid (APE) is a common sequence in serine/threonine kinases (Hanks and Hunter 1995). In the case of the Nek kinases this motif was replaced with serine-proline-glutamic acid and at least one non-polar conserved residue. The Pfnek-3 sequence was divergent from this in that tyrosine replaced the proline residue to form SYE. The second position from the glutamic acid is a non-polar residue.

Subdomain IX has an invariant aspartic acid close to the N-terminus followed by a cluster of non-polar residues with a nearly invariant glycine residue in the middle and a nearly invariant proline residue close to the C-terminus (Hanks and Hunter 1995). The aspartic acid and the glycine residues as well as several other amino acids were conserved in this subdomain. The Nek kinases had the consensus sequence KSDIWA whereas Pfnek-3 had the sequence LSDLFQ. The cluster of non-polar residues surrounding the conserved glycine residue was conserved between Pfnek-3 and the other Nek kinases with the sequence oGCoooEo. Furthermore threonine-leucine and phenylalanine-x-alanine sequences were conserved between Pfnek-3 and the other Nek sequences closer to the C-terminus.

In this alignment there was poor conservation in subdomain X with no identical sequences observed. Two non-polar residues were conserved. Subdomain XI had four (including the conserved leucine) of the five semi-conserved non-polar sequences and the consensus arginine residue. In this alignment the conserved N-terminus tyrosine residue was conserved followed by a semi-conserved serine/threonine residue. There is no conservation observed in the short carboxyl domain, consisting of 24 amino acids, that follows the kinase catalytic domain. Figure 3.31 is a graphical representation and comparison of the kinase catalytic region with all the major consensus sequences indicated.

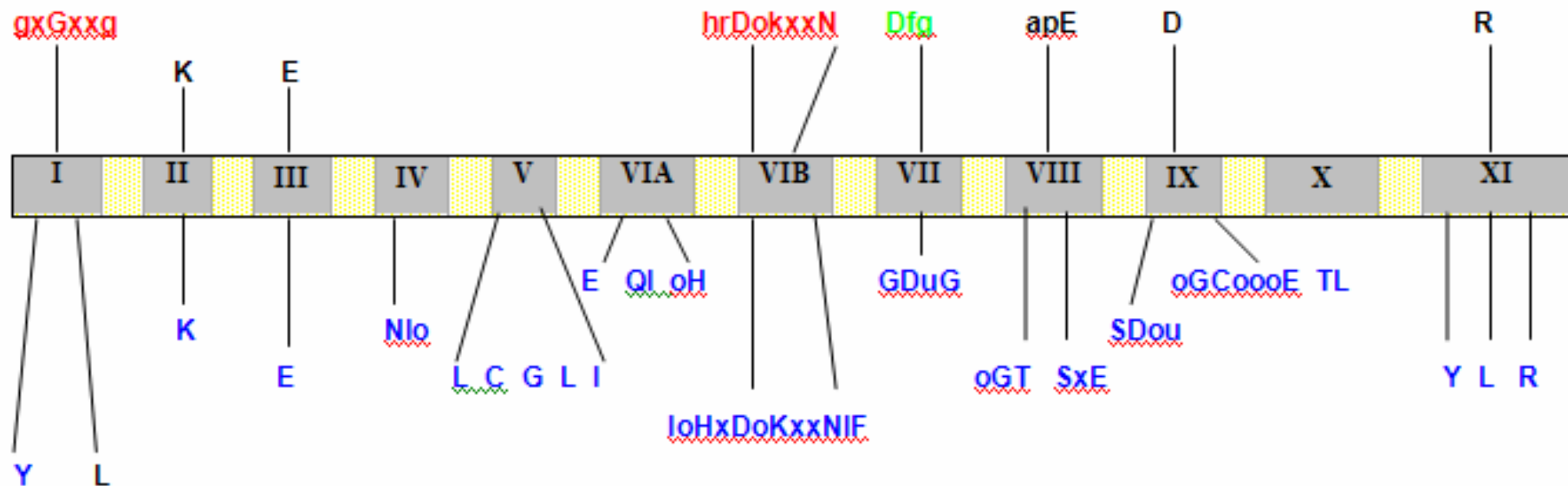


Figure 3.31 Comparison of the major consensus amino acid motifs in the catalytic domain of Ser / Thr protein kinases with Pfnek-3 and a group of homologous Nek proteins. The twelve subdomains of the protein kinase catalytic domain are indicated by the central bar and consensus residues for protein kinases are indicated above the central bar. The consensus motifs in red are conserved in the Nek homologues but not in Pfnek-3. The **DFG** consensus is conserved in Pfnek-3 but not in the aligned Neks. The corresponding conserved residues and motifs for Pfnek-3 and other Nek residues are indicated beneath the central bar. Conserved amino acids are shown in capital letters , o refers to semi- conserved non-polar residues, u refers to aromatic amino acid residues, x indicates any amino acid.

3.8.3 Prosite motif search of Pfnek-3

Figure 3.32 shows the various motifs of Pfnek-3 identified by the Prosite programme. The putative Pfnek-3 sequence has the obvious protein kinase domain, two potential N-glycosylation sites and three protein kinase C phosphorylation sites that are distinct from the four casein kinase phosphorylation sites. Not one of the predicted protein kinase C phosphorylation sites fulfilled the more stringent RxxS/TxR requirement. However, NIMA kinases have a conserved F_{K/R}K_{R/T} phosphorylation site where the phenylalanine and threonine residues are fully conserved (Pu *et al.*, 1992). Two such sites were identified in Pfnek-3. Close to the N-terminus, a signal sequence, a *Plasmodium* export element (PEXEL) motif and a transmembrane sequence was identified.

A putative signal sequence consistent with an endoplasmic reticulum localisation signal was identified between amino acid 2-19 by both Prosite and SignalP 3.0 (Hulo *et al.*, 2004, Bendtsen *et al.*, 2004). This was confirmed by the hydrophobic nature of this area (Figure 3.33) and the absence of a cluster of positive amino acids normally found in nuclear localisation signals. The putative signal sequence was followed by a PEXEL motif, which has the consensus core 'RxLxQ/E' (Templeton and Deitsch, 2005). Pfnek-3 has the sequence RILCN. The presence of the terminal asparagine (N) in place of a glutamate (E) or glutamine (Q) in Pexel motifs are not unusual and are described by Hiller *et al.*, 2004.

```

          *****
1    MVCIYLFCFL LFHLFVVNTR MRILCNLLLIV ICEYMFPPFYS FILFKFVKML SLLLPSILSC
61  EKKYQVYKNN GYKFETVLDF MTSDSEIHLI RSIESDEIYI SKVYDIYGIN EDDLNKYMNE
121 LYIMNKLRNC ENIVNIIDYI KENDTLSFIL EFCNQGDLHS DILRKKLNNE IYTESEIFNI
181 LHQILNGLNI IHQNGIIHGD LKSTNIFIKD NKIKIGDFGI SSEQSSNNNL GTLNCLSYES
          GIIHGD
241 IKFKKTNKLS DLFQVGCILY ELATLSSPFC ATNINDMISL FEDKNYKSYI IKNISSIYSQ
301 KLVNVISKLL SLNTLRERLEV ITNYNLETTQ HVHLENANKI ISCITVQ

```

Figure 3.32 The amino acid sequence of Pfnek-3 indicating the various motifs identified by the Prosite program. Blue = N-glycosylation sites, red = protein kinase C phosphorylation sites, orange = casein kinase phosphorylation site, pink = myristoylation site, purple = protein kinase site, green = lipid attachment site, dark green = signal sequence; aqua = transmembrane sequence; * = PEXEL motif . Lavender shading = NIMA consensus phosphorylation sites. Overlapping motifs are indicated below the amino acid sequence.

PEXEL motifs have also been termed vacuolar transport signals (VTS's). PEXEL sequences have been shown to target parasite encoded proteins across the parasitophorous vacuole to the erythrocyte membrane and cytoplasm (Marti *et al.*, 2004, Hiller *et al.*, 2004).

The PEXEL motif is followed by a stretch of predominantly hydrophobic amino acids from residue 25 to 55. The hydrophobicity plot in Figure 3.33 indicates that this region is hydrophobic and may represent a transmembrane region. PlasmoDB predicts a transmembrane region between acid sequence 21-43 (www.PlasmoDB.org). The PEXEL motif falls within this domain. The hydrophobic region is followed by a cluster of amino acids that is enriched in basic amino acids (55-60).

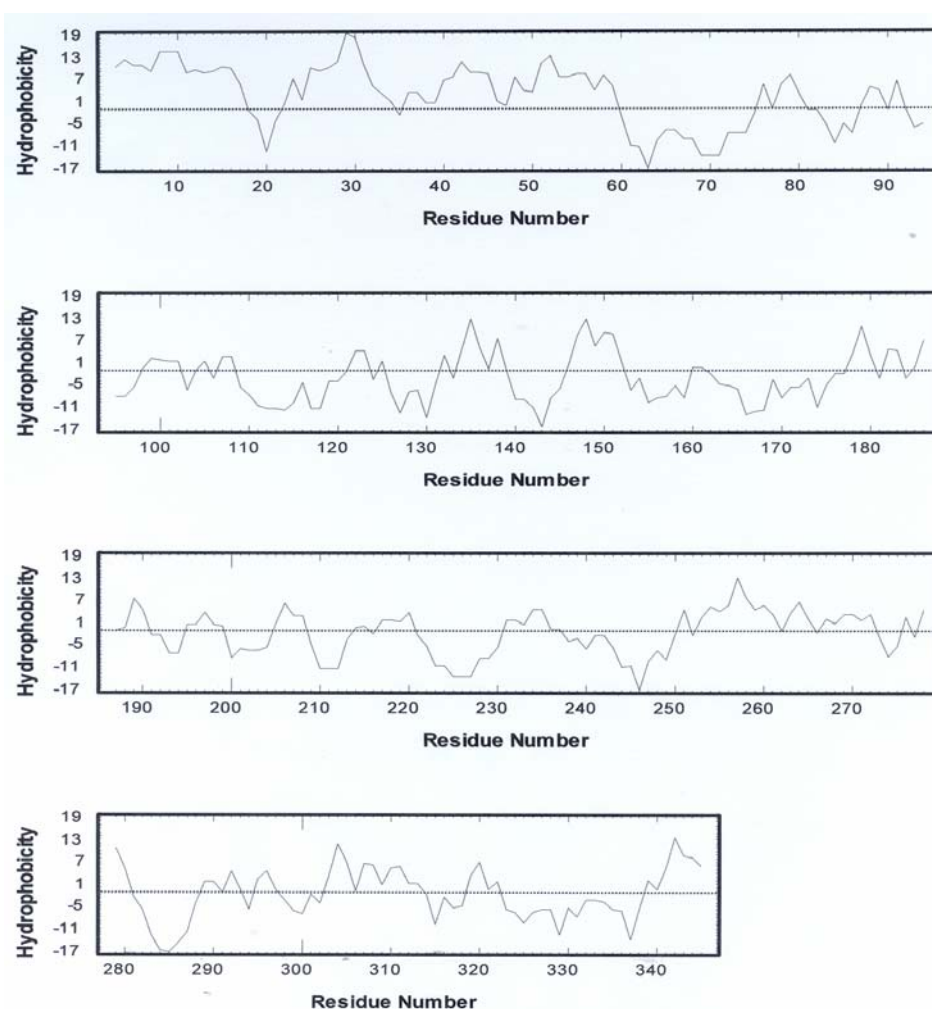


Figure 3.33 The hydrophobicity plot of Pfnek-3. The amino acid residue numbers are plotted on the X axis and Y values indicate their hydrophobicity index. Negative Y values indicate hydrophilic residues and positive values indicate hydrophobic residues.

Although two myristoylation sites were identified they are not likely to be of any significance since they are located in the catalytic domain and it is unlikely that this protein undergoes proteolytic degradation to an active form. Myristoylation takes place after proteolytic cleavage of a protein and is a mechanism used to anchor proteins into membranes. A prokaryotic membrane lipoprotein lipid attachment site was identified between amino acid residues 50 and 60. Although the sequence does not conform to all the stringent requirements it is possible that this sequence may form a hydrophobic region that interacts with membrane lipid molecules or forms part of a transmembrane sequence.

NIMA and Nek kinases frequently have a PEST sequence in the carboxyl end of the protein, which indicates the rapid proteolytic degradation of a protein. When Pfnek-3 was subjected to the PESTfind program (<http://www.at.emblnet.org/toolbox/pestfind/>) no potential PEST sequence was identified.

3.9 The effect of kinase inhibitors on *Plasmodium falciparum*

3.9.1 Screening and selection of compounds

In an initial screening process 565 compounds derived from known kinase inhibitors, mostly purine derivatives, were tested for antimalarial activity using the [³H]-hypoxanthine incorporation assay as described in chapter 2 section 2.11. The compounds were supplied by Dr. L. Meijer from the CNRS, Roscoff, France.

Compounds were screened at a concentration of 2 μ M over a single (48 hrs) and a near double (76 hrs) cycle of exposure. This allows for one cycle of reinvasion and subsequent maturation to schizonts. A subset of purines was selected for further study as a result of their inhibitory activity against the parasite or their structural similarity to purines with high activity. Some compounds showed large discrepancies in their inhibitory effect when the two exposure times were compared, therefore IC₅₀ values were determined after 76 hrs of exposure which include reinvasion by merozoites and near completion of a second round of DNA synthesis.

With the exception of isopentenyladenine and compound 101, all these selected molecules were 2,6,9-tri-substituted purine derivatives with an amino group at C2 and a benzylamino or anilino group at C6. The dose-response curves, IC₅₀ values, and the relevant chemical structures are presented in figures 3.34, 3.35, 3.36, and 3.37. The structure activity relationships as well as a comparison of antimalarial activity to their effect on the isolated CDK1/cyclin B system is discussed in chapter 4. A report describing this work was published in *Biochemical Pharmacology* (Harmse *et al.*, 2001).

3.9.2 Antimalarial activity of iso-olomoucine, olomoucine, roscovitine and isopentenyladenine

Olomoucine, roscovitine and isopentenyladenine displayed similar activity against *P. falciparum* with IC₅₀ values of 8.45, 5.36, and 8.34 μ M, respectively as shown in Figure 3.34. Iso-olomoucine, an isomer with no inhibitory effect on protein kinases, (personal communication, Dr. L. Meijer) showed minimal activity against *P. falciparum*, with an IC₅₀ value of 235 μ M (data not shown).

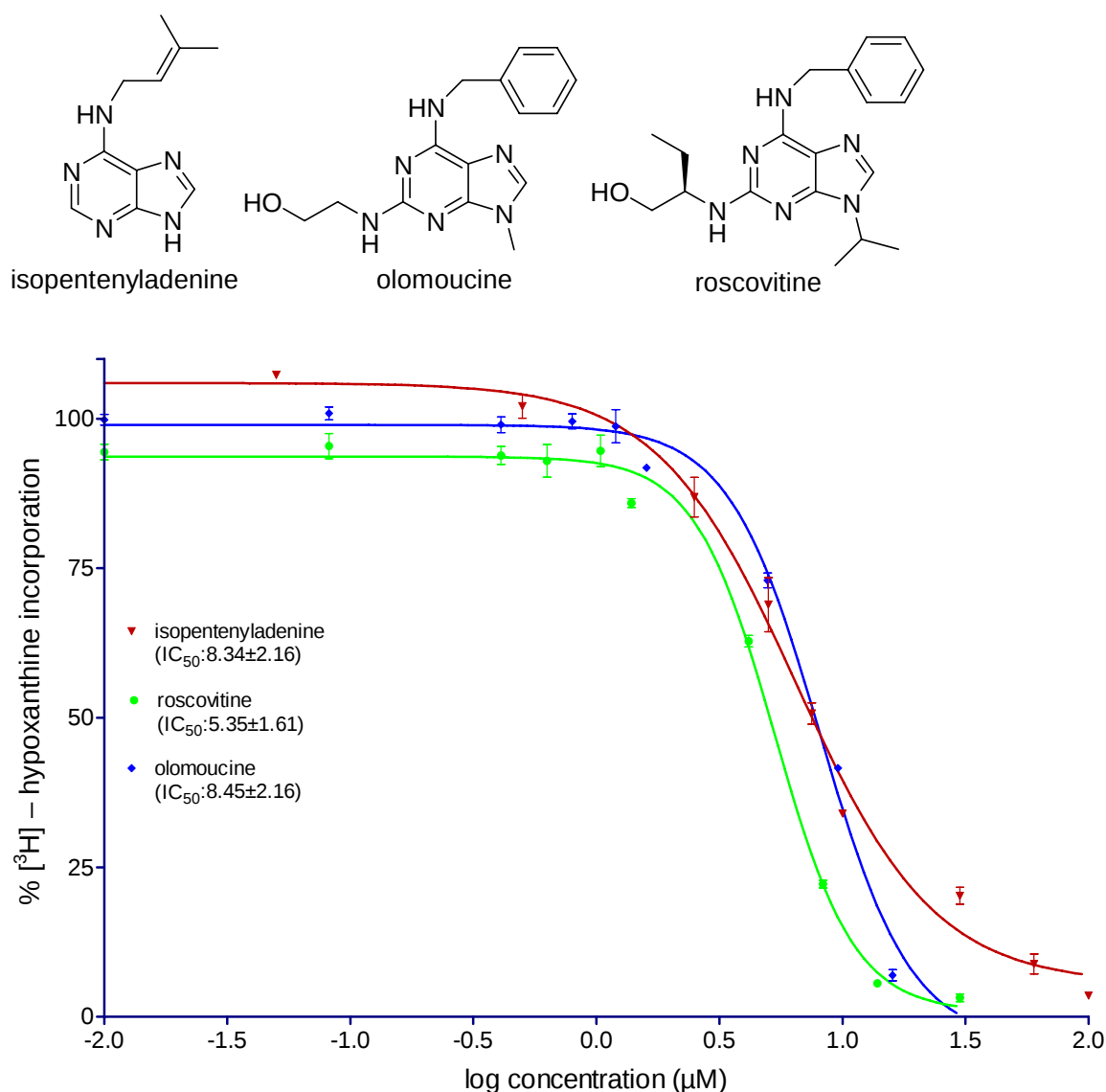


Figure 3.34 Dose response curves showing the effect of olomoucine, roscovitrine and isopentenyladenine on the *P. falciparum* FCR-3 strain. Each point represents the mean and standard deviations of the quadruplicate determinations of each specific concentration of one representative experiment.

3.9.3 Antimalarial activity of the purvalanol series of compounds

The purvalanol series of compounds had a wide range of inhibitory activity against *P. falciparum* as shown in Figure 3.35. Within this family, the *R* and *S* isomers of purvalanol A displayed similar inhibitory activity against *P. falciparum* proliferation, with IC₅₀ values of 0.55 and 0.42 μM, respectively.

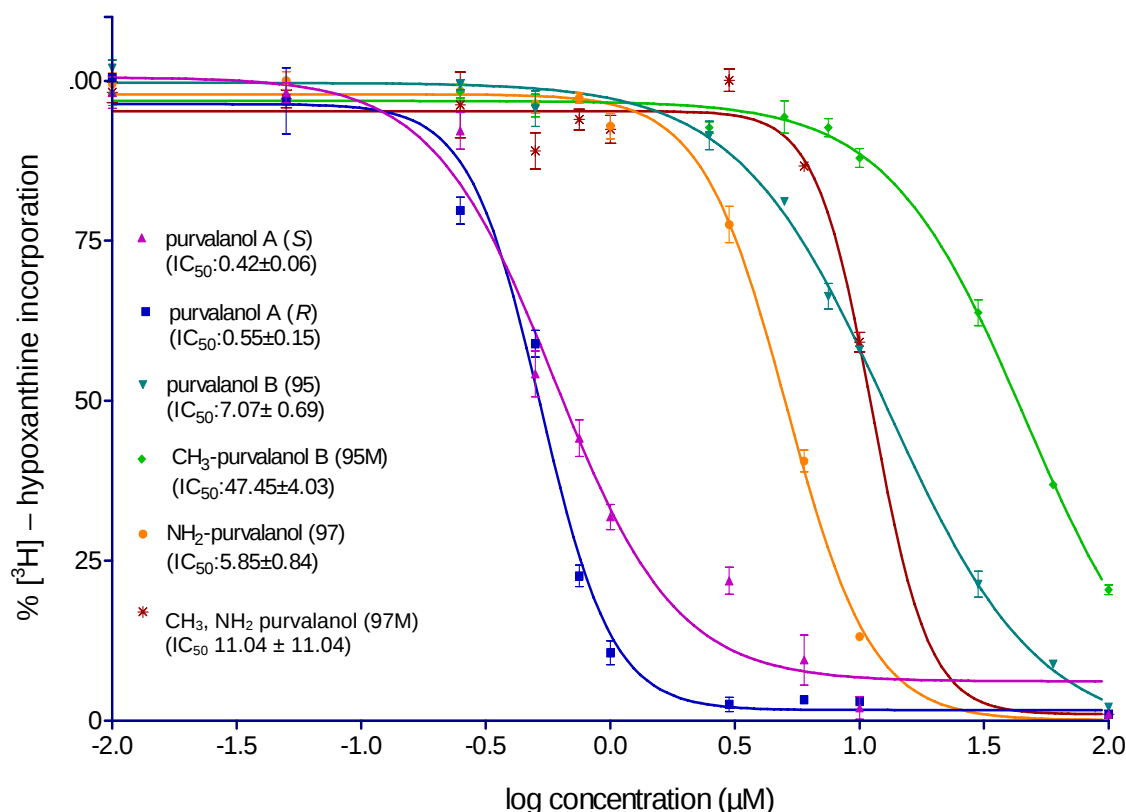
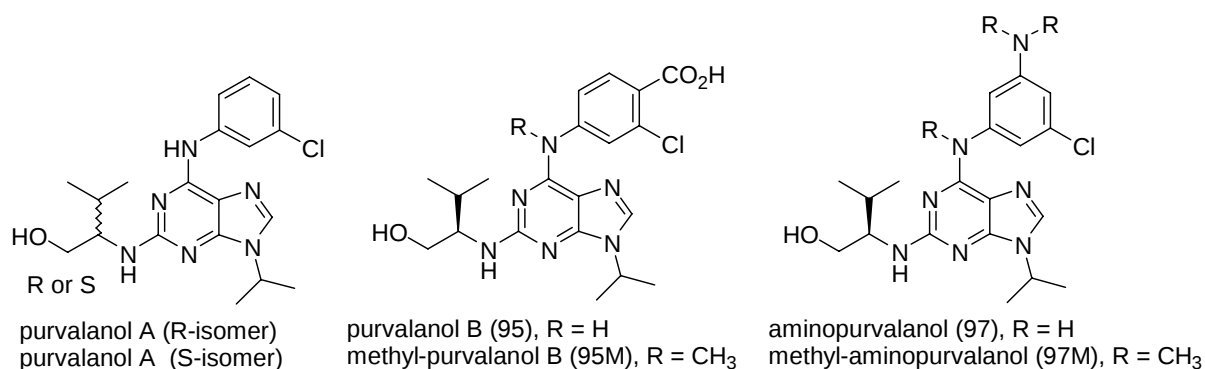


Figure 3.35 Dose response curves showing the effect of the various purvalanol derivatives on the *P. falciparum* FCR-3 strain. Each point represents the mean and standard deviation of the quadruplicate determinations of each concentration tested of one representative experiment.

Purvalanol B (95), a potent CDK inhibitor, had an average effect on *P. falciparum* with an IC₅₀ value of 7.0 μM. The methylated derivative of purvalanol B (95M), which has no protein kinase inhibition properties, showed relatively poor activity against *P. falciparum* with an IC₅₀ value of 47.5 μM. Amino-purvalanol (97) did not significantly increase the activity against *P. falciparum* compared to the carboxylated derivative (compound 95), and decreased the IC₅₀ value to 5.8 μM. Methylation of 97 (97M) decreased its antimalarial activity to an IC₅₀ value of 11.04 μM.

3.9.4 Selective inhibitory activity of compounds 99 and 101 against

P. falciparum

The two purine derivatives, compounds 99 and 101, shared common substitutions on C2 (3-trifluorobenzylamino) and C6 (4-aminobenzylamino). These compounds were active against *P. falciparum* at submicromolar concentrations, with IC_{50} values of 0.83 and 0.63 μM respectively, as shown in Figure 3.36).

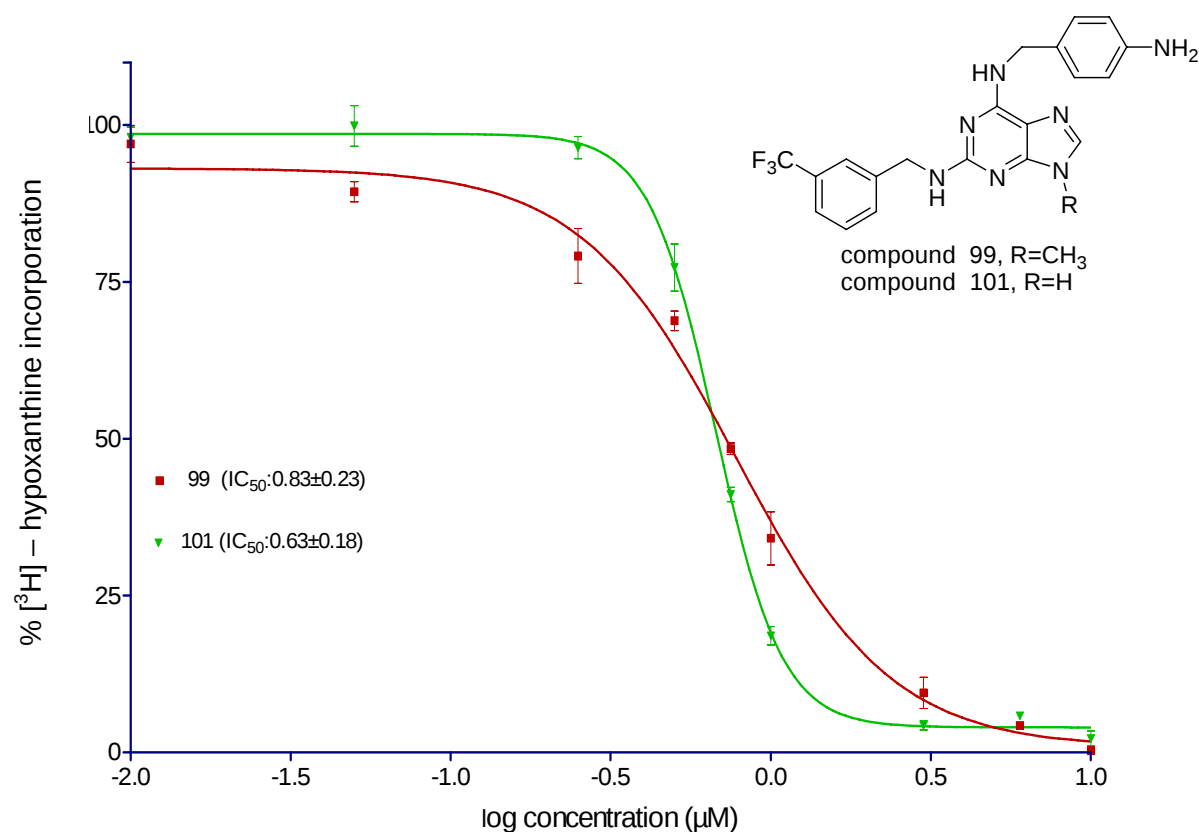


Figure 3.36 Dose response curves showing the effect of compound 99 and 101 on the *P. falciparum* FCR-3 strain. Each point represents the mean and standard deviation of the quadruplicate determinations of each concentration tested of one representative experiment.

3.9.5 Antimalarial activity of R and S phenyl substitutions on the

C2-hydroxyethyl side chain

Compounds 59 and 66 are enantiomers with bulky phenyl substitutions on the chiral carbon of the C2 - hydroxyethylamino side chain. Compound 59 was active against *P. falciparum* with an average IC_{50} value of 0.53 μM , whereas compound 66 showed no activity against the parasite at the highest concentration tested (10 μM) as shown in Figure 3.37.

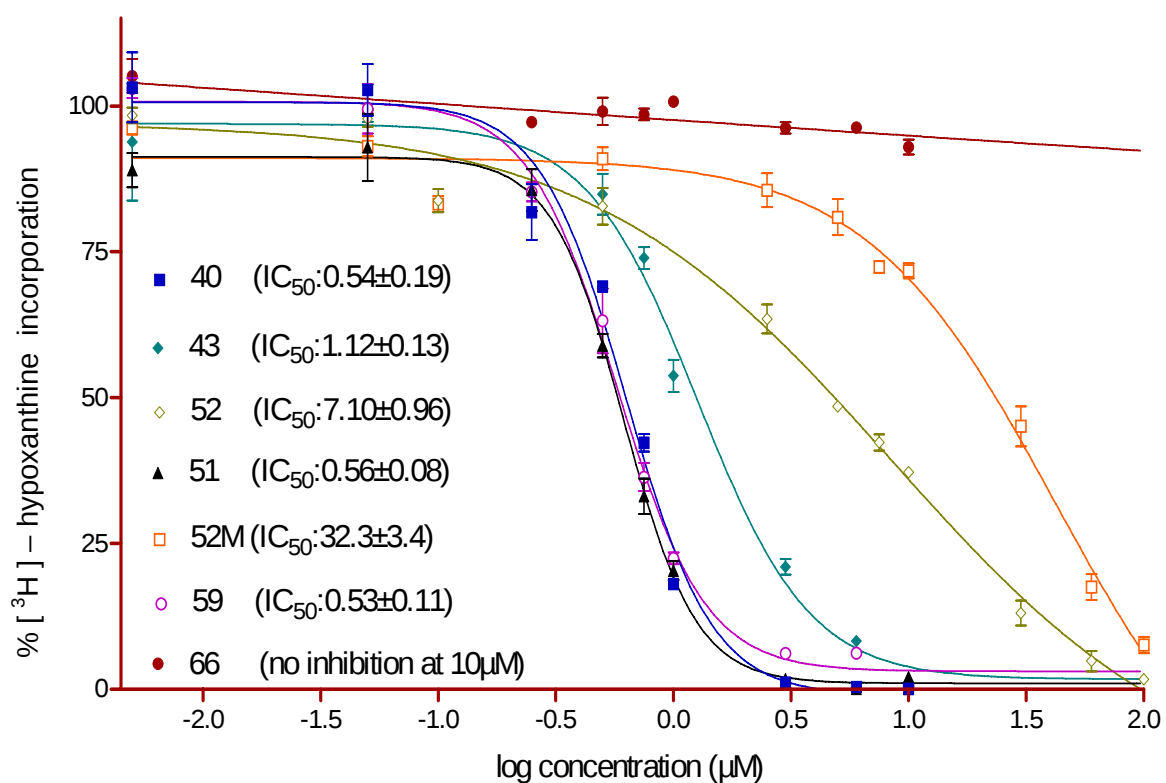
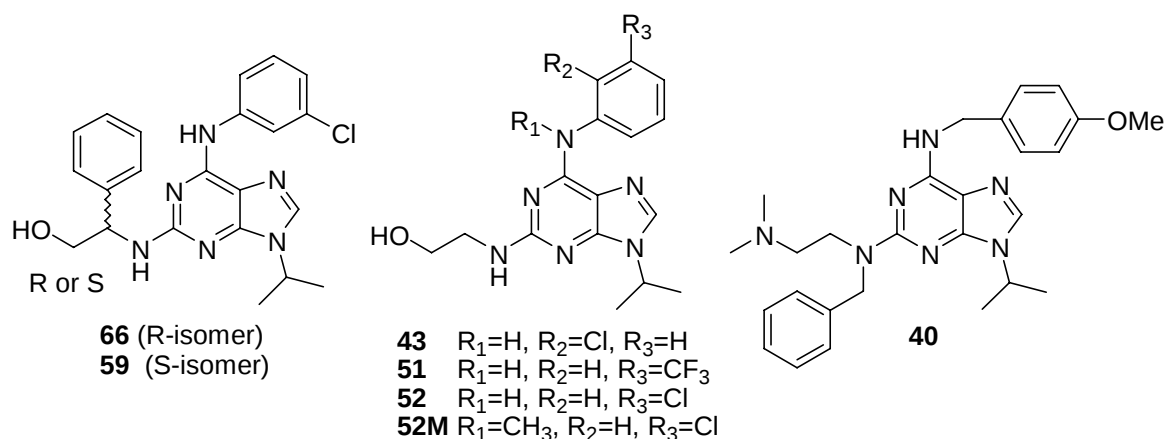


Figure 3.37 Dose response curves showing the effect of the various purine derivatives on the *P. falciparum* FCR-3 strain. Each point represents the mean and standard deviation of the quadruplicate determinations of each concentration tested of one representative experiment.

3.9.6 The effect of halogen substitutions on C6 on antimalarial activity

Compounds 43, 51 and 52 are structurally similar, differing only in the substitutions on C6 and are shown in Figure 3.37. Nevertheless, considerable variation in the IC_{50} values of these compounds on *P. falciparum* (IC_{50} values of 0.56 and 7.10 μM for compound 51 and 52, respectively) was observed. Compound 43 had an IC_{50} value of 1.12 μM when tested

against *P. falciparum*. This was higher than that of compound 51 but much lower than the IC_{50} of its isomer, compound 52 (IC_{50} :7.1 μ M). Methylation of compound 52 reduced the antimalarial activity to an IC_{50} value of 32.3 μ M.

3.9.7 Antimalarial activity of compound 40

This compound has an unusual structure with bulky substitutions on the C2 and C6 rings as seen in Figure 3.37. It inhibited the growth of *P. falciparum* with a sub-micromolar IC_{50} value of 0.54 μ M that compared well with that of the other purines tested.

3.9.8 The effect of purine inhibitors at an IC_{90} concentration on the development of the erythrocytic stages of parasite

The effects of roscovitine and olomoucine on the progression of the parasites through their normal erythrocytic cycle were assessed at a concentration equivalent to the IC_{90} and are presented in Figures 3.38 – 3.40. Both compounds caused a similar decrease in the total parasitaemia compared to the control as seen in Figure 3.38.

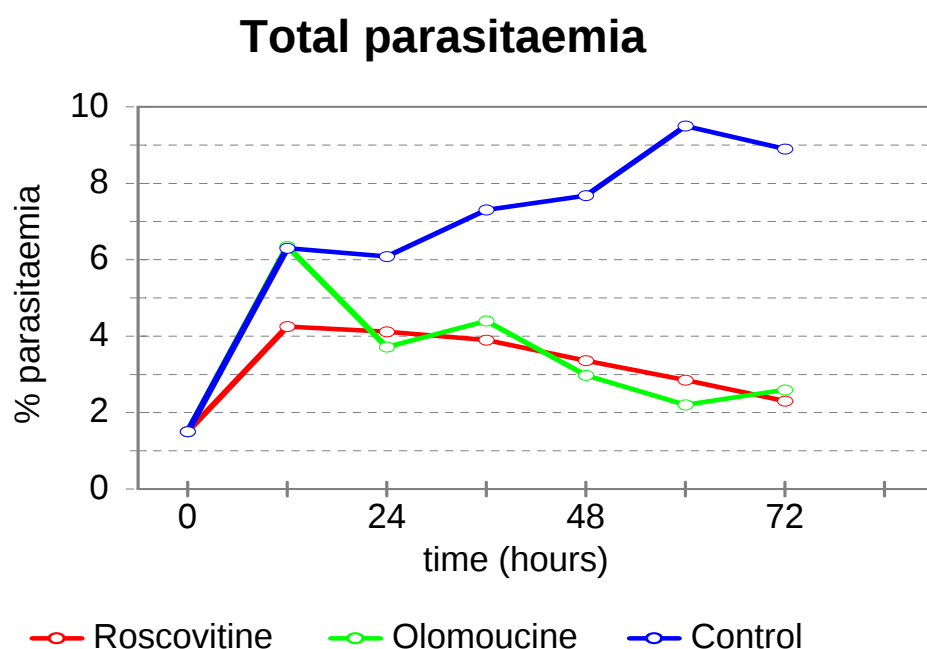
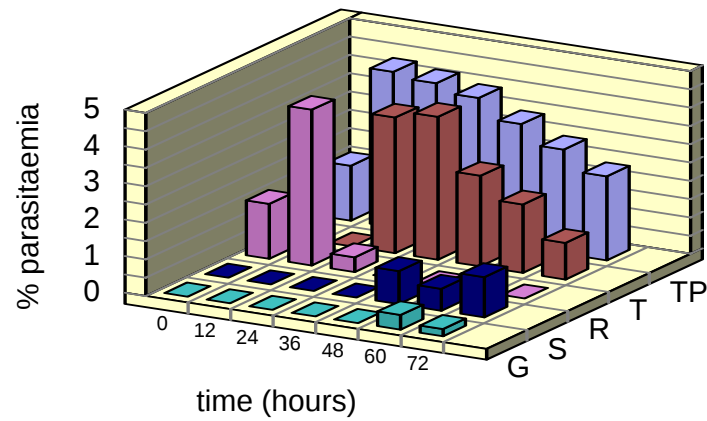
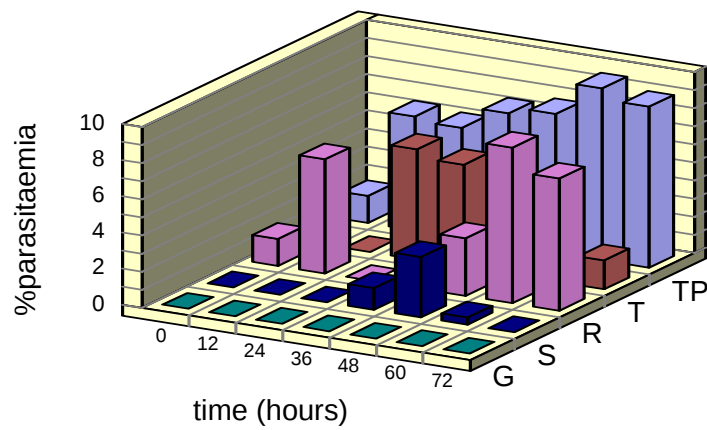


Figure 3.38 The effect of roscovitine and olomoucine at an IC_{90} on *P. falciparum* growth. At time 0 hrs parasites were mostly in the ring stage.

Roscovitine



Control culture



Olomoucine

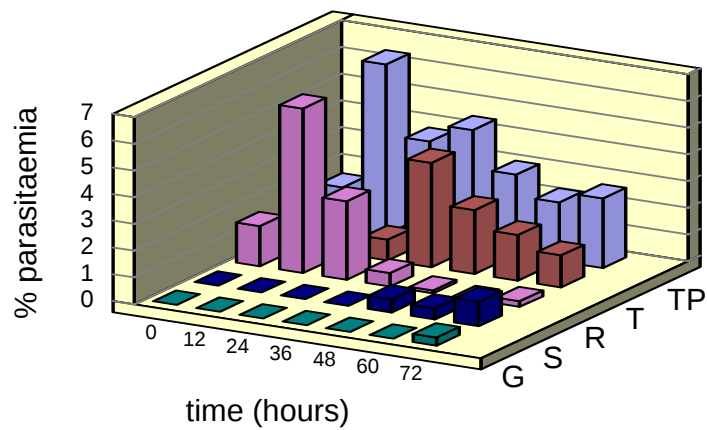
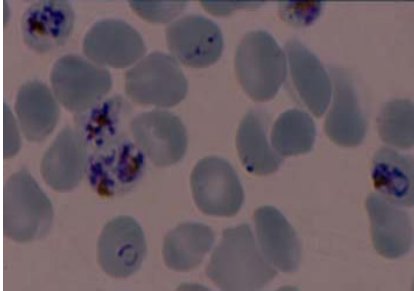
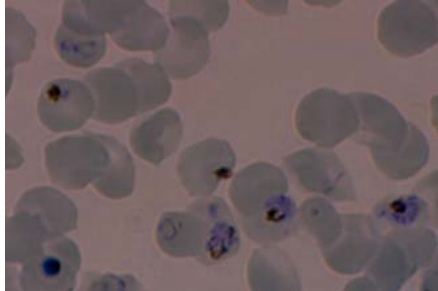


Figure 3.39 The effect of olomoucine and roscovitine at an IC_{90} on the stages of the erythrocytic cycle of the *P. falciparum* FCR3 strain. G=gametocytes ; S = schizonts
R = rings, T = trophozoites, TP = total parasitaemia

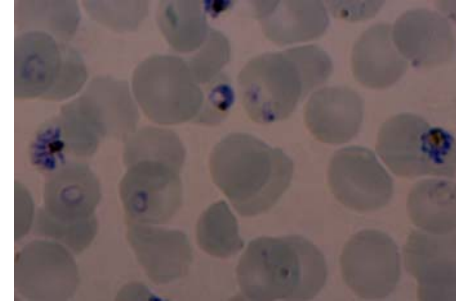
Control 36 hrs



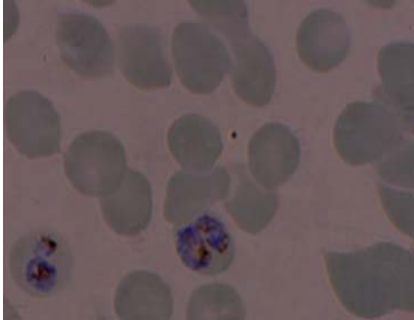
Control 60hrs



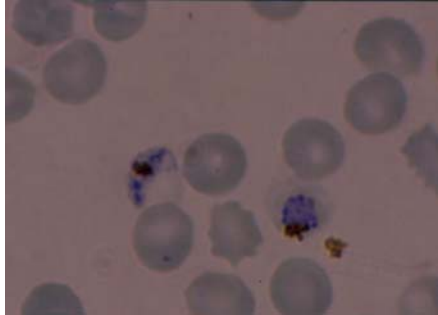
Control 72 hrs



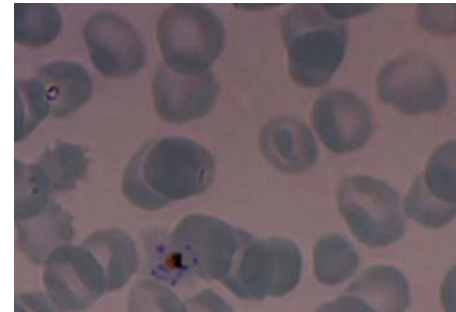
Olomoucine 36 hrs



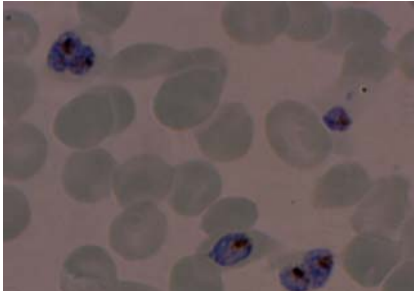
Olomoucine 60 hrs



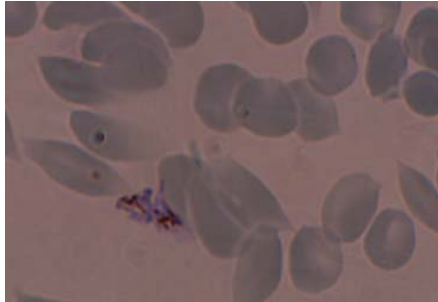
Olomoucine 72 hrs



Roscovitine 36 hrs



Roscovitine 60 hrs



Roscovitine 72 hrs

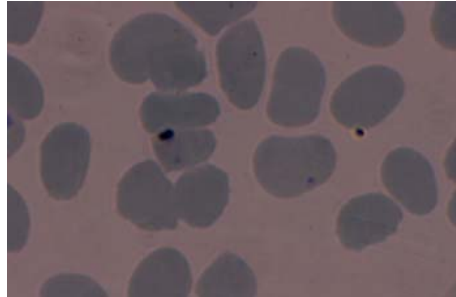


Figure 3.40 Photographs showing the effect of olomoucine and roscovitine at IC_{90} concentrations on *P. falciparum* erythrocytic stages.

The control culture cycled normally as seen in Figure 3.39. The number of rings increased to 6% after 12 hours due to the presence of some schizonts in the seeding culture. At 24 hrs post seeding, the number of trophozoites increased as the rings matured. Between 36 and 48 hours schizonts were present in the culture followed by an increase in rings from 48hrs to 72 hrs and a decline in the number of trophozoites. The total parasitaemia increased to above 9% after 60 hrs. The slight drop in total parasitaemia at 72 hours may have been due to nutrient depletion in the culture medium or to inaccurate quantification. Formation of gametocytes was not observed in this culture.

The roscovitine (Figure 3.39) treated culture initially followed a similar pattern to the control albeit with a decreased total parasitaemia compared to the control culture. The formation of rings from 48-72 hrs did not take place as in the control culture, although low numbers of schizonts were present. It appears as if the cultures progressed normally to the trophozoite stage but the progression to the schizont stage and next ring stage were inhibited. The culture exposed to olomoucine followed the same pattern as roscovitine as shown in Figure 3.39. It is possible that the kinase inhibitors act by affecting the viability of the schizonts or the ability of the merozoites to reinvade erythrocytes. This implies that the targets of the two kinase inhibitors are essential for the formation of schizonts and possibly the reinvasion process of the parasites.

Both olomoucine and roscovitine treated cultures also formed gametocytes at 60 hrs, which is a very rare occurrence in the *P. falciparum* FCR3 strain and is shown in Figure 3.40. Gametocytes normally form when the parasites are exposed to stressful conditions so it can be assumed that the presence of the purines provided stress triggers for gametocyte formation.

Chapter 4 Discussion

4.1 Introduction

In this study two novel kinases of *P. falciparum* were expressed as recombinant proteins and were found to have kinase activity. Based upon their structural similarity to known kinases, both PfcGAK and Pfnek-3, are thought to be involved in the regulation of *Plasmodium* cell proliferation. Both the ORFs coding for the kinases were identified from raw sequence data of chromosome 12 made available at the NCBI malaria website in 2000. The first corresponds to the first 1614 bp of the predicted gene sequence, PFL2280w (PfcGAK), which codes for a 909 amino acid residue protein in PlasmoDB version 4.4. The second ORF corresponds to the gene sequence of a 347 amino acid residue protein, PFL0080c (Pfnek-3) in PlasmoDB version 4.4.

4.2 PfcGAK

4.2.1 PfcGAK gene structure

The 1614 bp ORF initially identified as coding for PfcGAK was amplified by PCR from *P. falciparum* DNA, cloned into an expression vector and the nucleic acid sequence verified. The complete protein as predicted in PlasmoDB version 4.4 is encoded by a single exon, which is located between base numbers 1959598 and 1962327. The gene codes for a 909 amino acid residue protein with a predicted molecular weight of 105.5 kDa. The ORF amplified and cloned in this study encompassed the kinase catalytic domain as well a substantial number of flanking amino acids (including the lysine rich repeat region) at the C terminus. Since the main interest was to determine kinase activity it was decided not to pursue the cloning of the complete ORF once the sequence became available.

This approach was followed by Dorin *et al.*, 2001 who only expressed the catalytic region of Pfnek-1 and Graeser *et al.*, 1997, who partially expressed PfMap1. Expression of a partial *P. falciparum* cyclin gene has also been described by Merckx *et al.*, 2003. Furthermore, the full length PfcGAK protein has a molecular weight of 105 kDa, and heterologous expression of large *P.falciparum* proteins in *E.coli* has been shown to be problematic (Mehlin *et al.*, 2006). Unfortunately the strategy of partial cloning and

expression of a gene suffers from the disadvantage of preventing the study of the effects that the remainder of the protein would have had on activity.

4.2.2 Cloning and expression of recombinant PfcGAK

Cloning of PCR products suffers from the disadvantage of introducing mutations since *Taq* polymerase does not have efficient proofreading ability. A polymerase enzyme mixture with proofreading ability was therefore used to amplify the DNA and reduce the mis-incorporation of nucleotides. Several PfcGAK clones were screened for expression and two clones, PfcGAK14 and PfcGAK24, were selected for further analysis.

Full length sequencing of inserts showed the presence of one silent mutation in PfcGAK14 as well as in PfcGAK24. PfcGAK24 also had an in frame deletion of the nanomeric repetitive peptide (81 amino acids) that appears to have caused an increase in the yield of PfcGAK protein expressed in *E. coli*. It is likely that the presence of this repetitive region in some way hindered the translation of the PfcGAK mRNA transcripts. It is also possible that the repetitive nature of the nucleotide coding sequence prevented the transcription of the region. Experiments to check the presence of full-length PfcGAK in the insoluble fraction showed negligible amounts of protein (results not shown) thus the increase in yield could not be attributed to an improvement in protein solubility.

Since pHB6 code for fusion proteins with two epitope tags, the Western blots were probed with anti-histidine as well as anti-HA antibodies. Proteins reacting with both antibodies were considered to be complete. However, there were two proteins of different sizes reacting to both antibodies, namely ± 65 kDa (PfcGAK14) and ± 58 kDa (PfcGAK24). The anti-His antibody detected several smaller protein products (Figure 3.10), initially assumed to be caused by proteolytic degradation. Since prokaryotes like *E. coli* are known to initiate protein synthesis at internal AUG sites on mRNA this may have been the cause of some of the additional bands observed in the anti-His western blot (Preibisch *et al.*, 1988). This is especially possible if there are purine rich clusters (Shine-Dalgarno (SD) sequences) present approximately between 4 and 20 nucleotides upstream from the start codon (Zubay, 1993, p845). Recently non-Shine-Dalgarno translational initiators/enhancers with low guanine content have also been identified and may contribute to secondary site initiation (Kolev *et al.*, 2003).

An analysis of the internal ATG (Methionine) sites and their preceding nucleotide sequences suggest that secondary site initiation of translation is a possibility. Table 4.1 provides a summary of the information with the methionine residue number, a predicted molecular weight of the corresponding protein as well as 24 upstream nucleotides, with purine rich clusters underlined. However, the SD consensus sequence 5'-AGGAGGU-3', was not a prominent feature of the nucleotides upstream from the potential methionine initiation sites. Since the molecular weights of some of the bands in the anti-His probed Western blot correspond to the predicted weights it is possible that some of the bands were formed as result of faulty initiation. It is thus not unlikely that the truncated protein $\pm$ 5-7.5 kDa smaller than the complete protein, was a product of initiation at either Met₅₁ or Met₇₆.

Table 4.1 Internal methionine residues of PfcGAK as potential translation start sites

<i>Met posn.</i>	<i>kDa</i>	<i>5'nucleotide sequences</i>
Met₅₁	60	<u>gagggagcgtattcatttgtttat</u>
Met₇₆	57	<u>tgtcaagataaagagaaaatagat</u>
Met₁₁₂	52	<u>gagaataattataaaaattgttatt</u>
Met₁₁₄	52	<u>aattataaaaattgttattatgttg</u>
Met₂₃₈	38	<u>atatcttggaaaattgatatatgg</u>
Met₂₈₄	32	<u>aaaagaattatttctattcttctg</u>
Met₃₃₁	27	<u>agggttaaccaaatTTTTgaaaaa</u>
Met₃₃₆	26	<u>tttgaaaaaatgaatgaacttaac</u>
M₅₃₀	0.4	<u>attattgattatacagataataaa</u>

Underlined sequence = purine rich sequences

4.2.3 PfcGAK kinase activity

Exogenous substrates are usually used to test recombinant kinases for phosphotransfer activity. In this study an α - and β - casein mixture, myelin basic protein (MBP) and histone H1 protein were used as exogenous substrates. Recombinant PfcGAK was able to phosphorylate both α - and β -casein but not MBP or histone H1 protein. It is not uncommon for recombinant kinases to phosphorylate only one exogenous substrate. Pfnk-2 was able to phosphorylate both α - and β -casein but not MBP or histone H1 protein (Reininger *et al.*, 2005). PfPK5 phosphorylated only histone H1 (Le Roch *et al.*, 2000) and Pfmrk could

phosphorylate histone H1 only in the presence of a cyclin (Waters *et al.*, 2000). In contrast, Pfnk-1 phosphorylated both α - and β -casein as well as MBP (Dorin *et al.*, 2001). PfPKB phosphorylated histone H1 as well as a synthetic peptide called 'crosstide' (Kumar *et al.*, 2004). This trend is not restricted to *P. falciparum* protein kinases.

Recombinant PfcGAK showed kinase activity regardless of the presence or absence of the nanomeric repeat region which was located outside the catalytic site. However, the apparent kinase activity of PfcGAK24 was less than that of the full length PfcGAK14 (Figure 3.17), implying that the nanomeric repeat may influence kinase activity. It may be that the repeat region plays a role in the folding of the protein and / or the phosphorylation status of PfcGAK since several potential ser/ thr phosphorylation sites are found in the repeat region (See Figure 3.29) However, the role of this repeat region is currently entirely speculative.

A kinase assay using freshly prepared PfcGAK protein showed that the protein was able to undergo autophosphorylation (Figure 3.19). However, in this assay no phosphorylation of the casein substrate took place. Kinase activity was inconsistent between different batches of recombinant PfcGAK isolates of the same clone and freshly prepared proteins did not show any kinase activity. This may be due to differences between individual protein preparations. It is also not known how the newly synthesised kinase behaves under storage. It may be that storage in glycerol aids the correct folding of the protein.

Heterologous expression of eukaryotic proteins in *E. coli* is plagued by misfolding and proteolytic degradation. Misfolding occurs despite the presence of two molecular chaperone systems (DnaK-DnaJ-GrpE and GroEL-GroES) in *E. coli* (Thomas and Baneyx, 1996; Baneyx, 1999). Misfolding causes aggregation of heterologous proteins in inclusion bodies and is partially overcome by decreasing the growth temperature during protein induction. Misfolding also targets proteins for proteolytic degradation. It appears that the chaperone proteins become limiting once overproduction of heterologous proteins occurs. (Baneyx, 1999). It is possible that the *E. coli* chaperone systems are unable to facilitate the correct folding of the *P. falciparum* proteins. Chaperone proteins have been identified in *P. falciparum* but their actions are still under investigation (Monaghan and Bell, 2005, Wiser, 2003). In a recent study by Mehlin *et al.*, 2006, only 63 of 1000 genes were expressed as soluble proteins. Only one putative ser/thr protein kinase was soluble but was not evaluated

for function. All these factors may contribute to the unpredictable variation in biological activity observed for PfcGAK and Pfnek-3.

Several *P. falciparum* protein kinases have been expressed in *E.coli*, either as GST-fusion proteins, (Dorin *et al.*, 1999, Dorin *et al.*, 2001 Kumar *et al.*, 2004, Dorin *et al.*, 2005, Reininger *et al.*, 2005) or as hexa-histidine tagged proteins (Ross-MacDonald *et al.*, 1994, Bracchi-Richard *et al.*, 2000, Deng and Baker 2002). However, Pfcrk-3 and Pfcrk-4 have remained refractory to heterologous expression (Prof. C.Doerig, personal communication).

Efficient expression of PfcGAK and its *in vitro* kinase activity makes the enzyme an attractive target for inhibitor screening. However, the factors required for maximal enzyme activity, various kinetic parameters such as the K_m value and cofactor requirements will first have to be elucidated. More importantly, the function of the enzyme in the parasite requires further investigation by making use of antibodies and gene silencing methods. It is of no use to identify an effective PfcGAK inhibitor if the enzyme is not essential for the survival of the parasite.

4.2.4 Structural analysis of the putative PfcGAK protein (PFL2280w)

The kinase catalytic domain is typical of a serine/threonine kinase. The absence of the conserved glycine triad in subdomain I is not unique to this kinase. All the other invariant residues of the kinase catalytic domain are conserved as shown in Figures 3.21 and 3.23. PfcGAK shares similar levels of similarity to both cGAK and BIKE kinases especially in the catalytic domain. Clustal analysis of the carboxyl terminal region with either cGAK or BIKE kinases showed conservation of many single residues but no specific conserved motifs were identified in either analysis. BIKE kinases are induced by bone morphogenic proteins and appear to attenuate the differentiation to osteoblasts of MC3T3-E1 cells (Kearns *et al.*, 2001) A Prosite motif search of PfcGAK did not identify any of the commonly conserved cGAK motifs like the J-domain, clathrin binding sites or PTEN regions. DNA J domains are responsible for the interaction of proteins with heat shock protein70 and are important in the postulated function of cGAK (Young *et al.*, 2003).

There are numerous potential phosphorylation sites on PfcGAK, which indicate that the enzyme may form part of a regulatory network. The potential tyrosine kinase phosphorylation site close to the amino terminus is of great interest. No kinases with clear similarity to tyrosine kinases have been identified in PlasmoDB (Ward *et al.*, 2004) although tyrosine kinase activity in *Plasmodium* has previously been reported (Sharma, 2000).

The GOLD domain is of particular interest since it is thought to be involved in protein-protein interactions. In p24 proteins the GOLD domain is a conserved globular domain and presumed to have a function related to the Golgi complex, secretion or protein sorting. GOLD domains are typically located close to amino or carboxy termini of proteins. Generally two groups of proteins containing GOLD domains are found; the p24 type in which the GOLD domain is the only domain and the other group in which additional lipid interacting domains like the pleckstrin, Sec14p or FYVE domains are present (Anantharaman and Aravind, 2002). This is not true for PfcGAK since a kinase domain is present in the protein. With the exception of the carboxyl terminal region prenylation site, no obvious lipid interaction domains or motifs were identified.

The predicted prenylation site, or Ca₁a₂X box, on the C-terminus of PfcGAK points to a possible translational modification by the attachment of a farnesyl or geranyl group to the cysteine residue. The Ca₁a₂X box amino acid is typically located at the carboxyl terminus of a protein. Farnesyl (C15) or geranyl (C10) isoprenoid groups are used to anchor proteins to membranes and are common in nuclear proteins (Alberts *et al.*, 2002). The presence of phosphorylation sites combined with a possible prenylation site indicates that the protein may be involved in signal transduction processes. However, analysis with the PrePS program (Maurer-Stroh and Eisenhaber, 2005) indicated that the C terminus of PfcGAK is not likely to be a true prenylation site. It will be important to establish the cellular location of PfcGAK in the parasite as part of a future study. Currently no further proteomic data are available.

The numerous asparagine residues contribute to fifteen potential N-glycosylation sites of PfcGAK. Glycoproteins are often secreted or associate with the plasma membrane. However, N-glycosylation of *P. falciparum* proteins has not yet been described. An analysis of the main surface proteins of *P. falciparum* shows the absence of N- or O-

glycosylation despite the presence of consensus N-glycosylation sites. Interestingly, when the C-terminal domain of the major surface protein of the merozoite (MSP-1) was expressed in a mammalian host, the consensus sites were glycosylated. The parasite also appears to be deficient in galactose, galactosamine, fucose and sialic acid, which are required for N-glycosylation. Parasite surface proteins are stabilized by glycosylphosphatidylinositol (GPI) anchors which are not identical to their mammalian counterparts. The addition of GPI anchors are thought to be the only form of protein glycosylation that takes place in *Plasmodium*. (Davidson and Gowda, 2001).

Two potential tyrosine sulphation sites were identified on PfcGAK; the first site being Tyr 666 and the second site being Tyr 793. A subtle protein modification in the late Golgi compartment is the sulphation of selected tyrosine residues of proteins. The sulphate is obtained from the sulphate donor, 3'-phosphoadenosine - 5'-phosphosulphate. Sulphated tyrosine residues are found in secreted proteins and occasionally in the extracellular domains of plasma membrane proteins. (Alberts *et al.*, 2002, p446-448).

Human cGAK shares with PfcGAK the presence of a protein kinase catalytic domain, several possible N-glycosylation sites, casein kinase and protein kinase C phosphorylation sites, and an involucrin repeat region. The role and function of GAK in higher organisms is still unclear.

It is clear that additional work will have to be carried out to determine if PfcGAK is a true homologue of human cGAK. The hydrophobicity plot, (Figure 3.28), showed several interesting features. The lysine rich nanomeric repeat region was hydrophilic with a hydrophobicity score of -20 and was followed by another stretch (± 90) of mostly hydrophilic amino acids. The second repeat region was also hydrophilic. The remainder of the protein also appears to be predominantly hydrophilic with some hydrophobic interruptions.

The eighty one residue nanomeric repetitive peptide sequence is an interesting feature of PfcGAK. This peptide sequence was deleted from clone PfcGAK24 without compromising the reading frame of the protein. Repetitive peptide sequences occur commonly in *Plasmodium* proteins and their significance is not understood. The nanomeric repeat sequence (KLQND/E KD/VSG) did not have any meaningful similarity to any other

proteins in Genbank except for some similarity to a six times repeated nanomeric repeat region (GGLQHDSSES) found in the tissue factor pathway inhibitor of *Canis familiaris* (Figure 3.27). The presence of a number of potential phosphorylation sites in this region indicates that it may have a regulatory role. A Blastp search with the sequence KLQND/EKD/VSG showed that the motif KLQND occurs commonly in membrane associated proteins of prokaryotes.

4.3 Pfnek-3

4.3.1 Pfnek-3 gene structure

The 1041 bp ORF initially identified coding for PfNek-3 (PFL0080c) was amplified by PCR, cloned into an expression vector and the nucleic acid sequence verified. The protein is encoded by a single exon as predicted by the gene prediction algorithms in PlasmoDB version 4.4.

4.3.2 Expression of recombinant Pfnek-3

The expression of recombinant Pfnek-3 was problematic with very low levels of expression observed in all the clones evaluated. It may be that the hydrophobic nature of the N-terminus of the protein causes the hexa-histidine tag of the recombinant protein to be buried and unavailable for binding to the nickel matrix. However, isolation of Pfnek-3 under denaturing conditions did not improve the protein yield.

4.3.3 Pfnek-3 kinase activity

Although expression levels of Pfnek-3 were very low, one of the clones, Pfnek-3-N2(3), showed kinase activity against casein as shown in Figure 3.20. A Pfnek-3 negative isolate from the vector control culture did not show any activity against casein. No phospho-transfer activity was observed when MBP and H1 protein were used as substrates. Further work on Pfnek-3 is required to optimise the expression and verify the kinase activity of the protein. The various kinetic parameters such as K_m value and cofactor requirements, pH and temperature optima need to be established.

4.3.4 Structural analysis of the putative Pfnek-3 protein (PFL0080c)

P. falciparum Pfnek-3 has the highest similarity to the Nek9 enzymes of human, *Xenopus* and mouse origin with 30% identical amino acids and between 55 and 58 % of amino acids falling into the same class. The catalytic site is well conserved with all the invariant amino acid residues present except for the glycine triad in subdomain I. The alignment is shown in Figure 3.30.

Unlike the other Nek1 and Nek9 kinases that display the motif HRDIK in the catalytic domain VIB, Pfnek-3 has the sequence HGDLY. The common DFG motif in subdomain VII is conserved in Pfnek-3 but not for the other Nek9 kinases, which displays the motif DYG. It is not possible to predict the effects of these deviations with information currently available.

The Prosite motif search showed that the Pfnek-3 is likely to undergo phosphorylation by protein kinase C or casein kinase, with seven potential phosphorylation sites identified. However these predictions can be erroneous especially for short motifs such as PKC sites and may have occurred by chance. Pfnek-3 has two potential FXXT phosphorylation sites, which are essential for the activation of the NIMA related kinases (Pu *et al.*, 1992). Phosphorylation of the threonine residue is essential for the activation of the enzymes (Pu *et al.*, 1992). The first (FKKT) occurs at the C terminus end of subdomain VIII and the second (FCAT) at the C terminus of subdomain IX. Based on the Clustal alignment in Figure 3.30 it appears that the second site is more likely to be the true phosphorylation site since the first phenylalanine, the third alanine and the fourth threonine residues are conserved in the Pfnek-1 and Nek9 enzymes. However, the first site is more closely related to the consensus sequence F**K**/R**K**/R**T** of the NIMA related kinases.

Two potential N-glycosylation sites were identified as well as a potential membrane lipoprotein lipid attachment site positioned between the N-terminus and the catalytic domain. The sequence is also preceded by a stretch of amino acids which are rich in hydrophobic residues like Phe, Ile and Val. This amino acid region shows the presence of a signal sequence between residues 1-19 and a transmembrane motif between residues 21 and 43.

Pfnek-3 has in addition the newly identified PEXEL (*Plasmodium* export element) consensus sequence (RxLxQ/E/D) occurs between residues 22 and 26. PEXEL or vacuolar transport signals (VTS) are found on *Plasmodium* sequences targeted for export to the erythrocyte. (Marti *et al.*, 2004, Hiller *et al.*, 2004). Marti *et al.*, (2004), predicted the RxLxE/Q sequence, position 1 occupied by a positively charged hydrophilic amino acid (R or K), a hydrophobic amino acid in position 3 (L or I) and a less conserved amino acid in position 5 (Q/E/D) with uncharged amino acids in positions 2 and 4. The Pfnek-3 motif is RILCN. The C terminal asparagine residue falls in the same class of amino acids as glutamine. Hiller *et al.*, 2004, proposed the motif RxSRILAExxx to be the primary vacuolar targeting signal of the *Plasmodium* secretome. Pfnek-3 has the sequence **RMRLCNLLI**. Both groups proved that the PEXEL motif is required for the transport of both soluble and membrane proteins from the parasite to the erythrocyte and its membrane. The presence of this sequence in Pfnek-3 suggests that the protein may be exported to the parasitophorous vacuole (PV) or the erythrocyte.

Many NIMA related kinases have PEST sequences between the catalytic and carboxyl terminus. PEST sequences target proteins for ubiquitin mediated proteolysis. Pfnek-3 does not have any PEST sequences as determined by the PESTfind program thus indicating that the protein is not likely to have a short half-life as expected with proteins containing PEST sequences.

Plasmodium has a unusual class of protein kinases known as the FIKK kinases (Ward *et al.*, 2004), members of which were recently shown to have a PEXEL sequence (Schneider *et al.*, 2005). The *P. falciparum* trophozoite protein R45 is one of the FIKKs. Several other FIKK kinases were discovered in other protozoal parasites of the phylum *Apicomplexa*. Schneider *et al.*, 2005, refers to this class of kinases as the R45-FIKK family. In this group of kinases the catalytic domain is located at the carboxy terminus, while the N-terminus is poorly conserved. All the genes in this kinase group have a three exon structure and encode a protein with a signal or anchor sequence preceding a PEXEL motif. Pfnek-3, by contrast, is encoded for by a single exon and consists of a signal sequence, PEXEL motif, putative transmembrane region and a kinase domain. It would be of great interest to determine the subcellular location of the native Pfnek-3 protein.

The *Plasmodium* parasite resides in the parasitophorous vacuole throughout its intracellular life cycle. The parasite must have a transport system within the erythrocyte cytoplasm to

mediate the uptake of nutrients from the host blood stream and for the display of parasite derived proteins on the erythrocyte cell membrane. The parasite has an extensive protein trafficking network that consists of an endoplasmic reticulum that is continuous with the nuclear membrane, and a rudimentary Golgi complex. (Cooke *et al.*, 2004). Apparent homologues of proteins involved in endosome formation like clathrin, adaptin and dynamin have also been identified in PlasmoDB. Transport to and from the Golgi complex is mediated by vesicles.

Transport across the PV is thought to be mediated by single or mixed cargo vesicles. Proteins destined for secretion in the PV appear to have the classical hydrophobic N-terminal sequence, followed by a PEXEL motif. It appears that all proteins destined for the erythrocyte is secreted into the PV although there is evidence for the existence of subcompartments in the PV. Proteins destined for the erythrocyte cytosol and its membrane have been found to have longer hydrophobic signal sequences and may also have recessed signals and appear to accumulate in Maurer's clefts. (Cooke *et al.*, 2004). A number of known secretory proteins lack all the obvious signal sequences and are thought to be transported to the erythrocyte by vesicles. A study with antibodies targeted against Pfnk-3 or *in vivo* tagging will give information regarding the subcellular location of Pfnk-3 and clues as to its possible function.

The hydrophobicity plot of Pfnk-3 supports the theory of a predicted signal sequence followed by a transmembrane domain since the hydrophobicity index is the highest between amino acid 1-18 and 24 to 32 as seen in Figure 3.33. The region between amino acid 40 and 60 is also hydrophobic and supports the prediction of a transmembrane sequence. The remainder of the protein displays hydrophilic regions interspersed with hydrophobic regions as is usual for globular proteins.

Apart from the catalytic domain, Pfnk-3 does not appear to share any of the other conserved features of Nek-9 such as the central domain with similarity to RCC1 and a C terminal region coiled coil domain. Nercc1/Nek8/Nek9 prefer β -casein as exogenous substrate and undergo autophosphorylation. Similar to Nek9 *Plasmodium* Pfnk-3 also preferred casein as an exogenous substrate but it is not known if it undergoes autophosphorylation.

4.4 Identification of cell cycle regulatory kinases using a homology based approach

Several model organisms have been used to study the cell cycle including yeast, *Drosophila*, *Xenopus*, starfish oocytes and mammalian cells. It is generally assumed that the molecules involved in the regulation of the cell cycle are conserved throughout the eukaryotic kingdom making it possible to use a structural similarity based approach to identify these molecules in various organisms. (This has been loosely called a homology based approach) In this context numerous proteins, including cell cycle kinases, have been identified and characterized.

However, the model systems referred to above all belong to the same phylogenetic lineage, the Opisthokonta which is very distant from the unicellular malaria parasite which belongs to the phylum Apicomplexa. (Doerig, 2004). It is therefore likely that homology based approaches may miss proteins specific to the cell cycle of the malaria parasite. It is also likely that an incorrect function may be assigned to a specific protein based on its similarity to a well characterised protein of the model system. PfPKB may be an example of this; the protein is similar to mammalian PKB in the kinase catalytic region but lacks the pleckstrin homology domain, which is essential for the function of the enzyme in the AKT/PI3-K signal transduction pathway. It is therefore important to be cautious when functions are assigned to proteins based purely on their structural similarity.

Predicted functions have been assigned to both PfcGAK and Pfnek-3 based only on their respective similarities to the known proteins. This similarity lies mainly in the kinase catalytic domain. Outside the catalytic domain both enzymes are poorly conserved compared to cGAK, Nek1 or to Nek9. It is clear that extensive work is still required to identify the proper substrates of both enzymes and their possible role in the *Plasmodium* life cycle.

4.5 *Plasmodium falciparum* kinases as potential drug targets

The protein kinases of *P. falciparum* have unusual structural properties when compared to those of higher organisms, which may render them good parasite selective drug targets. For example the malarial CDKs and MAPKs have large C-terminal or N-terminal extensions.

Some malarial parasite kinases have large insertions in their catalytic domain (Doerig, 2004).

The *Plasmodial* cGMP - dependent kinases have three instead of the two characteristic cGMP binding sites found in higher eukaryotes. *Plasmodium* also has a unusual class of kinases called the CDPK kinases. CDPKs are found in plants and Alveolates (which includes both the ciliates and the Apicomplexa). The FIKK group of kinases are found only in Apicomplexa and may represent an ideal parasite selective drug target (Doerig, 2004).

The phosphorylation of tyrosine residues has been observed in *P. falciparum* (Sharma *et al.*, 2000) but the genome database has not revealed any proteins with similarity to the classical tyrosine kinases found in higher organisms (Doerig 2004). Is it possible that *P.falciparum* tyrosine phosphorylation is carried out by the so called “dual” – specificity kinases, which belong to the ser/thr group but can also phosphorylate tyrosine. An investigation into the mechanism of tyrosine phosphorylation may reveal unusual drug targets.

Although a number of putative signal transduction kinases have been identified in *P. falciparum*, subsequent experimental data and sequence analysis have supplied evidence that question the correctness of their originally assigned functions. An example is the kinases related to MAPKs. PfPK6 has sequence features found both in MAPKs and CDKs but displays kinase activity independent of cyclin. Furthermore Pfmap-2 has an atypical activation site and there is no MAPKK homologue predicted in the PlasmoDB. Nevertheless Pfnek-1, a NIMA related kinase, has an activation site that is similar to MAPKKs and is able to phosphorylate Pfmap-2 (but no other MAPK) *in vitro*. It is not known if this is a feature of kinase substrate promiscuity or if Pfnek-1 is the true activator of Pfmap-2. PfPK7 is most closely related to MAPKKs but has no residue able to undergo phosphorylation at the activation position. The available evidence suggests that the typical MAPKKK-MAPKK-MAPK module that exists in higher organisms does not exist in *P. falciparum* (Doerig, 2004).

Homologues of the cell cycle machinery identified in *P. falciparum* presents the same dilemma. Although the overall similarity does exist there is no clear “homology” in terms of function. The CDK related kinases identified in parasites do not share obvious homologies to those in higher organisms. For instance there is no definite homologue to the

CDK4/6 family. PfPK5 shares similarity with both CDK1 and CDK5. CDK5 does not function directly as a cell cycle regulator but is involved in neuronal maintenance (Maccioni *et al.*, 2001). It is also surprising the PfPK5 is promiscuous in terms of activating cyclins since it is activated by both human cyclin H or Pfcyc-1, which is homologous to human cyclin H. Cyclin H activates CDK7, which is a CDK activating kinase that functions in DNA repair and participates in RNA transcription, in human cells. Furthermore PfPK5 activity can also be stimulated by other cyclins and cyclin related proteins such as human cyclin A (Holton *et al.*, 2003) and p25, the cyclin like activator of CDK5 in neuronal cells. Hence it appears that a prediction of the precise function of proteins such as kinases and cyclins cannot be predicted from sequence similarity alone.

4.6 Inhibitory effects and structure activity relationships of purine derived kinase inhibitors on the FCR-3 strain of *P. falciparum*

The IC₅₀ values obtained by the hypoxanthine incorporation assay for *P. falciparum* were compared to the IC₅₀ values obtained by screening the compounds in the starfish oocyte *in vitro* model of the CDK1/cyclin B system developed by Rialet and Meijer (1991) in Roscoff, France. IC₅₀ values of less than one µM for the hypoxanthine assay are generally considered to be promising. The IC₅₀ values for chloroquine are less than 10 nM for a single and double replicative cycle and for quinine they are less than 50 nM for a single and double replicative cycle.

In contrast, when compounds are tested for activity against cancer cells, IC₅₀ values in the 10 – 20 µM range are considered to be potent since the treatment strategies for cancer patients are vastly different to any other therapeutic regimens used in infectious diseases. Cancer chemotherapeutic regimes use drug dosages that are orders of magnitude higher than those used in conventional drug therapies to achieve a total cell kill effect. This unfortunately contributes to the poor adverse effect profile and the high cost of chemotherapy.

The starfish oocyte data for the CDK1/cyclin B assay were kindly made available by Dr. L Meijer in the context of a joint project. Table 4.2 provides a direct comparison of *P. falciparum* IC₅₀ values and that of the starfish oocyte CDK1/cyclinB complex.

Table 4.2 Comparison of IC₅₀ values of purine inhibitors tested on *P. falciparum* with those obtained from testing on the starfish oocyte CDK1/cyclin B complex.

Compound	IC₅₀ values: <i>P. falciparum</i> proliferation (μM)	IC₅₀ values: CDK1/cyclin B activity (μM)
iso-olomoucine	235 ± 25.0	> 500
olomoucine	8.45 ± 1.15	7
roscovitine	5.36 ± 1.61	0.45
isopentenyl-adenine	8.34 ± 2.16	55
Purvanolol A (R)	0.55 ± 0.15	0.03
Purvalonol A (S)	0.42 ± 0.06	0.5
Purvalonol B (95)	7.07 ± 0.69	0.02
Methyl-purvalonol (95M)	47.45 ± 4.03	inactive
Amino-purvalonol (97)	5.85 ± 0.84	0.03
Methyl amino purvanolol (97M)	11.04 ± 0.92	inactive
51	0.56 ± 0.08	0.42
52	7.1 ± 0.96	0.22
52M	32.32 ± 2.4	inactive
59	0.53 ± 0.11	0.8
66	No inhibition at 10 μM	0.4
40	0.5 ± 0.19	5.0
43	1.12 ± 0.13	4.3
99	0.83 ± 0.23	25
101	0.63 ± 0.18	42

With the exception of isopentenyl-adenine and compound 101, all the compounds tested were 2,6,9-tri-substituted purine derivatives with an amino group at C2 and a benzylamino

or anilino group at C6. To simplify the structure activity comparisons, the structures of the compounds are presented again.

4.6.1 Antimalarial activity of iso-olomoucine, olomoucine, roscovitine and isopentenyladenine

Olomoucine, roscovitine and isopentenyladenine (Figure 4.1), displayed similar activity against *P. falciparum* with IC_{50} values of 8.45, 5.36, and 8.34 μ M, respectively. This is in sharp contrast with the effects of these compounds on an isolated CDK1/cyclin B system where sensitivities to the three inhibitors were very different, with isopentenyladenine being the least effective (IC_{50} :55 μ M), and roscovitine the most effective (IC_{50} :0.45 μ M). Olomoucine had a IC_{50} value of 7.0 μ M

Iso-olomoucine, an isomer that does not inhibit protein kinases, showed minimal activity against *P. falciparum*, with an IC_{50} value of 235 μ M that was consistent with the lack of inhibition on the CDK1/cyclin B system. Roscovitine has bulkier groups on C2 and N9 that may contribute to its submicromolar effect on the isolated enzyme system (Figure 4.1). In contrast, these structural differences between roscovitine, olomoucine and isopentenyladenine did not appear to influence their antimalarial activity.

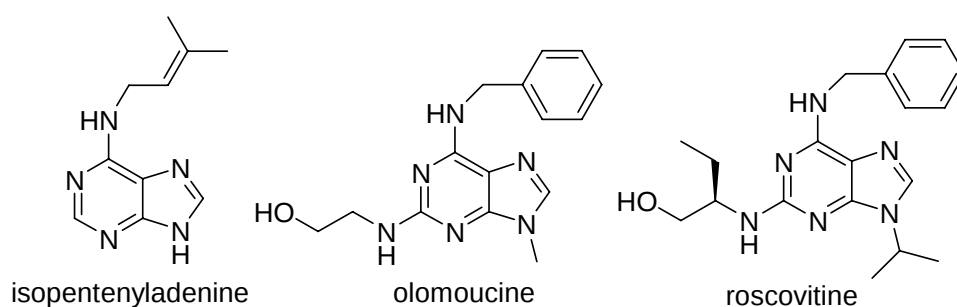


Figure 4.1 The chemical structures of isopentenyladenine, olomoucine and roscovitine.

The selectivity of roscovitine for cyclin dependent kinases above other protein kinases has recently been confirmed (Bach *et al.*, 2005). However, the same study also showed that roscovitine can bind to the enzyme pyridoxal kinase. This enzyme is a non-protein kinase and is responsible for the activation of vitamin B6 by phosphorylation. This finding has

implications for the interpretation of the data obtained from the ongoing clinical trials of roscovitine. (Bach *et al.*, 2005).

4.6.2 Antimalarial activity of the purvalanol series of compounds

The purvalanol series of compounds showed a wide range of inhibitory activities against *P. falciparum*. Molecules of this family are characterized by the presence of a chiral valinol substituent at C2 (Gray *et al.*, 1999). Within this family, the *R* and *S* isomers of purvalanol A (Figure 4.2), displayed similar submicromolar levels of inhibitory activity against *P. falciparum* proliferation, with IC₅₀ values of 0.55 and 0.42 µM, respectively (Table 4.2). In contrast, there was a 15-fold difference in activity between the two isomers on purified CDK1/cyclin B, with the *R* isomer being the most active with an IC₅₀ value of 0.03 µM compared to 0.5 µM of the *S* isomer. It is also important to note that the IC₅₀ value of purvalanol A (*R*) on recombinant PfPK5.p25 activity is 8 µM (Le Roch *et al.*, 2000) which indicates that this enzyme is probably not the molecular target of purvalanol A (*R*) in *P. falciparum*.

Purvalanol B contains a carboxyl group on the 3-chloroanilino C6 substituent of purvalanol A (Figure 4.2) which makes it a very potent inhibitor when tested on the isolated CDK1/cyclin B enzyme system with an IC₅₀ value of 0.02 µM. Although it is a potent CDK inhibitor, it had an average effect on *P. falciparum* with an IC₅₀ value of 7.0 µM. It is nevertheless noteworthy that this compound has a minimal inhibitory effect on a large number of human cancer cell lines (IC₅₀ > 100 µM), presumed to be the result of the hydrophilicity conferred by the carboxyl group and hence shows some specificity towards *P. falciparum* (Nathaneal Gray, unpublished observations).

However, the presence of a carboxyl group does not prevent a compound from crossing membranes since there will be equilibrium between the uncharged and charged species of the carboxyl group, ($\text{RCOOH} \leftrightarrow \text{RCOO}^- + \text{H}^+$). The undissociated RCOOH is able to cross cell membranes allowing the intracellular accumulation of the compound. Furthermore, purines are also transported across membranes into *P. falciparum* by specific transport mechanisms (Ginsburg and Stein, 2005).

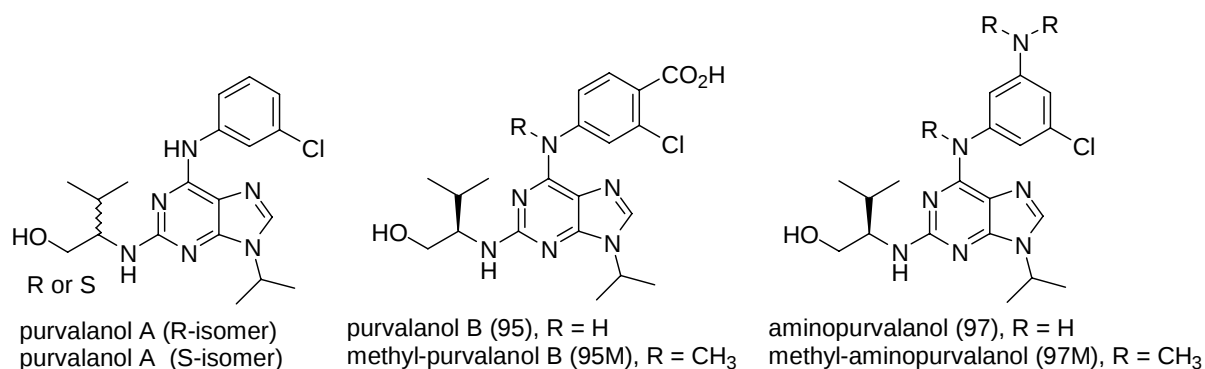


Figure 4.2 The chemical structures of the purvalanol group of compounds.

The methylated derivative of purvalanol B, which was inactive in the CDK1/cyclin B kinase assay, was also less active against *P. falciparum* with an IC₅₀ value of 47.5 μ M. An amino substituent on the chloroanilino group of purvalanol A (forming compound 97) did not significantly increase the activity against *P. falciparum* compared to the carboxylated derivative (compound 95), decreasing the IC₅₀ value to 5.8 μ M which was in contrast with its high activity against the isolated CDK1/cyclin B system (IC₅₀:0.03 μ M). Tri-methylation of 97 (97M) abolishes the strong inhibitory effect against isolated CDK1/cyclin B, and also decreased its antimalarial activity (IC₅₀:11.04 μ M). It appears that substitutions of the chloroanilino substituent of purvalanol with either amino, aminomethyl or carboxyl groups had a detrimental effect on the antimalarial activity of these compounds.

4.6.3 Selective inhibitory activity of compounds 99 and 101 on *P. falciparum*

The two purine derivatives, compounds 99 and 101, share common substitutions on C2 (3-trifluoromethyl-benzylamino) and C6 (4-aminobenzylamino), as seen in Figure 4.3. These compounds displayed poor activity against purified CDK1/cyclin B, with IC₅₀ values of 25 and 42 μ M, respectively, while they were active against *P. falciparum* at sub-micromolar concentrations, with IC₅₀ values of 0.83 and 0.63 μ M, respectively (Table 4.2). The *P. falciparum* inhibitory data do not allow the activity of these compounds to be attributed to either one of the two substitutions on the C2 or C6 groups. However, considering the structure-activity profiles of the compounds discussed above (the olomoucine and purvalanol families), it seems likely that the increased activity is related to the C2 trifluoromethyl-benzylamino substituent.

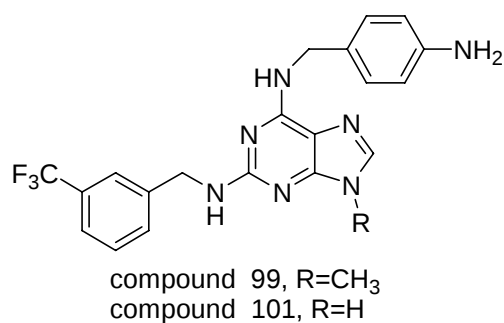


Figure 4.3 The chemical structures of the compound 99 and 101

4.6.4 Antimalarial activity of R and S phenyl substitutions on the C2-hydroxyethyl-amino side chain

Compounds 59 and 66 are enantiomers with bulky phenyl substitutions on the chiral carbon of the C2 hydroxyethylamino substituent (Figure 4.4). Compound 59 was active against *P. falciparum* with an average IC₅₀ value of 0.53 μ M, whereas compound 66 showed no activity against the parasite at the highest concentration tested (10 μ M) (Table 4.2). This was the highest concentration available for testing.

Both compounds showed sub-micromolar activity against the CDK1/cyclin B complex with compound 66 being more active with an IC₅₀ value of 0.4 versus 0.8 μ M for compound 59. This indicates that the spatial orientation of the phenyl ring has a major effect on the antimalarial activity but not against the isolated enzyme, and provides evidence that the parasite target is probably not identical in its three dimensional structure to the starfish CDK1/cyclin B enzyme system. This was in contrast with the effect observed for the *R* and *S* isomers of purvalanol A, where the substitution on the same chiral carbon is the aliphatic isopropyl group of the amino acid valine, which displayed similar antimalarial activities in the sub-micromolar range.

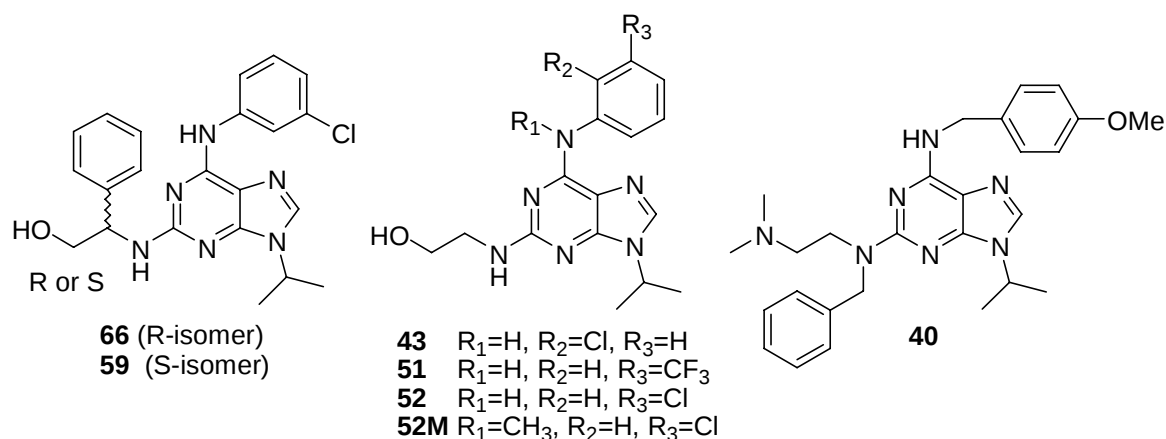


Figure 4.4 The structures of various purine derivatives.

4.6.5 The effect of halogen substitutions on C6 on antimalarial activity

Compounds 43, 51 and 52 are structurally similar, differing only in the substitutions on C6 (Figure 4.4). Nevertheless, we observed a considerable variation in the IC_{50} values of these compounds on *P. falciparum* (IC_{50} values of 0.56 and 7.10 μM for compound 51 and 52, respectively). This was in contrast with the submicromolar IC_{50} values of 0.42 and 0.22 μM for 51 and 52, respectively, for these compounds on the CDK1/cyclin B system. Compound 43 had an IC_{50} value of 1.12 μM when tested against *P. falciparum* which was approximately four times lower than that of the CDK1/cyclin B system of 4.3 μM . This was higher than that of compound 51 but lower than that of its isomer, compound 52 (IC_{50} : 7.1 μM). Methylation of compound 52 reduced the antimalarial activity to an IC_{50} value of 32.3 μM , which was consistent with its lack of activity against the CDK1/cyclin B enzyme system. This indicates that compound 51 with a CF_3 group on position R_3 is a good candidate lead compound.

4.6.6 Antimalarial activity of compound 40

This compound has a unusual structure with bulky substitutions on the C2 and C6 rings (Figure 4.4). It had a sub-micromolar IC_{50} value of 0.54 μM which compared well with that of the other purines tested. The IC_{50} value of this compound against *P. falciparum* was ten fold lower than its IC_{50} value against the CDK1/cyclin B enzyme system of 5 μM . Interestingly this compound also has antifungal activity (N. Gray unpublished observations, personal communication).

4.6.7 The effect of purine inhibitors at an IC₉₀ concentration on the erythrocytic stages of the parasite

The effects of roscovitine and olomoucine on the progression of the parasites through their normal erythrocytic cycle were assessed at a concentration equivalent to the IC₉₀. These two compounds were the only ones available in sufficient quantities for this assessment. At an IC₉₀ both compounds caused inhibition of the parasite growth as can be seen in Figures 3.38 and 3.39, which was consistent with their IC₅₀ values.

The roscovitine and olomoucine treated cultures (Figure 3.38) showed a decreased total parasitaemia compared to the control culture. The parasites appeared to progress normally to the trophozoite stage but the formation of schizonts and the release of merozoites were suppressed since rings were not observed. Interestingly, the purine treated cultures revealed the presence of gametocytes, (Figure 3.40) which was a very rare occurrence in the *P. falciparum* FCR-3 strain. Gametocytes form in culture under stressful conditions and it may be that olomoucine and roscovitine provided the stress triggers for gametocyte formation. It may be that the inhibitors cause inhibition of an unknown kinase that is involved in the regulation of gametogenesis.

4.6.8 Concluding remarks regarding the purine derived inhibitors

Some purine-derived inhibitors were able to interfere with parasite growth. The effect of these agents was measured against survival of the parasite. Although the precise molecular targets for these inhibitors are not known it is important that they do inhibit the proliferation of the parasite, which is the ultimate clinical outcome expected from a potential therapeutic agent. The structure activity relationships derived from the data collected here can provide useful information for the generation of parasite selective inhibitory compounds.

It is also important not to make assumptions regarding the specificity of the compounds tested. Immobilized purvalonol B was used to identify the target in *P. falciparum*. A 32 kDa band was identified to be *P. falciparum* casein kinase 1. Purvalonol B was able to inhibit this enzyme with an IC₅₀ of around 200 nM (Doerig, 2004) but could only inhibit the parasite at an IC₅₀ of 7 µM. The discrepancy between the IC₅₀ of the whole parasite and the isolated drug target is important. This indicates that caution must be exercised in

applying data obtained from high throughput screening of inhibitors against recombinant enzymes. The efficacy of a potential drug must be evaluated against a whole organism and not isolated enzymes before proceeding with pre-clinical toxicity studies. It is of no use to efficiently inhibit a specific enzyme and not kill the parasite at a host-friendly inhibitor concentration.

Purvalonol B also binds casein kinase homologues in *Leishmania* and *Toxoplasma*. In extracts from Metazoans (mammalian cells, *Xenopus* and starfish oocytes) purvalonol B bound kinase enzymes from the CMG group, which includes CDKs and MAPKs (Knockaert *et al.*, 2002). Caution must be exercised in assigning a potential compound to a specific class of kinase inhibitor without it being tested against a wide variety of enzymes. It is possible that the compound may inhibit crucial host enzymes and cause serious adverse effects. A sensible approach is to do immobilized inhibitor ligand binding studies to identify potential critical ligands for any new molecule as described by Godl *et al.*, (2003) and Knockaert *et al.*, (2002).

Chapter 5 Summary

5.1 PfcGAK

5.1.1 PfcGAK expression

In this study the prediction that the putative gene PFL2280W codes for a kinase was confirmed. The nucleic acid sequence coding for part of the putative PfcGAK gene was amplified and cloned into an expression vector and the sequence confirmed. In some clones the expression construct lost the nanomeric repetitive region. It is presumed that this coding region was not compatible with the normal functioning of *E. coli* since a clone, PfcGAK24, which was deficient in this repeat region, expressed protein to a much higher level than the clones containing the nanomeric repetitive region.

The recombinant protein showed kinase activity against casein but not against myelin basic protein and histone H1 protein. Freshly prepared protein did not show any kinase activity but was shown to undergo autophosphorylation. It is possible that site-specific phosphorylation is required for maximal activity.

5.1.2 PfcGAK protein structure analysis

The identity of gene PFL2280W as a homologue of PfcGAK was assigned based on the similarity of the kinase catalytic domain to characterised and putative cGAK proteins. The kinase catalytic domain is well conserved with the only major anomaly being the absence of the common glycine triad in subdomain I. An analysis of the complete predicted protein could not identify any of the other domains found in cGAK proteins like the J-domain, a PTEN site or clathrin binding sites.

Clustal analysis showed that PfcGAK has a similar percentage identity to both cGAK and to BIKE kinases, which are involved in the inhibition of osteoclast differentiation. It would thus not be unreasonable to conclude that PfcGAK may not be a true homologue of human cGAK. Unfortunately this study did not provide experimental evidence as to the *in vivo* function of PfcGAK but has provided worthwhile clues to follow in subsequent studies.

Analysis of the primary amino acid sequence revealed several interesting features like the presence of a unique nanomeric repetitive sequence and a number of motifs. The numerous potential N-glycosylation sites would normally indicate that the protein is secreted or associated with the outer membrane of an organism. However, since protein glycosylation is not common in *P. falciparum* it is not known if glycosylation sites are accurate predictors of protein localization in the *Plasmodium* proteome. There are no strong indicators of the presence of signal sequences or transmembrane regions, although they may be hidden and only be revealed once the protein is folded properly. The protein has several potential phosphorylation sites, including a tyrosine phosphorylation site, which indicates that it may be regulated by phosphorylation reactions. There are no predicted classical tyrosine kinases in PlasmoDB. The potential carboxyl terminus prenylation site indicates that the protein is most likely anchored via a farnesyl or geranyl group to a membrane. However this needs to be confirmed experimentally.

PfcGAK does not have any of the catalytic domain motifs that would allow its classification to any of the kinase subclasses as described by Hanks, 2003. Ward *et al.*, 2004 grouped this kinase in the ‘other’ group. Analysis of the protein structure performed in this study confirms this classification.

Future studies on PfcGAK

A future study making use of PfcGAK antibodies can provide information regarding the cellular location of PfcGAK and confirm the accuracy of the prenylation and glycosylation predictions. The function of the nanomeric repetitive region can be explored using the two clones obtained in this study to carry out protein-protein interaction studies.

5.2 Pfnek-3

The coding region of Pfnek-3 (PFL0080c) was amplified successfully and the nucleic acid sequence in PlasmoDB version 4.4 was confirmed. Expression of this clone was problematic despite the use of the BL21 Codon Plus strain of *E. coli*. Only one of the clones displayed kinase activity despite the fact that the open reading frame was inserted correctly. Pfnek-3 appeared to have kinase activity against casein but not histone 1 or myelin basic protein. Further work is required on this protein to improve its expression and optimise the kinase activity.

Pfnek-3 is a protein of 41 kDa and consists mainly of a putative signal sequence, a PEXEL motif followed by a putative transmembrane region and a kinase catalytic domain. All the critical residues of the kinase catalytic domain are conserved with the exception of the common glycine triad in subdomain I. Of particular interest is the presence of the PEXEL motif, which is found in *Plasmodium* proteins exported to the parasitophorous vesicle and erythrocyte cytoplasm and membrane. Most of these proteins have a signal sequence coded for by a separate exon but this is not the case for Pfnek-3, which is encoded by a single exon. Pfnek-3 specific antibodies will be useful in determining the subcellular location of Pfnek-3.

The Nek kinases of higher organisms are classified in the “other” group of kinases (Hanks, 2003). Likewise, the Nek kinases of *P. falciparum* form a separate cluster of kinases with structural features that do not allow them to be grouped with any of the major kinase groups (Ward *et al.*, 2004).

5.3 Purine derived inhibitors

Some purine-derived inhibitors were able to interfere with parasite growth. This is important since the effect of these agents was measured against survival of the parasite. Although the precise molecular targets for these inhibitors are not known it is important that some of them do inhibit the proliferation of the parasite, which is the ultimate clinical effect desired from a new compound. The structure activity relationships derived from them provided useful lead compounds for design of parasite specific compounds. This is true especially for the enantiomers 66 and 59 where the orientation of a phenyl group changed the antimalarial activity of the compound dramatically.

An IC₅₀ value above 1000 nM is not considered to be of great significance in the treatment of infectious diseases since the dose required to obtain a meaningful clinical effect would be too high. This in turn will cause other problems like unacceptable toxicity, problems with administration and is also likely not to be cost effective. Six compounds had IC₅₀ values of less than 1000 nM and warrant further evaluation. They include purvalonol A and compounds 51, 59, 40, 99 and 101. Unfortunately their IC₅₀ values are above the range

normally considered to be effective in malaria. However, a favourable toxicity profile may still render these compounds useful alternatives in the treatment of malaria.

Although drug screening is a relatively simple process it does provide a useful tool in the fight against malaria. It is important to realize that despite the remarkable progress made on understanding the basic molecular processes of the parasite a simple screening test will give a reliable indication with regards to drug efficacy against the parasites. In the clinical situation, the patient survives when the parasite is killed, regardless of the mechanism of killing exercised by the drug. The most important factors are clinical efficacy with minimal toxicity to the host and delayed development of drug resistance.

5.4 Conclusion

Cloning and expression of two putative *P. falciparum* kinases confirmed that the genes indeed code for kinases. The study raised intriguing questions regarding the function and structural features of the two kinases. The data obtained from the study were not sufficient to draw any conclusions regarding their specific roles in the *Plasmodium* cell cycle. When the available information regarding the regulation of the *Plasmodium* cell cycle is compared to that of mammalian cells it is clear that the study of the *Plasmodium* cell cycle regulators is still in its infancy. The *Plasmodium* genome database is an invaluable resource for parasitologists and is providing a useful tool to unravel the regulatory components of the cell cycle and provide clues to interfere successfully with the erythrocytic and sexual cycles. Screening libraries of compounds for activity still plays a major role in the discovery of new anti-malarial agents despite the huge progress made in understanding the basic molecular processes of the parasite and the value of the relatively simplistic screening method must not be underestimated.

References

- Abraham RT. (2001) Cell cycle checkpoint signalling through the ATM and ATR kinases *Genes Dev.* 15:2177-2196
- Ahn JS, Radhakrishnan ML, Mapelli M, Choi S, Tidor B, Cuny GD, Musacchio A, Yeh LA, Kosik KS. (2005) Defining CDK5 ligand chemical space with small molecule inhibitors of Tau phosphorylation. *Chem. Biol.* 12(7): 811-823
- Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P. (2002) The molecular biology of the cell, 4th edition Garland Science, New York.
- Al-Khedery B, Barnwell JW, Galinski MR. (1999) Stage specific expression of 14-3-3 in asexual blood stage *Plasmodium*. *Mol. Biochem. Parasitol.* 102:117-130
- Andrews PD, Knatko E, Moore WJ, Swedlow JR. (2003) Mitotic mechanics: the auroras come into view. *Curr. Opin. Cell. Biol.* 15:672-683
- Anantharaman V. Aravind L. (2002) The GOLD domain, a novel protein module involved in Golgi function and secretion. *Genome Biol.* 3:1-7
- Arnot, DE, Gull K.(1998) The *Plasmodium* cell cycle: facts and questions. *Ann. Trop. Med. Parasitol.* 92(4): 361-5
- Bach S, Knockaert M, Reinhardt J, Lozach O, Schmitt S, Baratte B, Koken M, Coburn SP, Tang L, Jiang T, Liang D, Galons H, Dierick J, Pinna L, Meggio F, Totzke F, Schachtele C, Lerman A, Carnero A, Wan, Y, Gray N, Meijer L. (2005) Roscovitine targets, protein kinases and pyridoxal kinase. *J. Biol. Chem.* 280: 31208-31219
- Barik S, Taylor RE, Chakrabarti D. (1997) Identification, cloning and mutational analysis of the casein kinase 1 cDNA of the malaria parasite, *Plasmodium falciparum*. Stage specific expression of the gene. *J. Biol. Chem.* 272: 26132-26138
- Banyex F. (1999) Recombinant protein expression in *Escherichia coli*. *Curr. Opin. Biotech.* 10:411-421.
- Belham C, Comb M J, Avruch J. (2001) Identification of the NIMA family kinases NEK6/7 as regulators of the p70 ribosomal S6 kinase. *Curr. Biol.* 11:1155-1167
- Belham C, Roig J, Caldwell JA, Aoyama Y, Kemp BE, Comb M, Avruch J. (2003) A mitotic cascade of NIMA family kinases. Ncc1/Nek9 activates the Nek6 and Nek7 kinases. *J. Biol. Chem.* 278:34897-909.
- Bendtsen JD, Nielsen H, Von Heijne G, Brunak S. (2004) Improved prediction of signal peptides: SignalP 3.0. *J. Mol. Biol.* 340: 783-795
- Billker O, Dechamps S, Tewari R, Wenig G, Franke-Fayard B, Brinkman V. (2004) Calcium and calcium-dependent protein kinase regulate gamete formation and mosquito transmission in a malaria parasite. *Cell.* 117:503-514

Blow JJ, Hodgson B (2002) Replication licensing – defining the proliferative state? *Trends Cell. Biol.* 12: 72-78

Bowers AJ, Boylan JF. (2004) Nek8, a NIMA family kinase member, is overexpressed in primary human breast tumors. *Gene.* 328:135-42.

Bozdech Z, Llinas M, Pulliam BL, Wong ED, Zhu J, De Risi JL. (2003) The transcriptome of the intraerythrocytic developmental cycle of *Plasmodium falciparum*. *PLoS Biol.* 1(1): E5

Bracchi-Ricard V, Barik S, Delvecchio C, Doerig C, Chakrabarti R, Chakrabarti D. (2000) PfPK6, a novel cyclin-dependent kinase/mitogen-activated protein kinase-related protein kinase from *Plasmodium falciparum*. *Biochem. J.* 347:255-63.

Bradford MM. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein using the principle of protein dye binding. *Anal. Biochem.* 72:248-54

Bunz F, Dutriaux A, Lengauer C, Waldman T, Zhou S, Brown JP, Sedivy JM, Kinzler KW, Vogelstein B. (1998) Requirement for p53 and p21 to sustain G2 arrest after DNA damage. *Science.* 282: 1497-501

Carucci DJ, Witney AA, Muhia DK, Warhurst DC, Schaap P, Meima M. (2000) Guanylyl cyclase activity associated with putative bifunctional integral membrane proteins in *Plasmodium falciparum*. *J. Biol. Chem.* 275: 22147-22156

Chan TA, Hwang PM, Hermeking H, Kinzler KW, Vogelstein B. (2000) Cooperative effects of genes controlling the G2/M checkpoint. *Genes Dev.* 14:1584-1588

Chang YT, Gray NG, Rosania G, Sutherlin DP, Kwon S, Norman TC, Sarohia R, Leost M, Meijer L, Schultz PG. (1999) Synthesis and application of functionally diverse 2,6,9-trisubstituted purine libraries as CDK inhibitors. *Chem. Biol.* 6:361-75

Chan M, Sim T. (2005) Functional analysis, overexpression, and kinetic characterization of pyruvate kinase from *Plasmodium falciparum*. *Bioch. Biophys. Res. Comm.* 326: 188-196

Chen DS, Zhu NL, Hung G, Skotzko MJ, Hinton DR, Tolo V, Hall FL, Anderson WF, Gordon EM. (1997) Retroviral vector-mediated transfer of an antisense cyclin G1 construct inhibits osteosarcoma tumor growth in nude mice. *Hum. Gene Ther.* 8(14):1667-74.

Chen A, Yanai A, Arama E, Kilfin G, Motro B. (1999) NIMA-related kinases: isolation and characterization of murine nek3 and nek4 cDNAs, and chromosomal localization of nek1, nek2 and nek3. *Gene.* 234(1):127-37.

Chen X. (2002) Cyclin G: a regulator of the p53-Mdm2 network. *Dev. Cell.* 5:518-9.

Chen Y, Riley D, Zheng L, Chen P, Lee W. (2002) Phosphorylation of the mitotic regulator protein hec1 by Nek2 kinase is essential for faithful chromosome segregation. *J. Biol. Chem.* 277: 49408-49416

Chishti AH, Maalouf GJ, Marfatia S, Palek J, Wang W, Fisher D, Lui SC (1994) Phosphorylation of protein 4.1 in *Plasmodium falciparum* infected human red blood cells. *Blood* 83: 3339-3345

- Cobrinik D. (2005) Pocket proteins and cell cycle control. *Oncogene*. 24(17):2796-809
- Conner SD and Schmid SL. (2003) Differential requirements for AP-2 in clathrin mediated endocytosis. *J. Cell. Biol.* 162(5):773-9
- Cooke BM, Lingelbach K, Bannister LH, Tilley L. (2004) Protein trafficking in *Plasmodium falciparum* – infected red blood cells. *Trends in Parasitology* 20: 581-589
- Dai Y, Grant S (2004) Small molecule inhibitors targeting cyclin-dependent kinases as anticancer agents. *Curr. Oncol. Rep.* 6:123-30
- Davies J R, Osmani A H, de Souza C P C, Bachewich C, Osmani S A (2004) Potential link between the NIMA mitotic kinase and nuclear membrane fission during mitotic exit in *Aspergillus nidulans*. *Eukaryotic Cell.* 3(6): 1433-1444
- Davidson EA, Gowda DC. (2001) Glycobiology of *Plasmodium falciparum*. *Biochimie.* 83, 601-604
- De Azevedo WF, Leclerc S, Meijer L, Havlicek L, Strnad M, Kim S.(1997) Inhibition of cyclin-dependent kinases by purine analogues: crystal structure of human cdk2 complexed with roscovitine. *Eur. J. Biochem.* 243:518-26
- De Bondt HL, Rosenblatt J, Jancarik J, Jones HD, Morgan DO, Kim SH (1993) Crystal structure of cyclin dependent kinase 2. *Nature* 363:595-602
- De Souza CPC, Osmani AH, Wu L, Spotts JL, Osmani SA. (2000) Mitotic histone H3 phosphorylation by the NIMA kinase in *Aspergillus nidulans*. *Cell.* 102:293-302
- Deng W, Baker DA. (2002) A novel cyclic GMP-dependent protein kinase is expressed in the ring stage of the *Plasmodium falciparum* life cycle. *Mol. Microbiol.* 44(5): 1141-1151
- Deng W, Parbu-Patel A, Meyer DJ, Baker D. (2003) The role of two novel regulatory sites in the activation of the cGMP-dependent protein kinase from *Plasmodium falciparum*. *Bioch. J.* 374:559-565
- Desjardins RE, Canfield CJ, Haynes JD, Chulay JD. (1979) Quantitative assessment of antimalarial activity *in vitro* by a semiautomated microdilution technique. *Antimicrob. Agents Chemother.* 16:710-18
- Di Agostino S, Rossi P, Geremia R, Sette C. (2002) The MAPK pathway triggers activation of Nek2 during chromosome condensation in mouse spermatocytes. *Development* 129: 1715-1727.
- Djian P, Phillips M, Easley K, Huang E, Simon M, Rice RH, Green H. (1993) The involucrin genes of the mouse and the rat: study of their shared repeats. *Mol. Biol. Evol.* 10(6): 1136-1149
- Dluzewski AR, Garcia CR. (1996) Inhibition of invasion and intraerythrocytic development of *Plasmodium falciparum* by kinase inhibitors. *Experientia* 52: 621-623

- Doerig C, Doerig C, Horrocks P, Coyle J, Carlton J, Sultan A, Arnot D, Carter R. (1995) Pfcrk-1, a developmentally regulated cdc2-related protein kinase of *Plasmodium falciparum*. *Mol. Biochem. Parasitol.* 70:167-74.
- Doerig C, Chakrabarti D, Kappes B, Matthews K (1999) The cell cycle in protozoan parasites. *Prog. Cell. Cycle. Res.* 4:163-84
- Doerig C, Endicott J, Chakrabarti D. (2002) Cyclin-dependent kinase homologues of *Plasmodium falciparum*. *Int. J. Parasit.* 32:1575-1585
- Doerig C, Chakrabarti D. (2004) Cell cycle control in *Plasmodium falciparum*: A genomics perspective. Chapter 10 p 249-287 in *Malaria parasites: Genomes and Molecular Biology*. Eds . Waters AP & Janse CP. Caister Academic Press, Wymondham, UK.
- Doerig C. (2004) Protein kinases as targets for anti-parasitic chemotherapy. *Biochim. Biophys. Acta.* 1697: 155-168
- Doerig CM, Parzy D, Langsley G, Horrocks P, Carter R, Doerig CD. (1996) A MAP kinase homologue from the human malaria parasite, *Plasmodium falciparum*. *Gene.* 177:1-6
- Dorin D, Alano P, Boccaccio I, Ciceron L, Doerig C, Sulpice R, Parzy D Doerig C. (1999) An atypical mitogen-activated protein kinase (MAPK) homologue expressed in gametocytes of the human malaria parasite. *J. Biol. Chem.* 274: 29912-29920
- Dorin D, Le Roch K, Sallicandro P, Alano P, Parzy D, Pouillet P, Meijer L, Doerig C. (2001) PfNek-1, a NIMA- related kinase from the human malaria parasite *Plasmodium falciparum*. *Eur. J. Biochem.* 268:2600-2608
- Dorin D, Semblat J, Pouillet P, Alano P, Goldring D, Whittle C, Patterson S, Chakrabarti D, Doerig C. (2005) PfPK7, an atypical MEK-related protein kinase, reflects the absence of classical three-component MAPK pathways in the human malaria parasite *Plasmodium falciparum*. *Mol. Microbiol.* 55:184-186
- Droucheau E, Primot A, Thomas V, Mattei D, Knockaert M, Richardson C, Sallicandro P, Alano P, Jafarshad A, Baratte B, Kunick C, Parzy D, Pearl L Doerig C, Meijer L. (2004) *Plasmodium falciparum* glycogen synthase kinase-3; molecular mode, expression, intracellular localization and selective inhibitors. *Biochim. Biophys. Acta.* 1697: 181-196
- Drucker BJ, Lydon NB. (2000) Lessons learnt from the development of an *abl* tyrosine kinase inhibitor for chronic myelogenous leukaemia. *J. Clin. Invest.* 105:3-7
- Elia AE, Cantley LC, Yaffe MB. (2003) Proteomic screen finds pSer/pThr-binding domain localizing Plk1 to mitotic substrates. *Science* 299: 1228-1231
- Endo Y, Fujita T, Tamura K, Tsuruga H, Nojima H.(1996) Structure and chromosomal assignment of the human cyclin G gene. *Genomics.* 38(1):92-5.
- Färber PM, Greaser R, Franklin RM, Kappes B. (1997) Molecular cloning and characterization of a second calcium-dependent protein kinase of *Plasmodium falciparum*. *Mol. Bioch. Parasitol.* 87: 211-216

- Fisher RP, Morgan DO. (1994) A novel cyclin associates with MO15/CDK7 to form the CDK-activating kinase. *Cell* 78:713-24
- Flatt PM, Tang LJ, Scatena CD, Szak ST, Pietenpol JA. (2000) p53 regulation of G(2) checkpoint is retinoblastoma protein dependent. *Mol. Cell Biol.* 12:4210-23.
- Foote SJ, Cowman A. (1994) The mode of action and the mechanisms of resistance to antimalarial drugs. *Acta Tropica* 56:157-71
- Fry AM, Meraldi P, Nigg EA. (1998) A centrosomal function for the human Nek2 protein kinase, a member of the Nima family of cell cycle regulators. *EMBO J.* 17:470-481
- Fry AM, Major T, Meraldi P, Stierhof YD, Tanaka K, Nigg EA. (2000) C-Nap1 a novel centrosomal coiled coil protein and candidate substrate for the cell cycle regulated protein kinase Nek2. *J. Cell. Biol.* 141: 1563-1574
- Fry AM. (2002) The Nek2 protein kinase: a novel regulator of centrosome structure. *Oncogene* 21: 6184-6194
- Gardner MJ, Hall N, Fung E, White O, Berriman M, Hyman RW, Carlton JM, Pain A, Nelson KE, Bowman S, Paulsen IT, James K, Eisen JA, Rutherford K, Salzberg SL, Craig A, Kyes S, Chan MS, Nene V, Shallom SJ, Suh B, Peterson J, Angiuoli S, Pertea M, Allen J, Selengut J, Haft D, Mather MW, Vaidya AB, Martin DM, Fairlamb AH, Fraunholz MJ, Roos DS, Ralph SA, McFadden GI, Cummings LM, Subramanian GM, Mungall C, Venter JC, Carucci DJ, Hoffman SL, Newbold C, Davis RW, Fraser CM, Barrell B. (2002) Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature* 419(6906):498-511.
- Garnham PCC. (1988) Malaria parasites of man: life cycles and morphology. In *Malaria; Principles and practise of malariology* (Eds Wersdorfer WH, McGregor I) Churchill Livingstone; 61-96
- Girard TJ, Gailani D, Broze GJ Jr. (1994) Complementary DNA sequencing of canine tissue factor pathway reveals a unique nanomeric repetitive sequence between the second and third Kunitz domains. *Biochem J.* 303: 923-928
- Grallert A, Hagan IM. (2002) *Schizosaccharomyces pombe* Nima related kinase, Fin1, regulates spindle formation and an affinity of Polo for the SPB. *EMBO J.* 21: 3096-3107
- Grana X, Reddy EP. (1995) Cell cycle control in mammalian cells: role of cyclins, cyclin dependent kinases (CDKs), growth suppressor genes and cyclin dependent kinase inhibitors (CKIs). *Oncogene.* 11:211-219
- Graeser R, Franklin RM, Kappes B. (1996) (a) Mechanisms of activation of the cdc2-related kinase PfPK5 from *Plasmodium falciparum*. *Mol. Biochem. Parasitol.* 79:125-7.
- Graeser R, Wernli B, Franklin RM, Kappes B. (b) (1996) *Plasmodium falciparum* protein kinase 5 and the malarial nuclear division cycles. *Mol. Biochem. Parasitol.* 82:37-49.
- Graeser R, Küry P, Franklin RM, Kappes B (1997) Characterization of a mitogen-activated protein (MAP) kinase from *Plasmodium falciparum*. *Mol. Microbiol.* 23(1) 151-159

Grant S, Roberts JD. (2003) The use of cyclin dependent kinase inhibitors alone or in combination with established cytotoxic drugs in cancer chemotherapy. *Drug Resist. Updat.* 6(1): 15-26

Gray NS, Wodicka L, Thunnissen AWH, Norman TC, Kwon S, Espinoza FH, Morgan DO, Barnes G, Leclerc S, Meijer L, Kim S, Lockhart DJ, Schultz PG. (1998) Exploiting chemical libraries, structure, and genomics in the search for kinase inhibitors. *Science.* 281:533-38

Gray NS, D tivaud L, Doerig C, Meijer L.(1999) ATP-site directed inhibitors of cyclin-dependent kinases. *Curr. Medicin. Chem.* 6(9):859-75

Greener T, Zhao X, Nojima H, Eisenberg E, Greene LE. (2000) Role of cyclin G-associated kinase in uncoating clathrin –coated vesicles from non-neuronal cells. *J. Biol. Chem.* 275;1365-1370

Giet R, Petretti C, Prigent C. (2005) Aurora kinases, aneuploidy and cancer : a coincidence or a real link? *Trends. Cell. Biol.* 15:241-250

Ginsburg H, Stein WD. (2005) How many functional transport pathways does *Plasmodium falciparum* induce in the membrane of its host erythrocyte? *Trends Parasitol.* 21:118-121

Godl K, Wissing J, Kurtenbach A, Habenberger P, Blencke S, Gutbrod H, Salassidis K, Stein-Gerlach M, Missio A, Cotten M, Daub H. (2003) An efficient proteomics method to identify the cellular targets of protein kinase inhibitors. *Proc. Natl. Acad. Sci. USA.* 100(26):15434-9

Hames RS, Wattam SL, Yamano H, Bacchieri R, Fry AM. (2001) APC/C-mediated destruction of the centrosomal kinase Nek2A occurs early in mitosis and depends upon a cyclin A-type D-box. *EMBO J.* 20:7117-7127

Hanks SK, Hunter T. (1995) The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification. *FASEB J.* 9: 576-595

Hanks SK. (2003) Genomic analysis of the eukaryotic protein kinase superfamily: a perspective. *Genome Biology.* 14 (111):1-7

Harmse L, Van Zyl R, Gray N, Schultz P, Leclerc S, Meijer L, Doerig C, Havlik I. (2001) Structure activity relationships and inhibitory effects of various purine derivatives on the in vitro growth of *Plasmodium falciparum*. *Biochem. Pharmacol.* 62: 341-348

Harris SL, Levine AJ. (2005) The p53 pathway: positive and negative feedback loops. *Oncogene.* 24(17):2899-908.

Hayward RE. (2000) *Plasmodium falciparum* phosphoenolpyruvate carboxykinase is developmentally regulated in gametocytes. *Mol. Bioch. Parasit.* 107: 227-240

Hashimoto Y, Akita H, Hibino M, Kohri K, Nakanishi M. (2002) Identification and characterization of Nek6 protein kinase, a potential human homolog of NIMA histone H3 kinase. *Bioch. Biophys. Res. Commun.* 293:753- 758

- Helps NR, Luo X, Barker HM, Cohen PTW. (2000) NIMA related kinase 2 (Nek2), a cell cycle regulated protein kinase localized to centrosomes, is complexed to protein phosphatase1. *Biochem. J.* 349:509-518
- Hermeking H, Lengauer C, Polyak K, Zhang L, Tiagalingam S, Kinzler KW, Vogelstein B. (1997) 14-3-3 sigma is a p53-regulated inhibitor of G2/M progression. *Mol. Cell.* 1(1):3-11.
- Hirai H, Kawanishi N, Iwasawa Y. (2005) Recent advances in the development of selective small molecule inhibitors for cyclin dependent kinases. *Curr. Top. Med. Chem.* 5(2):167-179
- Hiller NL, Bhattacharjee S, Van Ooij C, Liolios K, Harrison T, Lopez-Estrano C, Haldar K. (2004) A host-targeting signal in virulence proteins reveals a secretome in malarial infection. *Science.* 306: 1934-1937
- Holland PM, Milne A, Garka K, Johnson RS, Willis C, Sims JE, Rauch CT, Bird TA, Virca GD. (2002) Purification, cloning, and characterization of Nek8, a novel NIMA-related kinase, and its candidate substrate Bicd2. *J. Biol. Chem.* 277(18):16229-40.
- Holton S, Merckx A, Burgess D, Doerig C, Noble M, Endicott. (2003) Structures of *P. falciparum* PfPK5: testing the CDK regulation paradigm and mechanisms of small molecules inhibition. *Structure.* 11: 1329-1337
- Horne MC, Goolsby GL, Donaldson KL, Tran D, Neubauer M, Wahl AF. (1996) Cyclin G1 and cyclin G2 comprise a new family of cyclins with contrasting tissue-specific and cell cycle-regulated expression. *J. Biol. Chem.* 271(11):6050-61.
- Hulo N, Sigrist CJA, Le Saux V, Langendijk-Genevaux PS, Bordoli L, Gattiker A, De Castro E, Bucher P, Bairoch A. (2004) Recent improvements to the PROSITE database. *Nucleic Acids Research.* Vol 32: Database issue D134-D137
- Innocente SA. (1999) p53 regulates a G2 checkpoint through cyclin B1. *Proc. Natl. Acad. Sci. USA* 96, 2147-2152
- Janse CJ, Ponnudurai T, Lensen AH, Meuwissen JH, Ramesar J, Van der Ploeg M, Overdulve JP. (1988) DNA synthesis in gametocytes of *Plasmodium falciparum*. *Parasitol.* 96(1):1-7
- Jeffrey PD, Russo AA, Polyak K, Gibbs E, Hurwitz J, Massague J, Pavletich NP. (1995) Mechanism of CDK activation revealed by the structure of a Cyclin A-CDK2 complex. *Nature.* 376: 313-320
- Jensen JB, Trager W. (1977) *Plasmodium falciparum* in culture: use of outdated erythrocytes and description of the candle jar method. *J. Parasitol.* 63:883-86
- Johnson DG, Walker CL. Cyclins and cell cycle checkpoints. (1999) *Annu. Rev. Pharmacol. Toxicol.* 39:295-312
- Kanaoka Y, Kimura SH, Okazaki I, Ikeda M, Nojima H. (1997) GAK: a cyclin G associated kinase contains a tensin/auxilin-like domain. *FEBS Lett.* 402(1):73-80.

Kawabe T, Saganuma M, Ando T, Kimura M, Hori H, Okamoto T. (2002) CDC25C interacts with PCNA at G2/M transition. *Oncogene*. 21:1717-26

Kawamoto F, Fujioka H, Marakami RI, Hagiwara M, Ishikawa T, Hidaka H. (1993) The roles of Ca^{2+} / calmodulin and cGMP pathways in gametocytogenesis of a rodent malaria parasite, *Plasmodium bergeri*. *Eur. J. Cell. Biol.* 60: 101-107

Kearns AE, Donohue MM, Sanyal B, Demay MB. (2001) Cloning and characterization of a novel protein kinase that impairs osteoblast differentiation *in vitro*. *J. Biol. Chem.* 276: 42213-42218

Keenan SM, Welsh WJ. (2004) Characteristics of the *Plasmodium falciparum* PK5 ATP-binding site: implications for the design of novel antimalarial agents. *J. Mol. Graph. Model.* 22;241-247

Kehoe JW, Bertozzi CR. (2000) Tyrosine sulfation: a modulator of extracellular protein-protein interactions. *Chem. & Biol.* 7: R57-R61

Kelly TJ, Brown GW. (2000) Regulation of chromosome replication. *Annu. Rev. Biochem.* 69:829-80

Kilbey BJ, Fraser I, McAleese S, Goman M, Ridley RG. (1993) Molecular characterisation and stage specific expression of proliferating cell nuclear antigen (PCNA) from the malarial parasite, *Plasmodium falciparum*. *Nucleic Acids Res.* 21: 239-243

Kim YH, Choi JY, Jeong Y, Wolgemuth DJ, Rhee K. (2002) Nek2 localizes to multiple sites in mitotic cells suggesting its involvement in multiple cellular functions during the cell cycle. *Bioch. Biophys. Res. Comm.* 290:730-736

Kimura SH, Tsuruga H, Yabuta N, Endo Y, Nojima H. (1997) Structure, expression, and chromosomal localization of human GAK. *Genomics*. 44:179-187

Knockaert M, Gray N, Damiens E, Chang YT, Grellier P, Grant K, Fergusson D, Mottram J, Soete M, Dubremetz J, LeRoch K, Doerig C, Schultz P G, Meijer L. (2000) Intracellular targets of cyclin-dependent kinase inhibitors: identification by affinity chromatography using immobilised ligands. *Chem. Biol.* 7:411-422

Knockaert M, Meijer L. (2002) Identifying *in vivo* targets of cyclin-dependent kinase inhibitors by affinity chromatography. *Biochem. Pharmacol.* 64:819-25.

Koepp DM, Harper JW, Elledge SJ. (1999) How the cyclin became a cyclin: regulated proteolysis in the cell cycle. *Cell*. 97(4):431-4.

Kolev V, Ivanov I, Berzal-Herranz A, Ivanov IG. (2003) Non-Shine Dalgarno initiators selected from combinatorial DNA libraries. *J. Mol. Microbiol. Biotech.* 5(3): 154-60

Korolchuck VI, Banting G. (2002) CK2 and GAK/auxilin2 are major protein kinases in clathrin coated vesicles. *Traffic* 3:428-439

- Krien MJE, West RR, John UP, Koniaras K, McIntosh J R, O'Connel. (2002) The fission yeast Nima kinase Fin1p is required for spindle function and nuclear envelope integrity. *EMBO J.* 21:1713-1722
- Kumar A, Vaid A, Syin C, Sharma P. (2004) PfPKB a novel protein kinase B-like enzyme from *Plasmodium falciparum*. *J. Biol. Chem.* 279:24255- 24264
- Lacrima K, Valentini A, Lambertini C, Tadorelli M, Rinaldi A, Zucca E, Catapano C, Cavalli F, Gianella-Boradori A, McCallum DE, Bertoni F. (2005) *In vitro* activity of cyclin dependent kinase inhibitor CYC202 (Seliciclib, R-roscovitine) in mantle cell lymphomas. *Ann. Oncol.* 16(7):1169-76
- Laemmli UK. (1970) Cleavage of structural proteins during the assembly of the head of *Bacteriophage T4*. *Nature.* 227: 680-685
- Lambros C, Vanderberg JP. (1979) Synchronization of *Plasmodium falciparum* erythrocytic stages in culture. *J. Parasitol.* 65:418-20
- Larochelle S, Pandur J, Fisher R P, Salz H K, Suter B. (1998) CDK7 is essential for mitosis and for *in vivo* Cdk-activating kinase activity. *Genes. Dev.* 12:370-381
- Lee BH and Amon A. (2003) Polo kinase: meiotic cell cycle coordinator. *Cell Cycle.* 2:400-402
- Lee MH, Yang HY. (2001) Negative regulators of cyclin dependent kinases and their roles in cancers. *Cell. Mol. Life Sci.* 58:1907-1922
- Leete TH, Rubin H. (1996) Malaria and the cell cycle. *Parasitol. Today.* 11:442-4
- Le Roch K, Sestier C, Dorin D, Waters N, Kappes B, Chakrabarti D, Meijer L, Doerig C. (2000) Activation of a *Plasmodium falciparum* cdc2-related kinase by heterologous p25 and cyclin H. Functional characterization of a *P. falciparum* cyclin homologue. *J. Biol. Chem.* 275:8952-8.
- Lei M and Tye BK. (2001) Initiating DNA synthesis: from recruiting to activating the MCM complex. *J. Cell. Sci.* 114:1447-1454
- Letwin K, Mizzen L, Motro B, Ben-David Y, Bernstein A, Pawson T. (1992) A mammalian dual specificity protein kinase, Nek1, is related to the NIMA cell cycle regulator and highly expressed in meiotic germ cells. *EMBO J.* 11(10):3521-31.
- Loyer P, Trembley JH, Katona R, Kidd VJ, Lathi JM. (2005) Role of CDK/cyclin complexes in transcription and RNA splicing. *Cellular Signalling.* 17: 1033-1051
- Li Z, Le Roch K, Geyer JA, Woodard CL, Prigge ST, Koh J, Doerig C, Waters NC. (2001) Influence of human p16(INK4) and p21(CIP1) on the *in vitro* activity of recombinant *Plasmodium falciparum* cyclin-dependent protein kinases. *Biochem. Biophys. Res. Commun.* 288:1207-11.

- Li JL, Robson KJ, Chen JL, Targett GA, Baker DA. (1996) Pfmrk, a MO15-related protein kinase from *Plasmodium falciparum*. Gene cloning, sequence, stage-specific expression and chromosome localization. *Eur. J. Biochem.* 241:805-13. Erratum in: *Eur. J. Biochem* 243: 537.
- Li JL, Baker D, Cox LS. (2000a) Sexual stage specific expression of a third calcium-dependent protein kinase from *Plasmodium falciparum*. *Biochim. Biophys. Acta.* 1491: 341-349
- Li JL and Cox LS. (2000b) Isolation and characterisation of a cAMP-dependent protein kinase catalytic subunit gene from *Plasmodium falciparum*. *Mol. Bioch. Parasitol* 109: 157-163
- Li JL and Cox LS. (2001) Identification of an MCM homologue expressed specifically in the sexual stage of *Plasmodium falciparum*. *Int. J. Parasitol.* 31:1246-1252
- Li JL and Cox LS. (2003) Characterisation of a sexual stage-specific gene encoding ORC1 homologue in the human malaria parasite *Plasmodium falciparum*. *Parasitol. Int.* 52:41-52
- Lin DT, Goldman ND, Syin C. (1996) Stage specific expression of a *Plasmodium falciparum* protein related to the eukaryotic mitogen activated protein kinases. *Mol. Bioch. Parasit.* 78:67-77
- Lizcano JM, Deak M, Morrice N, Keiloch A, Hastie CJ, Dong L, Schutkowski M, Reimer U, Alessi DR. (2002) Molecular basis for the substrate specificity of NIMA-related kinase 6 (Nek6). *J. Biol. Chem.* 277: 27839-49
- Liu S, Lu W, Obara T, Kuida S, Lehoczy J, Dewar K, Drummond IA, Beier DR. (2002) A defect in a novel Nek-family kinase causes cystic kidney disease in the mouse and in zebrafish. *Development.* 129(24):5839-46.
- Maccioni RB, Otth C, Concha II and Munoz JP. (2001) The protein kinase CDK5: Structural aspects, roles in neurogenesis and involvement in Alzheimer's pathology. *Eur. J. Biochem.* 268: 1518-1527
- Mahjoub MR, Montpetit B, Zhao L, Finst R J, Goh B, Kim AC, Quarmby LM. (2002) The FA2 gene of *Chlamydomonas* encodes a NIMA family kinase with roles in cell cycle progression and microtubule severing during deflagellation. *J. Cell. Sci.* 115:1759-1768
- Malumbres M, Barbacid M. (2001) To cycle or not to cycle: a critical decision in cancer. *Nat. Rev. Cancer.* 1:222-231
- Marti M, Good RT, Rug M, Knuepfer E, Cowman AF. (2004) Targeting malaria virulence and remodelling proteins to the host erythrocyte. *Science.* 306:1930-1933
- Maurer-Stroh S, Eisenhaber F. (2005) Refinement and prediction of protein prenylation motifs. *Genome Biology.* 6:R55

Mehlin C, Boni E, Buckner FS, Engel L, Feist T, Gelb MH, Haji L, Kim D, Lui C, Mueller N, Myler PJ, Reddy JT, Sampson JN, Subramanian E, Van Voorhis WC, Worthey E, Zucker F, Hol WGJ. (2006) Heterologous expression of proteins from *Plasmodium falciparum*: Results from 1000 genes. *Mol. Biochem. Parasitology*. 148:144-160

Meijer L, Kim S. (1997) Chemical inhibitors of cyclin-dependent kinases. *Meth. Enzymology*. 283:113-28

Meijer L, Raymond E. (2003) Roscovitine and other purines as kinase inhibitors. From starfish oocytes to clinical trials *Acc. Chem. Res.* 36: 417-425

Meijer L, Leclerc S, Leost M. (1999) Properties and potential applications of chemical inhibitors of cyclin-dependent kinases. *Pharmacol. Ther.* 82:279-84

Meijer L, Borgne A, Mulner O, Chong JPI, Blow JJ, Inagaki N, Inagaki M, Delcros JG, Moulinoux JP. (1997) Biochemical and cellular effects of roscovitine, a potent and selective inhibitor of the cyclin-dependent kinases cdc2, cdk2 and cdk5. *Eur. J. Biochem.* 243:527-36

Meraldi P, Honda R, Nigg EA. (2004) Aurora kinases link chromosome segregation and cell division to cancer susceptibility. *Curr. Opin. Gen. Dev.* 14:29-36

Merckx A, Le Roch K, Nivez MP, Dorin D, Alano P, Gutierrez GJ, Nebreda AR, Goldring D, Whittle C, Patterson S, Chakrabarti D, Doerig C. (2003) Identification and initial characterization of three novel cyclin-related proteins of the human malaria parasite *Plasmodium falciparum*. *J. Biol. Chem.* 278:39839-50.

Michael D, Oren M. (2003) The p53-mdm2 module and the ubiquitin system. *Sem. Cancer Biol.* 13:49-58

Miletich JP, Broze GJ Jr. (1990) Beta protein C is not glycosylated at asparagine 329. The rate of translation may influence the frequency of usage at asparagine-X-cysteine sites. *J. Biol. Chem.* 265(19): 11397-404

Mockenhaupt FP. (1995) Mefloquine resistance in *Plasmodium falciparum*. *Parasitol. Today*. 11(7):248-53

Molinari M (2000) Cell cycle checkpoints and their inactivation in human cancer. *Cell Prolif.* 33:261-74.

Monaco EA, Vallano ML (2003) Cyclin dependent kinase inhibitors: cancer killers to neuronal guardians. *Curr. Med. Chem.* 10(5):367-79

Monaghan P, Bell A. (2005) A *Plasmodium falciparum* FK506-binding protein (FKBP) with peptidyl-prolyl cis-trans isomerase and chaperone activities. *Mol. Biochem. Parasitol.* 139(2):185-95

Moores SL, Schaber MD, Mosser SD, Rands E, O'Hara M, Garsky V, Marshall MS, Pompliano DL, Gibbs J. (1991) Sequence dependence of protein isoprenylation. *J. Biol. Chem.* 266:14603-14610.

- Morgan DO. (1995) Principles of CDK regulation. *Nature*. 374:131-4
- Morgan DO. (1997) Cyclin dependent kinases: engines, clocks and microprocessors. *Annu. Rev. Cell. Dev. Biol.* 13: 261-91
- Morita N, Kiryu S, Kiyama H. (1996) p53-independent cyclin G expression in a group of mature neurons and its enhanced expression during nerve regeneration. *J. Neurosci.* 16 (19):5961-6.
- Möskes C, Burghaus PA, Wernli B, Sauder U, Durrenberger M, Kappes B. (2004) Export of *Plasmodium falciparum* calcium dependent protein kinase 1 to the parasitophorous vacuole is dependent on the three N-terminal membrane anchor motifs. *Mol. Micro.* 54(3):676-691
- Musacchio A and Hardwick KG. (2002) The spindle checkpoint : structural insights into dynamic signalling. *Nat. Rev. Mol. Cell. Biol.* 3:731-741
- Noguchi K, Fukazawa H, Murakami Y, Uehara Y. (2002) Nek11, a new member of the NIMA family of kinases, involved in DNA replication and genotoxic stress responses. *J. Biol. Chem.* 277: 39655-39665
- Nigg EA. (2001) Mitotic kinases as regulators of cell division and its checkpoints. *Nat. Rev. Mol. Cell. Biol.* 2:21-32
- O'Connell MJ, Norbury C, Nurse P. (1994) Premature chromatin condensation upon accumulation of NIMA. *EMBO J* .13: 4926- 4937
- O'Connell MJ, Krien MJE, Hunter T. (2003) Never say never. The NIMA related kinases in mitotic control. *Trends in Cell. Biol.* 13: 221-228
- Ohkura H. (2003) Phosphorylation: Polo kinase joins an elite club. *Curr. Biol.* 13: R912-R914.
- Oren M, Damalas A, Gottlieb T, Michael D, Taplick J, Leal JFM, Maya R, Moas M, Seger R, Taya Y, Ben-Ze'ev A. (2002) Regulation of p53: intricate loops and delicate balances. *Bioch. Pharmacol.* 64: 865-871
- Ohtsuka T, Jensen MR, Kim HG, Kim KT, Lee SW. (2004) The negative role of cyclin G in ATM-dependent p53 activation. *Oncogene.* 23(31):5405-8
- Ohtsuka T, Ryu H, Minamishima YA, Ryo A, Lee SW. (2003) Modulation of p53 and p73 levels by cyclin G: implication of a negative feedback regulation. *Oncogene.* 11:1678-87.
- Olusanya O, Andrews PD, Swedlow JR, Smythe E. (2001) Phosphorylation of threonine 156 of the mu2 subunit of the AP2 complex is essential for endocytosis *in vitro* and *in vivo*. *Curr. Biol.* 11:896-900
- Okamoto K, Beach D. (1994) Cyclin G is a transcriptional target of the p53 tumour suppressor protein. *EMBO J.* 13(20):4816-22.

- Okamoto K, Li H, Jensen MR, Zhang T, Taya Y, Thorgeirsson S, Prives C. (2002) Cyclin G recruits PP2A to dephosphorylate Mdm2. *Mol. Cell* 9:761-771
- Orjih AU (1994) Saponin haemolysis for increasing concentration of *Plasmodium falciparum* infected erythrocytes. *The Lancet*. 343:295
- Osmani SA, Pu RT, Morris NR. (1988) Mitotic induction and maintenance by overexpression of a G2-specific gene that encodes a potential protein kinase. *Cell* 53, 237 - 244
- Osmani SA, May GS, Morris NR (1987) Regulation of the mRNA levels of nimA, a gene required for the G2-M transition in *Aspergillus nidulans*. *J. Cell. Biol.* 104:1495-1504
- Osmani A H, O'Donnel K, Pu RT, Osmani SA. (1991) Activation of the nimA protein kinase plays a unique role during mitosis that cannot be bypassed by absence of the bimE checkpoint. *EMBO J.* 10:2669-2679
- Patterson S, Whittle C, Robert C, Chakrabarti D. (2002) Molecular characterization and expression of an alternate proliferating cell nuclear antigen homologue, PfPCNA2 , in *Plasmodium falciparum*. *Biochem. Biophys. Res. Commun.* 298: 371-6
- Patterson S, Robert C, Whittle C, Chakrabarti R, Doerig C, Chakrabarti D. (2006) Pre-replication complex organization in the atypical DNA replication cycle of *Plasmodium falciparum*: Characterization of the mini-chromosome maintenance (MCM) complex formation. *Mol. Biochem. Parasitol.* 145: 50-59
- Pal B, Pybus B, Muccio DD, Chattopadhyay D. (2004) Biochemical characterization and crystallization of recombinant 3-phosphoglycerate kinase of *Plasmodium falciparum*. *Bioch. Biophys. Acta* 1699: 277-280
- Paulovich A. (1997) When checkpoints fail. *Cell*. 88: 315-321
- Petrocelli T, Poon R, Drucker DJ, Slingerland JM, Rosen CF. (1996) UVB radiation induces p21Cip1/WAF1 and mediates G1 and S phase checkpoints. *Oncogene*. 12(7):1387-96.
- Perez R, Wu N, Klipfel AA, Beart RW. (2003) A better cell cycle target for gene therapy of colorectal cancer: cyclin G. *J. Gastrointest. Surg.* 7:884-9.
- Polci R, Peng A, Chen PL, Riley DJ, Chen Y.(2004) NIMA-related protein kinase 1 is involved early in the ionising radiation-induced DNA damage response. *Cancer Res.* 64(24):8800-3.
- Pollard TD, Earnshaw EC. (2004) *Cell Biology*. Saunders, Elsevier Science, USA
- Preibisch G, Ishihara H, Tripier D, Leineweber M. (1988) Unexpected translation initiation within the coding region of eukaryotic genes expressed in *Eschericia coli*. *Gene*. 72: 179-86
- Prigent C, Glover D, Giet R. (2005) *Drosophila* Nek2 protein kinase knockdown leads to centrosome maturation defects while overexpression causes centrosome fragmentation and cytokinesis failure. *Exp. Cell. Res.* 303:1-13

- Pu RT, Gang X, Wu L, O'Donnel K, Ye XS, Osmani SA. (1992) Isolation of a functional homolog of the cell cycle specific NIMA protein kinase of *Aspergillus nidulans* and functional analysis of conserved residues. *J. Biol. Chem.* 270, 18110-18116
- Pu RT, Osmani SA. (1995) Mitotic destruction of the cell cycle regulated NIMA protein kinase of *Aspergillus nidulans* is required for mitotic exit. *EMBO J.* 14: 995-1003
- Ray MR, Wafa LA, Cheng H, Snoek R, Fazli L, Gleave M, Rennie PS. (2006) Cyclin G associated kinase: A novel androgen receptor-interacting transcriptional coactivator that is overexpressed in hormone refractory prostate cancer. *Int. J. Cancer.* 118(5): 1108-19
- Read LK, Mikkelsen RB. (1991) *Plasmodium falciparum* infected erythrocytes contain an adenylate cyclase with properties which differ from those of the host enzyme. *Mol. Bioch. Parasitol.* 45: 109-120
- Read M, Sherwin T, Holloway SP, Gull K, Hyde JE. (1993) Microtubular organization visualized by immunofluorescence microscopy during erythrocytic schizogony in *Plasmodium falciparum* and investigation of post-translational modifications of parasite tubulin. *Parasitol* 106: 223-232
- Read M, Hyde JE. (1996) On malaria and the cell cycle. *Parasitol. Today.* 12:442-4.
- Reimer CL, Borrás AM, Kurdistan SK, Garreau JR, Chung M, Aaronson SA, Lee SW. (1999) Altered regulation of cyclin G in human breast cancer and its specific localization at replication foci in response to DNA damage in p53^{+/+} cells. *J. Biol. Chem.* 274:11022-11029
- Reininger L, Billker O, Tewari R, Mukhopadhyay A, Fennel C, Dorin-Semblatt D, Doerig C, Goldring D, Harmse L, Ranford-Cartwright L, Packer J, Doerig C (2005) A NIMA related protein kinase is essential for completion of the sexual cycle of malaria parasites. *J. Biol. Chem.* 280 (36): 31957-64
- Rialet V, Meijer L.(1991) A new screening test for antimitotic compounds using the universal M phase-specific protein kinase, p34^{cdc2}/cyclin B^{cdc13}, affinity-immobilized on p13^{suc1}-coated microtitration plates. *Anticancer Res.* 11:1581-590
- Ridley RG, White JH, McAleese SM, Goman M, Alano P, De Vries E, Kilbey BJ. (1991) DNA polymerase delta: gene sequences from *Plasmodium falciparum* indicate that this enzyme is more highly conserved than DNA polymerase alpha. *Nucleic Acids Res.* 19: 6731-6736
- Ridley RG. (2002) Medical need, scientific opportunity and the drive for antimalarial drugs. *Nature* 415:686-693
- Roig J, Mikhailov A, Belham C, Avruch J. (2002) Nercc1, a mammalian NIMA-family kinase, binds the Ran GTPase and regulates mitotic progression. *Genes Dev.* 16(13):1640-58.
- Roskoski R(Jr) (2003) Protein prenylation: a pivotal posttranslational process. *Bioch. Biophys. Res. Commun.* 303:1-7

- Ross-Macdonald PB, Graeser R, Kappes B, Franklin R, Williamson DH. (1994) Isolation and expression of a gene specifying a cdc2-like protein kinase from the human malaria parasite *Plasmodium falciparum*. *Eur. J. Biochem.* 220, 693-701.
- Sambrook J, Fritsch EF, Maniatis T (1989) Molecular Cloning: A laboratory manual. 2nd edition. Cold Spring Harbor Laboratory Press, USA
- Sanger F, Nicklen S, Coulson AR. (1977) DNA sequencing with chain terminating inhibitors. *Proc. Natl. Acad. Sci. USA* 74:5463-5467
- Schenk PW, Snaar-Jagalska BE. (1999) Signal perception and transduction: the role of protein kinases. *Biochim. Biophys. Acta.* 1449:1-24
- Schneider AG, Mercereau-Puijalon O. (2005) A new Apicomplexa-specific protein kinase family: multiple members in *Plasmodium falciparum*, all with an export signature. *BMC Genomics.* 6:30
- Scholey JM, Brust-Masher I, Mogilner A. (2003) Cell division. *Nature* 422:746-752
- Shapiro GI, Edwards CD, Rollins BJ. (2000) The physiology of p16(INK4a)-mediated G1 proliferative arrest. *Cell. Biochem. Biophys.* 33:189-197
- Sharma A. (2000) Protein tyrosine kinase activity in human malaria parasite *Plasmodium falciparum*. *Indian J. Exp. Biol.* 38(12):1222-6
- Sherr CJ and Roberts JM. (1999) CDK inhibitors : positive and negative regulators of G1 phase progression. *Genes Dev.* 13: 1501-1512
- Shimizu A, Nishida J, Ueoka Y, Kato K, Hachiya T, Kuriaki Y, Wake N. (1998) Cyclin G contributes to G2/M arrest of cells in response to DNA damage. *Biochem. Biophys. Res. Commun.* 242(3):529-33.
- Sina BJ and Aultman SK. (2001) Resisting resistance. *Trends Parasitol.* 17: 305-306
- Sinden RE, Butcher GA, Billker O, Fleck SL. (1996) Regulation of infectivity of *Plasmodium* to the mosquito vector. *Adv. Parasitol.* 1996;38:53-117.
- Skalhegg BS and Tasken K. (2000) Specificity on the cAMP/ PKA signalling pathway. Differential expression, regulation, and subcellular localization of subunits of PKA. *Front. Biosci.* 5: D678-693
- Skotzko M, Wu L, Anderson WF, Gordon EM, Hall FL.(1995) Retroviral vector-mediated gene transfer of antisense cyclin G1 (CYCG1) inhibits proliferation of human osteogenic sarcoma cells. *Cancer Res.* 55(23):5493-8.
- Smith ML, Kontny HU, Bortnick R, Fornace AJ Jr.(1997) The p53-regulated cyclin G gene promotes cell growth: p53 downstream effectors cyclin G and Gadd45 exert different effects on cisplatin chemosensitivity. *Exp. Cell Res.* 230(1):61-8.
- SmitsVAJ and Medema RH. (2001) Checking out the G2/M transition. *Biochim. Biophys. Acta.* 1519:1-12

- Smythe E. (2002) Regulating the clathrin coated vesicle cycle by AP2 subunit phosphorylation. *Trends in Cell Biol.* 12:352-354
- Stewart ZA and Pietsenpol JA. (2001) p53 signalling and cell cycle checkpoints. *Chem. Res. Toxicol.* 14: 243-263
- Stewart ZA, Westfall MD and Pietsenpol JA. (2003) Cell cycle dysregulation and anticancer therapy. *Trends in Pharm. Sci.* 24: 139-145
- Syin C, Parzy D, Traincard F, Boccaccio I, Joshi MB, Lin DT, Yang X, Assemet K, Doerig C, Langsley G. (2001) The H89 cAMP –dependent protein kinase inhibitor blocks *Plasmodium falciparum* development in infected erythrocytes. *Eur. J. Biochem.* 268: 4842-4849
- Surpili M, Delben T M, Kobarg J. (2003) Identification of proteins that interact with the central coiled-coil region of the human protein kinase Nek1. *Biochemistry.* 42:15369-15376
- Tan BC, Lee SC. (2004) Nek9, a novel FACT-associated protein, modulates interphase progression. *J. Biol. Chem.* 279(10):9321-30.
- Tanaka K, Nigg EA. (1999) Cloning and characterization of the murine Nek3 protein kinase, a novel member of the NIMA family of putative cell cycle regulators. *J. Biol. Chem.* 274: 13491-13497
- Tamura K, Kanaoka Y, Jinno S, Nagata A, Ogiso Y, Shimizu K, Hayakawa T, Nojima H, Okayama H. (1993) Cyclin G: a new mammalian cyclin with homology to fission yeast Cig1. *Oncogene.* 8(8):2113-8.
- Taylor SL, Kinchington PR, Brooks A, Moffat JF. (2004) Roscovitine, a cyclin dependent kinase inhibitor, prevents replication of varicella-zoster virus. *J. Virol.* 78(6):2853-62
- Tassan JP, Jaquenoud M, Fry AM, Frutiger S, Hughes GJ, Nigg EA. (1995) *In vitro* assembly of a functional human CDK7-cyclin H complex requires MAT1, a novel 36 kDa RING finger protein. *EMBO J.* 14(22):5608-17.
- Templeton TJ, Deitsch KW.(2005) Targeting malaria parasite proteins to the erythrocyte. *Trends in Parasitology.* 21:399-402
- Thomas JG, Baneyx F. (1996) Protein misfolding and inclusion body formation in recombinant *Escherichia coli* cells overexpressing heat shock proteins. *J. Biol. Chem.* 271, 11141-11147
- Umeda A, Meyerholz A, Ungewickell E. (2000) Identification of the universal cofactor (auxilin 2) in clathrin coat dissociation. *Eur. J. Cell. Biol.* 79: 336-342
- Upadhyay P, Birkenmeier EH, Birkenmeier CS, Barker JE. (2000) Mutations in a NIMA-related kinase gene, Nek1, cause pleiotropic effects including a progressive polycystic kidney disease in mice. *Proc. Natl. Acad. Sci. USA.* 97(1):217-21.
- Vesely J, Havlicek L, Strnad M, Blow JJ, Donella-Deana A, Pinna L, Letham DS, Kato JY, D tivaud L, Leclerc S, Meijer L. (1994) Inhibition of cyclin-dependent kinases by purine derivatives. *Eur. J. Biochem.* 224:771-86

Vivanco I, Sawyers CL. (2002) The phosphatidylinositol 3-kinase-AKT pathway in human cancer. *Nature Reviews/Cancer* 2:489-501

Voss TS, Mini T, Jenoe P, Beck HP. (2002) *Plasmodium falciparum* possesses a cell cycle regulated short-type replication protein A large subunit encoded by an unusual long transcript. *J. Biol. Chem.* 277:17493-501

Wang S, Nakashima S, Sakai H, Numata O, Fujiju K, Nozawa Y. (1998) Molecular cloning and cell cycle dependent expression of a novel NIMA-related protein kinase (TpNrk) in *Tetrahymena* cells. *Biochem. J.* 334: 197-203

Wang Y, Roach PJ. (1993) Purification and assay of mammalian protein (serine/ threonine) kinases in Protein phosphorylation : A Practical Approach. Ed. Hardie DG, Oxford University Press, New York, USA.

Ward P, Equinet L, Packer J, Doerig C. (2004) Protein kinases of the human malaria parasite *Plasmodium falciparum*: the kinome of a divergent eukaryote. *BMC Genomics.* 5:79

Waters NC, Woodard CL, Prigge ST. (2000) Cyclin H activation and drug susceptibility of the Pfmrk cyclin dependent protein kinase from *Plasmodium falciparum*. *Mol. Biochem. Parasitol.* 107, 45-55.

White JH, Kilbey BJ, De Vries E, Goman M, Alano P, Cheesman S, McAleese S, Ridley RG. (1993) The gene encoding DNA polymerase alpha from *Plasmodium falciparum*. *Nucleic Acids Res.* 21: 3643-3646

Willems AR, Lanker S, Patton EE, Craig KL, Nason TF, Mathias N, Kobayashi R, Wittenberg C, Tyers M. (1996) Cdc53 targets phosphorylated G1 cyclins for degradation by the ubiquitin proteolytic pathway. *Cell.* 86:453-63

Winstanley PA. (2000) Chemotherapy for *Falciparum* malaria: the armoury, the problems and the prospects. *Parasitol. Today.* 16:146-153

Wiser MF.(2003) A *Plasmodium* homologue of cochaperone p23 and its differential expression during the replicative cycle of the malaria parasite. *Parasitol Res.* 90:166-70

Wu L, Osmani SA, Mirabito PM. (1998) A role for NIMA in the nuclear localization of cyclin B in *Aspergillus nidulans*. *J. Cell Biol.* 141:1575-87

Yam CH, Fung TK, Poon RY. (2002) Cyclin A in cell cycle control and cancer. *Cell. Mol. Life. Sci.* 59: 1317-1326

Ye XS and Osmani SA. (1997) Regulation of p34^{cdc2}/cyclin B H1 and Nima kinases during the G2/M transition and checkpoint responses in *Aspergillus nidulans*. *Prog. Cell. Cycle. Res.* 3: 221-232

Ye XS, Lee S, Wolkow T D, McGuire S, Hamer J E, Wood C, Osmani S A. (1999) Interaction between developmental and cell cycle regulators is required for morphogenesis in *Aspergillus nidulans*. *EMBO J.* 18(24): 6994-7001

Yin X, Warner DR, Roberts E, Pisano MM, Greene RM. (2005) Identification of novel CBP interacting proteins in embryonic orofacial tissue. *Bioch. Biophys. Res Comm.* 329: 1010-1017

Young DA, Nickerson-Nutter CL. (2005) mTOR-beyond transplantation. *Curr. Opin. Pharmacol.* 5:418-423

Yoo J C, Chang J R, Kim S H, Jang S K, Wolgemuth D J, Kim K, Rhee K. (2004) NIP1/XB51/ NECAB3 is a potential substrate of Nek2, suggesting specific roles of Nek2 in Golgi. *Exp. Cell. Res.* 292:393-402

Zhao Y, Kappes B, Franklin RM. (1994) *Plasmodium falciparum* calcium-dependent protein kinase phosphorylates proteins of the host erythrocyte membrane. *Mol. Bioch. Parasitol.* 66: 329-43

Zhang CX, Engqvist-Goldstein AE, Carreno S, Owen DJ, Smythe E, Drubin DG. (2005) Multiple roles for cyclin G associated kinase in clathrin-mediated sorting events. *Traffic.* 6(12):1103-13

Zhang L, Gjoerup O, Roberts TM. (2004) The serine/threonine kinase cyclin G-associated kinase regulates epidermal growth factor receptor signalling. *Proc. Natl. Acad. Sci. USA.* 101 (28) : 10296-10301

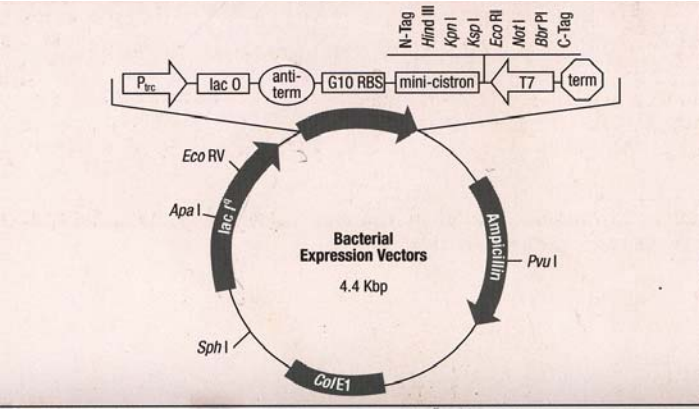
Zheng L and Lee WH. (2001) The retinoblastoma gene: a prototypic and multifunctional tumor suppressor. *Exp. Cell. Res.* 264:2-18

Zubay G. (1993) *Biochemistry*. WM. C. Brown Communications, Inc. Dubuque, USA

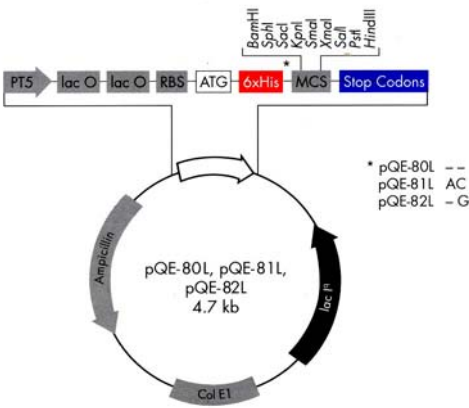
Appendix

Appendix A1 Vector maps

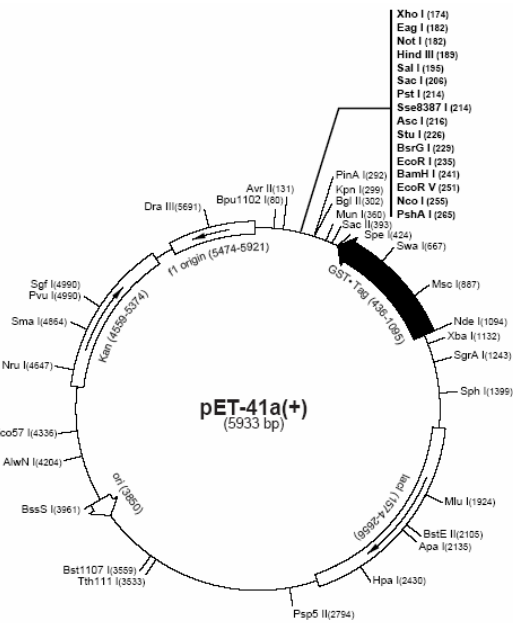
Vector map of pHB6



Vector map of pQ80L



Vector map of pET41a



Appendix A2 Solutions

Luria bertani broth :

Per liter:

To 900 ml of water add:

Bacto-tryptone	10 g
Yeast extract	5 g
NaCl	10 g

Mix until solutes are dissolved, adjust the pH to 7 with 1 M NaOH and adjust the volume to 1 liter with H₂O. Sterilize by autoclaving for 20 minutes.

2YT broth:

Per liter:

To 900 ml of water add:

Bacto-tryptone	16 g
Yeast extract	10 g
NaCl	5 g

Mix until solutes are dissolved, adjust the pH to 7 with 1 M NaOH and adjust the volume to 1 liter with H₂O. Sterilize by autoclaving for 20 minutes.

Appendix B

Appendix B1 PCR primers and PCR optimisation

Appendix B1.1 Primers for PCR

B1.1.1 Primers for cloning into the pHB6 vector

Nucleotide sequences in blue are vector sequence. Restriction sites for *Hind* III and *Eco*R I are underlined.

B1.1.1.1 PfcGAK primers

Sense strand primer

Amino acid sequence: **F R S I C T T L**

Primer sequence: 5' GCT GGA AGC TTG TTT CGA AGT ATT TGT ACT ACT TTA 3'

Hind III

Antisense strand primer

Amino acid sequence: **I S E N I Q V Q V***

Coding sequence: 5'ATA TCA GAA AAT ATT CAA GTT CAG GTG AAT TCG CGG 3'

Complementary

sequence: 3' TAT AGT CTT TTA TAA GTT CAA GTC CAC TTA AGC GCC 5'

Primer sequence: 5' CCG CGA ATT CAC CTG AAC TTG AAT ATT TTC TGA TAT 3'
EcoRI

Note: *The codon GTA was replaced by GTG that also codes for valine in the 3' valine residue.

Expected size of amplicon : 1629 bp

B1.1.1.2 Pfnek-3 primers

Sense strand primer

Amino acid sequence:

Primer sequence: 5' GCT GGA AGC TTG GTT TGC ATT TAC TTG TTT TGT TTT 3'
Hind III

Antisense strand primer

Amino acid sequence:

Coding sequence: 5' AAA ATA ATA TCA TGT ATA ACG GTT CAG AAT TCG CGG 3'
K I I S C I T V Q*

Complementary

sequence: 3' TTT TAT TAT AGT ACA TAT TGC CAA GTC TTA AGC GCC 5'

Primer sequence: 5' CCG CGA ATT CTG AAC CGT TAT ACA TGA TAT TAT TTT 3'
EcoRI

Note: The glutamine (Q) codon CAG was used in place of CAA to accommodate the *EcoRI* recognition sequence.

Expected size of amplicon : 1059 bp

B1.1.2 Primer design for cloning into pQ80L and pET41a vectors

The primer sequences were identical to those used for the pHB6 vector with the exception of the restriction sites. Since the pQ80L and pET41a vectors have the same restriction sites the same primer sets were used to generate amplicons for insertion into both vectors .

B1.1.2.1 PfcGAK primers

Sense strand primer

Amino acid sequence:

Primer sequence: 5' CAT CAC GGA TCC TTT CGA AGT ATT TGT ACT ACT TTA 3'
Bam HI

Antisense strand primer

Amino acid sequence:

Coding sequence: 5' ATA TCA GAA AAT ATT CAA GTT CAG GTG AAG CTT AAT TAG 3'
I S E N I Q V Q V

Complementary

sequence: 3' TAT AGT CTT TTA TAA GTT CAA GTC CAC TTC GAA TTA ATC 5'

Primer sequence: 5' CTA ATT AAG CTT CAC CTG AAC TTG AAT ATT TTC TGA TAT 3'
Hind III

Expected size of amplicon: 1629 bp

B1.1.2.2 Pfnek-3 primers

Sense strand primer

Amino acid sequence:

Primer sequence: 5' CAT CAC GGA TCC GTT TGC ATT TAC TTG TTT TGT TTT 3'
Bam HI

Antisense strand primer

Amino acid sequence:

K I I S C I T V

Coding sequence: 5' AAA ATA ATA TCA TGT ATA ACG GTT AAG CTT AAT TAG 3'

Complementary

sequence: 3' TTT TAT TAT AGT ACA TAT TGC CAA TTC GAA TTA ATC 5'

Primer sequence: 5' CTA ATT AAG CTT AAC CGT TAT ACA TGA TAT TAT TTT 3'
Hind III

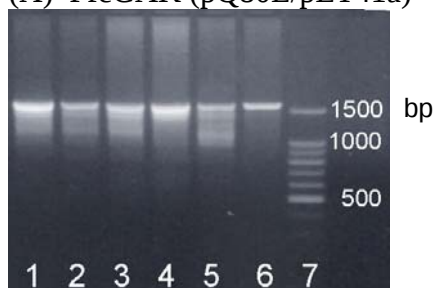
Expected size of amplicon : 1062 bp

Appendix B1.2 Optimisation of PCR conditions

B1.2.1. Annealing temperature

B1.2.1.1 The effect of annealing temperature on the PCR of PfcGAK and Pfn3 using the pQ80L and pET41a primers.

(A) PfcGAK (pQ80L/pET41a)



Legend:

1 = 55°C

2 = 58°C

3 = 60°C

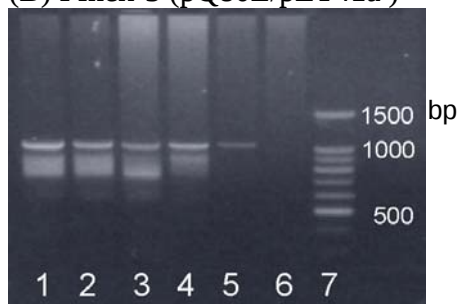
4 = 62°C

5 = 64°C

6 = 66°C

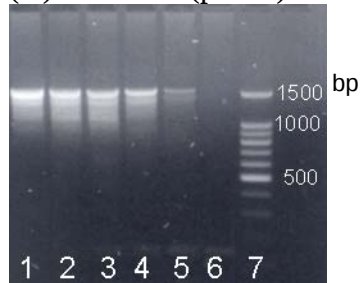
7 = 100 bp DNA size marker

(B) Pfn3 (pQ80L/pET41a)



B1.2.1.2 The effect of annealing temperature on the PCR of PfcGAK and Pfn3 using the pHB6 primers

(A) PfcGAK (pHB6)



Legend

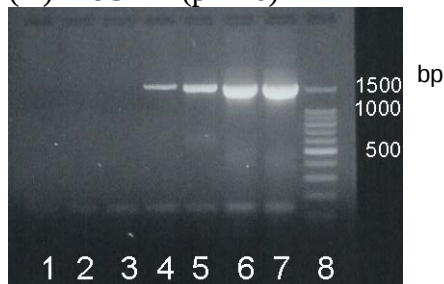
- 1 = 55°C
- 2 = 58°C
- 3 = 60°C
- 4 = 62°C
- 5 = 64°C
- 6 = 66°C
- 7 = 100 bp DNA size marker

(B) Pfn3 (pHB6)



B1.2.2 The effect of various MgCl₂ concentrations

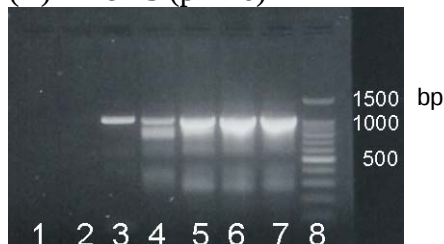
(A) PfcGAK (pHB6)



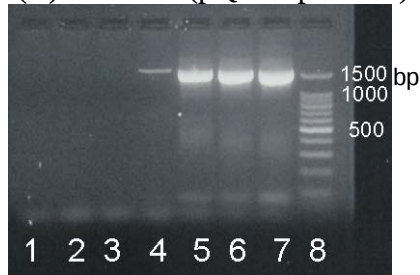
Legend:

- 1 = 1.0 mM MgCl₂
- 2 = 1.5 mM MgCl₂
- 3 = 2.0 mM MgCl₂
- 4 = 2.5 mM MgCl₂
- 5 = 3.0 mM MgCl₂
- 6 = 3.5 mM MgCl₂
- 7 = 4.0 mM MgCl₂
- 8 = 100bp DNA size marker

(B) Pfn3 (pHB6)



(C) PfcGAK (pQ80L/pET41a)



Legend:

- 1 = 1.0 mM MgCl₂
- 2 = 1.5 mM MgCl₂
- 3 = 2.0 mM MgCl₂
- 4 = 2.5 mM MgCl₂
- 5 = 3.0 mM MgCl₂
- 6 = 3.5 mM MgCl₂
- 7 = 4.0 mM MgCl₂
- 8 = 100bp DNA size marker

(D) Pfnk-3 (pQ80L/pET41a)



Appendix B2 The full-length ORF of the cloned PfcGAK construct showing the primers used for full-length sequencing

```

ATGGGTTACC CATAACGACGT CCCAGACTAC GCTGGAAGCT TGTTTCGAAG TATTGTACT ACTTTAA TCG
GTGGCAAGGT TTACAATATT AATGGGAAGA CGATAAGAGA AGAGAAATTA ATTTCTGAGG GAGCGTATTC
ATTTGTTTAT ATGGCGAAGG ATTTAAACAC GAATCGAGCT TATACTATAA AGAAGACCAT ATGTCAAGAT
AAAGAGAAAA TAGATATGGC GAATAAAGAA ATCAACATTG TAAAAAGTTT ACCTCCACAT AAAAATATAG
TACAATATTA TGGTTCCACT ATTATTTCTG AGAATAATTA TAAAATTGTT ATTATGTTGA TGGAGTATTG
TGAGAGAGGA AATTTATTAA ATATTTTCTA AAAAAATAAA GACAAGATTA AAGAAATTCA CATTGTAAAT
ATTTTAAAG ATATAATTAC CGGTTTATGT TTTTACATA ATCAAGAAAT CCCTATTATA CATCGTGATA
TTAAATTAGA AAAATTTTAA TGTGATAAAA ATGGTGTATA TAAAATT TGT GACTTCTGTT CACATTCTAT
ATCAAATTCT TTCTTTCCAA ATGATTTAAG TAAATCAGAA TTATCTAATT TGAAAGAAGA AATTGAGAGA
GATACCACAT TAATA TATAG ACCTCCCGAA TTAATTGATT TATATTCTAA TTTTGAAATA TCTTGAAAA
TTGATATATG GATGGTTGGT TGTATTTTAT ATCTACTTCT TTTTAAATC CATCCATTTC AAAATAATAA
TTTTTATCC ATTATTAATG CTAGTTTTAC TATTCCATAT TGTACGAAAT ATTCCAAAAG AATTATTTCT
ATTCTTCTGA TGA CTCTTAA TAAAAATCCT CAAAAAGAA TTGATACCTC TACTCTTTTA TTTATCCTTG
AAAATTATGC AGACCTCAAA AAATGGTTTG TG CATATTCC CGCAGATATC AAAAGAGGG TTAACCAAT
TTTTGAAAAA ATGAATGAAC TTAACATGAA TAATAACAC ACTGATCTTA TCGAAGTTAG TCATGAAAAAG
CTGATCGATG TAAAAATATC TTCACAAAAA TATGAAACCA ACCTTCCGAA GAGTAATAAT AATTTCTCA
AGTTTTTTAC AACCAGTAAT TTGAGTGGTG ATATTTTAA GAAAAAATC GAAAAGGAGG AAACATAAAA
AGAAATCAAG CAAGTAGGCA ACAAGCAAGT GAGCGAAACA GAAAAGGAAC CAAAGGATAA ATTGGAAAAAT
AATCCAGATG GTAATAAATT ACAAATGAT AAAGATAGTG GTAAATTACA AAATGATAAA GATAGTGGTA
AATTACAAAA TGAAAAAGAT AGTGGTAAAT TACAAAATGA TAAAGATAGT GGTAAATTAC AAAATGATAA
AGATAGCGGT AAATTACAAA ATGATAAAGA TAGTGGTAAA TTACAAAATG ATAAAGTTAG TGGTAAATTA
CAAAATGATA AAGATAGTGG TAAATTACAA AATGATAAAG TTAGTGGTAA ATTACAAAAT GATCGAGATG
GTAATAAATT ACAAATAAT CATATTATTG ATTATACAGA TAATAAATG GATAACGATA AAAAGAATAA
ATCTGATGAA ATATCAGAAA ATATTACAAG TTCAGGTGAA TTCGCGGCCG CTACACGTGC ATCATCATCA
TCATCATTA

```

Red = vector sequence

Plum = stop codon

Lime green = sense and antisense primer

Aqua = internal forward primer 1 sequence

Pink = internal forward primer 2 sequence

Blue = internal reverse primer

Appendix B3 Blastp analysis of PfcGAK - summary table

	DB:ID	Source	Length	% Id	% Pos	E ()
1	UNIPROT:Q8I4V7_PLAF7	Cyclin g-associated kinase, putative.	909	100	100	0.
2	UNIPROT:Q4YQ17_PLABE	Cyclin g-associated kinase, putative (Fragment).	299	80	87	4.3e-110
3	UNIPROT:Q4XUV7_PLACH	Cyclin g-associated kinase, putative (Fragment).	284	57	70	6.0e-79
4	UNIPROT:Q86HW6_DICDI	Hypothetical protein DDB0229347.	798	25	44	9.4e-57
5	UNIPROT:Q54VV7_DICDI	Hypothetical protein DDB0229350.	1126	35	60	9.9e-55
6	UNIPROT:Q2KJH2_BOVIN	Hypothetical protein.	1268	35	57	7.3e-39
7	UNIPROT:Q9VMY8_DROME	CG1107-PA, isoform A (Cg1107-pb, isoform b).	1165	34	55	1.0e-38
8	UNIPROT:Q5RJ35_BRARE	Novel protein similar to vertebrate cyclin G associated kinase (GAK) (Fragment).	845	34	58	1.2e-38
9	UNIPROT:Q53UK1_LOTJA	Ser/Thr protein kinase.	307	32	55	1.5e-38
10	UNIPROT:Q7QG29_ANOGA	ENSANGP00000005528 (Fragment).	1158	32	57	1.2e-37
11	UNIPROT:Q8VYC0_ARATH	Putative serine/threonine protein kinase.	650	31	55	1.6e-37
12	UNIPROT:O50071_ARATH	Putative serine/threonine protein kinase (Putative serine threonine protein kinase) (Putative serine/threonine kinase).	650	31	55	2.1e-37
13	UNIPROT:Q8H0Z0_ARATH	Putative serine/threonine protein kinase.	670	31	55	3.4e-37
14	UNIPROT:Q6H5H4_ORYSA	Putative AAK1 protein.	703	31	52	6.8e-37
15	UNIPROT:Q5RDC0_PONPY	Hypothetical protein DKFZp469F1313.	1311	34	57	8.6e-37
16	UNIPROT:GAK_HUMAN	Cyclin G-associated kinase (EC 2.7.1.-).	1311	34	57	1.1e-36
17	UNIPROT:Q3TY53_MOUSE	Visual cortex cDNA, RIKEN full-length enriched library, clone:K430355L13 product:similar to BMP-2 inducible protein kinase (EC 2.7.1.37) (BIKe).	352	32	58	1.7e-36
18	UNIPROT:Q69L75_ORYSA	Putative AAK1 protein.	308	32	53	2.2e-36
19	UNIPROT:Q9BUD9_HUMAN	AAK1 protein.	474	32	58	2.8e-36
20	UNIPROT:GAK_RAT	Cyclin G-associated kinase (EC 2.7.1.-).	1305	34	57	5.0e-36
21	UNIPROT:Q99KY4-2	(Q99KY4) Splice isoform 2 of Q99KY4	1256	34	57	9.2e-36
22	UNIPROT:Q3UIZ0_MOUSE	17 days pregnant adult female amnion cDNA, RIKEN full-length enriched library, clone:I920043D18 product:cyclin G associated kinase, full insert sequence.	1305	34	57	1.0e-35
23	UNIPROT:GAK_MOUSE	Cyclin G-associated kinase (EC 2.7.1.-).	1305	34	57	1.1e-35
24	UNIPROT:BMP2K_HUMAN	BMP-2-inducible protein kinase (EC 2.7.1.37) (BIKe).	1161	33	55	1.5e-35
25	UNIPROT:Q69L76_ORYSA	Cyclin G-associated kinase-like protein.	667	32	53	1.8e-35
26	UNIPROT:Q4W5H2_HUMAN	Hypothetical protein BMP2K.	662	33	55	3.7e-35

27	UNIPROT:Q9NSY1-2	(Q9NSY1) Splice isoform 2 of Q9NSY1	662	33	55	3.7e-35
28	UNIPROT:Q9NSY1-3	(Q9NSY1) Splice isoform 3 of Q9NSY1	677	33	55	4.3e-35
29	UNIPROT:BMP2K_MOUSE	BMP-2-inducible protein kinase (EC 2.7.1.37) (BIKe).	1138	33	55	1.0e-34
30	UNIPROT:AKL1_YEAST	Serine/threonine-protein kinase AKL1 (EC 2.7.1.37).	1108	27	49	1.1e-34
31	UNIPROT:Q91Z96-2	(Q91Z96) Splice isoform 2 of Q91Z96	649	33	55	1.6e-34
32	UNIPROT:Q4SJ98_TETNG	Chromosome 4 SCAF14575, whole genome shotgun sequence.	603	32	58	2.3e-34
33	UNIPROT:Q5PRC8_BRARE	Zgc:101641.	646	33	58	2.6e-34
34	UNIPROT:Q3UHJ0_MOUSE	CDNA, RIKEN full-length enriched library, clone:M5C1020L09 product:weakly similar to BMP-2 inducible protein kinase (EC 2.7.1.37) (BIKe) (HRIHFB2017).	959	32	59	5.3e-34
35	UNIPROT:Q6BMC6_DEBHA	Debaryomyces hansenii chromosome F of strain CBS767 of Debaryomyces hansenii.	1003	24	44	7.6e-34
36	UNIPROT:Q2M2I8_HUMAN	AP2 associated kinase 1.	863	32	58	8.6e-34
37	UNIPROT:Q9UPV4_HUMAN	KIAA1048 protein (Fragment).	897	32	58	9.6e-34
38	UNIPROT:Q5A5P2_CANAL	Likely protein kinase.	734	32	54	1.4e-33
39	UNIPROT:Q8MSQ7_DROME	SD07306p.	698	34	56	5.2e-33
40	UNIPROT:Q7KT38_DROME	CG10637-PC, isoform C.	698	34	56	5.2e-33
41	UNIPROT:Q6FPJ1_CANGA	Candida glabrata strain CBS138 chromosome J complete sequence.	622	24	44	7.7e-33
42	UNIPROT:Q6FRA3_CANGA	Candida glabrata strain CBS138 chromosome H complete sequence.	1027	31	53	2.1e-32
43	UNIPROT:Q9VJ30_DROME	CG10637-PA, isoform A.	1488	34	56	3.2e-32
44	UNIPROT:PRK1_YEAST	Actin-regulating kinase PRK1 (EC 2.7.1.37) (p53-regulating kinase 1).	810	30	54	3.4e-32
45	UNIPROT:Q50ZA0_ENTHI	Protein kinase, putative.	519	29	53	1.6e-31
46	UNIPROT:ARK1_YEAST	Actin-regulating kinase 1 (EC 2.7.1.37).	638	22	45	2.7e-31
47	UNIPROT:Q9U485_DROME	Numb-associated kinase.	1490	34	56	4.2e-31
48	UNIPROT:Q50ZL7_ENTHI	Protein kinase, putative.	552	25	44	5.5e-31
49	UNIPROT:Q2U6V2_ASPOR	ARK protein kinase family.	996	31	54	6.2e-31
50	UNIPROT:O43066_SCHPO	SPBC6B1.02 protein.	953	31	51	1.3e-30

Appendix B4 Blastp analysis of Pfnek-3 - summary table

	DB:ID	Source	Length	% Id	% Pos	E()
1	UNIPROT:Q8I629_PLAF7	Serine/threonine-protein kinase Nek1, putative.	347	100	100	9.4e-184
2	UNIPROT:Q4YS42_PLABE	Serine/threonine-protein kinase Nek1, putative (Fragment).	268	76	89	2.4e-107
3	UNIPROT:Q50MA8_ENTHI	Protein kinase, putative.	484	31	51	2.6e-26
4	UNIPROT:Q55BN8_DICDI	Protein serine/threonine kinase.	418	35	55	9.4e-26
5	UNIPROT:Q4Z3M6_PLABE	Serine/threonine protein kinase 2, putative.	310	32	53	1.2e-25

6	UNIPROT:Q4R7J4_MACFA	Testis cDNA, clone: QtsA-15092, similar to human NIMA (never in mitosis gene a)- related kinase 9 (NEK9), mRNA, RefSeq: NM_033116.3.	532	33	57	1.3e-25
7	UNIPROT:NEK9_XENLA	Serine/threonine-protein kinase Nek9 (EC 2.7.1.37) (NimA-related protein kinase 9) (XNek9) (Nercc1 kinase).	944	30	58	3.4e-25
8	UNIPROT:Q6DKD3_XENLA	LOC398600 protein.	944	30	58	3.4e-25
9	UNIPROT:NEK9_HUMAN	Serine/threonine-protein kinase Nek9 (EC 2.7.1.37) (NimA-related protein kinase 9) (Nercc1 kinase) (NIMA-related kinase 8) (Nek8).	979	33	57	6.0e-25
10	UNIPROT:Q52LK6_HUMAN	NIMA related kinase 9.	979	33	57	6.0e-25
11	UNIPROT:NEK9_MOUSE	Serine/threonine-protein kinase Nek9 (EC 2.7.1.37) (NimA-related protein kinase 9).	984	30	55	6.1e-25
12	UNIPROT:Q69Z43_MOUSE	MKIAA1995 protein (Fragment).	1005	30	55	6.3e-25
13	UNIPROT:Q4D7E4_TRYCR	Protein kinase, putative.	596	29	55	1.4e-23
14	UNIPROT:Q4DSP2_TRYCR	Protein kinase, putative.	598	29	55	1.4e-23
15	UNIPROT:Q4RLW8_TETNG	Chromosome 10 SCAF15019, whole genome shotgun sequence.	855	30	57	4.0e-23
16	UNIPROT:Q7R1M1_GIALA	GLP_28_62487_61384.	367	29	52	1.1e-22
17	UNIPROT:Q57WH3_9TRYP	Protein kinase, putative (EC 2.7.1.37).	596	28	55	1.3e-22
18	UNIPROT:Q8BSB6_MOUSE	12 days embryo embryonic body between diaphragm region and neck cDNA, RIKEN full-length enriched library, clone:9430031M01 product:NIMA (never in mitosis gene a)-related expressed kinase 1, full insert sequence.	302	29	52	2.3e-22
19	UNIPROT:Q7QUR6_GIALA	GLP_231_43409_44626.	405	29	53	3.8e-22
20	UNIPROT:Q8CD72_MOUSE	13 days embryo male testis cDNA, RIKEN full-length enriched library, clone:6030407P11 product:NIMA (never in mitosis gene a)-related expressed kinase 1, full insert sequence.	424	29	52	4.2e-22
21	UNIPROT:Q54HS0_DICDI	Hypothetical protein (Fragment).	498	31	52	7.2e-22
22	UNIPROT:Q86I06_DICDI	Similar to Mus musculus (Mouse). 13 days embryo male testis cDNA, RIKEN full-length enriched library, clone:6030407P11 product:NIMA (Never in mitosis gene a)-related expressed kinase 1, full insert sequence (Hypothetical protein DDB0220003).	1123	32	52	7.4e-22
23	UNIPROT:Q7R6C5_GIALA	GLP_574_51801_52931.	376	28	51	1.0e-21
24	UNIPROT:Q4YRJ3_PLABE	Serine/threonine-protein kinase Nek1, putative.	293	29	50	2.6e-21
25	UNIPROT:NEK1_MOUSE	Serine/threonine-protein kinase Nek1 (EC 2.7.1.37) (NimA-related protein kinase 1).	1203	29	52	5.8e-21
26	UNIPROT:Q69Z74_MOUSE	MKIAA1901 protein (Fragment).	1234	29	52	6.0e-21
27	UNIPROT:Q4QC33_LEIMA	Serine/threonine-protein kinase, putative.	801	29	50	8.1e-21
28	UNIPROT:Q5RFZ0_BRARE	Novel protein (Zgc:56141) (Fragment).	344	29	51	1.1e-20
29	UNIPROT:Q4QFJ2_LEIMA	Protein kinase, putative.	555	28	55	2.1e-20
30	UNIPROT:Q6P3R8_HUMAN	NIMA (Never in mitosis gene a)-related kinase 5.	708	30	53	2.2e-20
31	UNIPROT:Q91ZR4-2	(Q91ZR4) Splice isoform 2 of Q91ZR4	291	29	51	3.9e-20
32	UNIPROT:Q96PY6-2	(Q96PY6) Splice isoform 2 of Q96PY6	1214	28	52	4.2e-20

33	UNIPROT:Q19530_CAEEL	Hypothetical protein.	302	31	51	4.4e-20
34	UNIPROT:NEK1_HUMAN	Serine/threonine-protein kinase Nek1 (EC 2.7.1.37) (NimA-related protein kinase 1) (NY-REN-55 antigen).	1258	28	52	4.4e-20
35	UNIPROT:Q5H9T1_HUMAN	Hypothetical protein DKFZp686D06121.	1287	28	52	4.6e-20
36	UNIPROT:Q659I9_SPHGR	Nek protein.	893	28	49	5.5e-20
37	UNIPROT:Q4CVU2_TRYCR	Protein kinase, putative (EC 2.7.1.37).	763	30	51	6.9e-20
38	UNIPROT:Q7QQZ8_GIALA	GLP_442_9173_11869.	898	29	51	9.1e-20
39	UNIPROT:Q7ZUN2_BRARE	NIMA (Never in mitosis gene a)-related kinase 2 (Zgc:55602).	440	31	53	1.0e-19
40	UNIPROT:Q8N0P1_PARCA	Serine/threonine protein kinase 2 homolog.	356	34	51	1.2e-19
41	UNIPROT:Q4CXF6_TRYCR	Serine/threonine-protein kinase A, putative.	504	29	51	2.8e-19
42	UNIPROT:Q5M813_RAT	Nek4 protein.	309	27	50	3.5e-19
43	UNIPROT:Q3THY0_MOUSE	TIB-55 BB88 cDNA, RIKEN full-length enriched library, clone:1730057F21 product:NIMA (never in mitosis gene a)-related expressed kinase 4, full insert sequence.	274	27	50	4.4e-19
44	UNIPROT:Q54UZ1_DICDI	Hypothetical protein DDB0216386.	411	36	50	6.7e-19
45	UNIPROT:Q5RIE9_BRARE	Novel protein similar to vertebrate NIMA (Never in mitosis gene a)- related kinase 1 (NEK1).	1353	29	51	1.7e-18
46	UNIPROT:Q8I9T8_TETTH	Protein tyrosine kinase 1.	549	26	50	1.9e-18
47	UNIPROT:Q38CJ2_9TRYP	Protein kinase, putative (EC 2.7.1.37).	764	28	49	3.9e-18
48	UNIPROT:Q2IAZ0_9TRYP	NIMA-related kinase protein NRKC.	764	28	49	3.9e-18
49	UNIPROT:Q4Q5W2_LEIMA	Serine/threonine-protein kinase, putative (Protein kinase, putative) (EC 2.7.1.37).	441	28	51	4.3e-18
50	UNIPROT:Q6DII1_XENTR	Hypothetical protein.	442	30	49	4.3e-18

Appendix B5 Nucleic acid sequence alignment of PfcGAK inserts

Alignment of the nucleic acid sequences of PfcGAK, with the PfcGAK constructs of clone 24 and clone 14.

PfcGAK24	1	ATGGGTTACCCATACGACGTCCAGACTACGCTGGAA	GCTTG	TTTCGAAGT	51
PfcGAK	1		ATGTTGAAA	-----TTTCGAAGT	18
PfcGAK14	1	ATGGGTTACCCATACGACGTCCAGACTACGCTGGAA	GCTTG	TTTCGAAGT	51
PfcGAK24	52	ATTTGTACTACTTTAATCGGTGGCAAGGTTTACAATATTAATGGGAAGACG			102
PfcGAK	19	ATTTGTACTACTTTAATCGGTGGCAAGGTTTACAATATTAATGGGAAGACG			69
PfcGAK14	52	ATTTGTACTACTTTAATCGGTGGCAAGGTTTACAATATTAATGGGAAGACG			102
PfcGAK24	103	ATAAGAGAAGAGAAATTAATTTCTGAGGGAGCGTATTCATTTGTTTATATG			153
PfcGAK	70	ATAAGAGAAGAGAAATTAATTTCTGAGGGAGCGTATTCATTTGTTTATATG			120
PfcGAK14	103	ATAAGAGAAGAGAAATTAATTTCTGAGGGAGCGTATTCATTTGTTTATATG			153
PfcGAK24	154	GCGAAGGATTTAAACACGAATCGAGCTTATACTATAAAGAAGACCATATGT			204
PfcGAK	121	GCGAAGGATTTAAACACGAATCGAGCTTATACTATAAAGAAGACCATATGT			171
PfcGAK14	154	GCGAAGGATTTAAACACGAATCGAGCTTATACTATAAAGAAGACCATATGT			204

PfcGAK24	205	CAAGATAAAGAGAAAAATAGATATGGCGAATAAAGAAATCAACATTCTTAAA	255
PfcGAK	172	CAAGATAAAGAGAAAAATAGATATGGCGAATAAAGAAATCAACATTCTTAAA	222
PfcGAK14	205	CAAGATAAAGAGAAAAATAGATATGGCGAATAAAGAAATCAACATTCTTAAA	255
PfcGAK24	256	AGTTTACCTCCACATAAAAAATATAGTACAATATTATGGTTCCACTATTATT	306
PfcGAK	223	AGTTTACCTCCACATAAAAAATATAGTACAATATTATGGTTCCACTATTATT	273
PfcGAK14	256	AGTTTACCTCCACATAAAAAATATAGTACAATATTATGGTTCCACTATTATT	306
PfcGAK24	307	TCTGAGAATAATTATAAAAAATTGTTATTATGTTGATGGAGTATTGTGAGAGA	357
PfcGAK	274	TCTGAGAATAATTATAAAAAATTGTTATTATGTTGATGGAGTATTGTGAGAGA	324
PfcGAK14	307	TCTGAGAATAATTATAAAAAATTGTTATTATGTTGATGGAGTATTGTGAGAGA	357
PfcGAK24	358	GGAAATTTATTAAATATTTTTCAAAAAATAAAGACAAGATTAAAGAAATT	408
PfcGAK	325	GGAAATTTATTAAATATTTTTCAAAAAATAAAGACAAGATTAAAGAAATT	375
PfcGAK14	358	GGAAATTTATTAAATATTTTTCAAAAAATAAAGACAAGATTAAAGAAATT	408
PfcGAK24	409	CACATTGTAAATATTTTAAAAGATATAATTACCGGTTTATGTTTTTTACAT	459
PfcGAK	376	CACATTGTAAATATTTTAAAAGATATAATTACCGGTTTATGTTTTTTACAT	426
PfcGAK14	409	CACATTGTAAATATTTTAAAAGATATAATTACCGGTTTATGTTTTTTACAT	459
PfcGAK24	460	AATCAAGAAATCCCTATTATACATCGTGATATTAAATTAGAAAATATTTTA	510
PfcGAK	427	AATCAAGAAATCCCTATTATACATCGTGATATTAAATTAGAAAATATTTTA	477
PfcGAK14	460	AATCAAGAAATCCCTATTATACATCGTGATATTAAATTAGAAAATATTTTA	510
PfcGAK24	511	TGTGATAAAAAATGGTGTATATAAAATTTGTGACTTCTGTTCCACATTCTATA	561
PfcGAK	478	TGTGATAAAAAATGGTGTATATAAAATTTGTGACTTCTGTTCCACATTCTATA	528
PfcGAK14	511	TGTGATAAAAAATGGTGTATATAAAATTTGTGACTTCTGTTCCACATTCTATA	561
PfcGAK24	562	TCAAATTCTTTCTTTCCAAATGATTTAAGTAAATCAGAATTATCTAATTTG	612
PfcGAK	529	TCAAATTCTTTCTTTCCAAATGATTTAAGTAAATCAGAATTATCTAATTTG	579
PfcGAK14	562	TCAAATTCTTTCTTTCCAAATGATTTAAGTAAATCAGAATTATCTAATTTG	612
PfcGAK24	613	AAAGAAGAAATTGAGAGAGATACCACATTAATATATAGACCTCCCGAATTA	663
PfcGAK	580	AAAGAAGAAATTGAGAGAGATACCACATTAATATATAGACCTCCCGAATTA	630
PfcGAK14	613	AAAGAAGAAATTGAGAGAGATACCACATTAATATATAGACCTCCCGAATTA	663
PfcGAK24	664	ATTGATTTATATTCTAATTTTGAAATATCTTGAAAATTGATATATGGATG	714
PfcGAK	631	ATTGATTTATATTCTAATTTTGAAATATCTTGAAAATTGATATATGGATG	681
PfcGAK14	664	ATTGATTTATATTCTAATTTTGAAATATCTTGAAAATTGATATATGGATG	714
PfcGAK24	715	GTTGGTTGTATTTTATATCTACTTCTTTTAAAATCCATCCATTTCAAAAAT	765
PfcGAK	682	GTTGGTTGTATTTTATATCTACTTCTTTTAAAATCCATCCATTTCAAAAAT	732
PfcGAK14	715	GTTGGTTGTATTTTATATCTACTTCTTTTAAAATCCATCCATTTCAAAAAT	765
PfcGAK24	766	AATAATTTTTTATCCATTATTAATGCTAGTTTTACTATTCCATATTGTACG	816
PfcGAK	733	AATAATTTTTTATCCATTATTAATGCTAGTTTTACTATTCCATATTGTACG	783
PfcGAK14	766	AATAATTTTTTATCCATTATTAATGCTAGTTTTACTATTCCATATTGTACG	816
PfcGAK24	817	AAATATTCCAAAAGAATTATTTCTATTCTTCTGATGACTCTTAATAAAAAAT	867
PfcGAK	784	AAATATTCCAAAAGAATTATTTCTATTCTTCTGATGACTCTTAATAAAAAAT	834
PfcGAK14	817	AAATATTCCAAAAGAATTATTTCTATTCTTCTGATGACTCTTAATAAAAAAT	867
PfcGAK24	868	CCTCAAAAAAGAATTGATACCTCTACTCTTTTATTTATCCTTGAAAATTAT	918
PfcGAK	835	CCTCAAAAAAGAATTGATACCTCTACTCTTTTATTTATCCTTGAAAATTAT	885
PfcGAK14	868	CCTCAAAAAAGAATTGATACCTCTACTCTTTTATTTATCCTTGAAAATTAT	918
PfcGAK24	919	GCAGACCTCAAAAAATGGTTTGTGCATATCCCAGATATCAAAAAGAGG	969
PfcGAK	886	GCAGACCTCAAAAAATGGTTTGTGCATATCCCAGATATCAAAAAGAGG	936
PfcGAK14	919	GCAGACCTCAAAAAATGGTTTGTGCATATCCCAGATATCAAAAAGAGG	969
PfcGAK24	970	GTTAACCAAATTTTTGAAAAAATGAATGAACCTAACATGAATAATAAACAC	1020
PfcGAK	937	GTTAACCAAATTTTTGAAAAAATGAATGAACCTAACATGAATAATAAACAC	987
PfcGAK14	970	GTTAACCAAATTTTTGAAAAAATGAATGAACCTAACATGAATAATAAACAC	1020
PfcGAK24	1021	ACTGATCTTATCGAAGTTAGTCATGAAAAGCTGATCGATGTTAAAATATCT	1071
PfcGAK	988	ACTGATCTTATCGAAGTTAGTCATGAAAAGCTGATCGATGTTAAAATATCT	1038
PfcGAK14	1021	ACTGATCTTATCGAAGTTAGTCATGAAAAGCTGATCGATGTTAAAATATCT	1071

PfcGAK24	1072	TCACAAAAATATGAAACCAACCTTCCGAAGAGTAATAATAATTTCTCAAG	1122
PfcGAK	1039	TCACAAAAATATGAAACCAACCTTCCGAAGAGTAATAATAATTTCTCAAG	1089
PfcGAK14	1072	TCACAAAAATATGAAACCAACCTTCCGAAGAGTAATAATAATTTCTCAAG	1122
PfcGAK24	1123	TTTTTTACAACCAAGTAATTTGAGTGGTGATATTTTTAAGAAAAAAATCGAA	1173
PfcGAK	1090	TTTTTTACAACCAAGTAATTTGAGTGGTGATATTTTTAAGAAAAAAATCGAA	1140
PfcGAK14	1123	TTTTTTACAACCAAGTAATTTGAGTGGTGATATTTTTAAGAAAAAAATCGAA	1173
PfcGAK24	1174	AAGGAGGAAACTAAAAAGAAATCAAGCAAGTAGGCAACAAGCAAGTGAGC	1224
PfcGAK	1141	AAGGAGGAAACTAAAAAGAAATCAAGCAAGTAGGCAACAAGCAAGTGAGC	1191
PfcGAK14	1174	AAGGAGGAAACTAAAAAGAAATCAAGCAAGTAGGCAACAAGCAAGTGAGC	1224
PfcGAK24	1225	GAAACAGAAAAGGAACCAAAGGATAAATTGGAAAATAATCCAGATGGTAAT	1275
PfcGAK	1192	GAAACAGAAAAGGAACCAAAGGATAAATTGGAAAATAATCCAGATGGTAAT	1242
PfcGAK14	1225	GAAACAGAAAAGGAACCAAAGGATAAATTGGAAAATAATCCAGATGGTAAT	1275
PfcGAK24	1276	AAATTACAAAATGAT	1291
PfcGAK	1243	AAATTACAAAATGATAAAGATAGTGGTAAATTACAAAATGATAAAGATAGT	1293
PfcGAK14	1276	AAATTACAAAATGATAAAGATAGTGGTAAATTACAAAATGATAAAGATAGT	1326
PfcGAK24	1292	-----	1291
PfcGAK	1294	GGTAAATTACAAAATGAAAAAGATAGTGGTAAATTACAAAATGATAAAGAT	1344
PfcGAK14	1327	GGTAAATTACAAAATGAAAAAGATAGTGGTAAATTACAAAATGATAAAGAT	1377
PfcGAK24	1292	-----	1291
PfcGAK	1345	AGTGGTAAATTACAAAATGATAAAGATAGCGGTAAATTACAAAATGATAAA	1395
PfcGAK14	1378	AGTGGTAAATTACAAAATGATAAAGATAGCGGTAAATTACAAAATGATAAA	1428
PfcGAK24	1292	-----	1291
PfcGAK	1396	GATAGTGGTAAATTACAAAATGATAAAGTTAGTGGTAAATTACAAAATGAT	1446
PfcGAK14	1429	GATAGTGGTAAATTACAAAATGATAAAGTTAGTGGTAAATTACAAAATGAT	1479
PfcGAK24	1292	-----	1291
PfcGAK	1447	AAAGATAGTGGTAAATTACAAAATGATAAAGTTAGTGGTAAATTACAAAAT	1497
PfcGAK14	1480	AAAGATAGTGGTAAATTACAAAATGATAAAGTTAGTGGTAAATTACAAAAT	1530
PfcGAK24	1292	--CGAGATGGTAATAAATTACAAAATAATCATATTATTGATTATACAGAT	1339
PfcGAK	1498	GATCGAGATGGTAATAAATTACAAAATAATCATATTATTGATTATACAGAT	1548
PfcGAK14	1531	GATCGAGATGGTAATAAATTACAAAATAATCATATTATTGATTATACAGAT	1581
PfcGAK24	1340	AATATATTGATTATACAGATAATAAAATGGATAACGATAAAAAAGAATAAATCT	1390
PfcGAK	1549	AATATATTGATTATACAGATAATAAAATGGATAACGATAAAAAAGAATAAATCT	1599
PfcGAK14	1582	AATATATTGATTATACAGATAATAAAATGGATAACGATAAAAAAGAATAAATCT	1632
PfcGAK24	1391	GATGAAATATCAGAAAAATTTCAAGTTCAGGTGAATTCGCGGCCGCTACACGTG	1440
PfcGAK	1600	GATGAAATATCAGAAAAATTTACAAGTTCAGGTGATAAACAAAATTCTGTACCT	1649
PfcGAK14	1633	GATGAAATATCAGAAAAATTTCAAGTTCAGGTGAATTCGCGGCCGCTACACGTG	1683
PfcGAK24	1441	CATCATCATCATCATCATTAA	1462
PfcGAK	1650	GACCAGTCAGAAAATCAATCAAAAAGTCATTCATACAATTTAATAAAAATTCAA	1699
PfGAK14	1684	CATCATCATCATCATCATTAA	1705

Legend: Blue sequence = vector sequence, shaded sequence = identical sequence

Appendix B6 Nucleic acid sequence alignment of Pfnek-3 inserts

Pfnek-3	1	ATG-GTTTGCATTTACTTGTTTTGTTTTTATTAT	34
Pfnek-3(2)	1	CATCACCATCACCATCACGGATCCGTTTGCATTTACTTGTTTTGTTTTTATTAT	55
Pfnek-3(4)	1	CATCACCATCACCATCACGGATCCGTTTGCATTTACTTGTTTTGTTTTTATTAT	55
Pfnek-3	35	TCCATTTGTTTGTTGTTAATACTCGTATGAGGATATTGTGTAATTTGTTGATAGT	89
Pfnek-3(2)	56	TCCATTTGTTTGTTGTTAATACTCGTATGAGGATATTGTGTAATTTGTTGATAGT	110
Pfnek-3(4)	56	TCCATTTGTTTGTTGTTAATACTCGTATGAGGATATTGTGTAATTTGTTGATAGT	110

Pfnek-3	90	TATTTGTGAGTATATGTTCCCTTTTTACTCTTTTATATTGTTTAAATTTGTTAAG	144
Pfnek-3(2)	111	TATTTGTGAGTATATGTTCCCTTTTTACTCTTTTATATTGTTTAAATTTGTTAAG	165
Pfnek-3(4)	111	TATTTGTGAGTATATGTTCCCTTTTTACTCTTTTATATTGTTTAAATTTGTTAAG	165
Pfnek-3	145	ATGTTATCTTTATTGTTACCGTCAATATTATCGTGTGAGAAGAAATACCAGGTGT	199
Pfnek-3(2)	166	ATGTTATCTTTATTGTTACCGTCAATATTATCGTGTGAGAAGAAATACCAGGTGT	220
Pfnek-3(4)	166	ATGTTATCTTTATTGTTACCGTCAATATTATCGTGTGAGAAGAAATACCAGGTGT	220
Pfnek-3	200	ATAAGAATAATGGATATAAATTTGAGACCGTGTTAGATTTCATGACATCAGATAG	254
Pfnek-3(2)	221	ATAAGAATAATGGATATAAATTTGAGACCGTGTTAGATTTCATGACATCAGATAG	275
Pfnek-3(4)	221	ATAAGAATAATGGATATAAATTTGAGACCGTGTTAGATTTCATGACATCAGATAG	275
Pfnek-3	255	TGAAATTCATTTGATAAGATCTATAGAAAGTGATGAGATATATATATCTAAGGTA	309
Pfnek-3(2)	276	TGAAATTCATTTGATAAGATCTATAGAAAGTGATGAGATATATATATCTAAGGTA	330
Pfnek-3(4)	276	TGAAATTCATTTGATAAGATCTATAGAAAGTGATGAGATATATATATCTAAGGTA	330
Pfnek-3	310	TATGATATATATGGTATAAATGAAGATGATTTGAATAAATATATGAATGAATTAT	364
Pfnek-3(2)	331	TATGATATATATGGTATAAATGAAGATGATTTGAATAAATATATGAATGAATTAT	385
Pfnek-3(4)	331	TATGATATATATGGTATAAATGAAGATGATTTGAATAAATATATGAATGAATTAT	385
Pfnek-3	365	ATATAATGAATAAGTTAAGAAATTGTGAGAATATTGTAAATATAATAGATTATAT	419
Pfnek-3(2)	386	ATATAATGAATAAGTTAAGAAATTGTGAGAATATTGTAAATATAATAGATTATAT	440
Pfnek-3(4)	386	ATATAATGAATAAGTTAAGAAATTGTGAGAATATTGTAAATATAATAGATTATAT	440
Pfnek-3	420	TAAAGAAAATGATACATTATCTTTTATTCTTGAGTTTTGTAATCAAGGAGATTTA	474
Pfnek-3(2)	441	TAAAGAAAATGATACATTATCTTTTATTCTTGAGTTTTGTAATCAAGGAGATTTA	495
Pfnek-3(4)	441	TAAAGAAAATGATACATTATCTTTTATTCTTGAGTTTTGTAATCAAGGAGATTTA	495
Pfnek-3	475	CATTCAGATATCTTAAGAAAGAAATTAATAATGAAATATATACAGAAAGTGAAA	529
Pfnek-3(2)	496	CATTCAGATATCTTAAGAAAGAAATTAATAATGAAATATATACAGAAAGTGAAA	550
Pfnek-3(4)	496	CATTCAGATATCTTAAGAAAGAAATTAATAATGAAATATATACAGAAAGTGAAA	550
Pfnek-3	530	TATTTAATATATTACATCAAATATTAATGGTTTTAAATATAATTCATCAGAATGG	584
Pfnek-3(2)	551	TATTTAATATATTACATCAAATATTAATGGTTTTAAATATAATTCATCAGAATGG	605
Pfnek-3(4)	551	TATTTAATATATTACATCAAATATTAATGGTTTTAAATATAATTCATCAGAATGG	605
Pfnek-3	585	TATAATACATGGAGATTTAAAAAGTACTAATATATTTATTTAAAGATAATAAAATA	639
Pfnek-3(2)	606	TATAATACATGGAGATTTAAAAAGTACTAATATATTTATTTAAAGATAATAAAATA	660
Pfnek-3(4)	606	TATAATACATGGAGATTTAAAAAGTACTAATATATTTATTTAAAGATAATAAAATA	660
Pfnek-3	640	AAGATTGGTGATTTTCGGTATATCATCAGAACAAAGTTCAAATAATAATTTAGGAA	694
Pfnek-3(2)	661	AAGATTGGTGATTTTCGGTATATCATCAGAACAAAGTTCAAATAATAATTTAGGAA	715
Pfnek-3(4)	661	AAGATTGGTGATTTTCGGTATATCATCAGAACAAAGTTCAAATAATAATTTAGGAA	715
Pfnek-3	695	CATTAAATTGTTTAAAGTTATGAATCAATTAATTTAAGAAAACAAATAAATTAAG	749
Pfnek-3(2)	716	CATTAAATTGTTTAAAGTTATGAATCAATTAATTTAAGAAAACAAATAAATTAAG	770
Pfnek-3(4)	716	CATTAAATTGTTTAAAGTTATGAATCAATTAATTTAAGAAAACAAATAAATTAAG	770
Pfnek-3	750	TGATTTATTTCAAGTTGGATGTATTTTATATGAAC TTGCCACTTTGTCATCACCC	804
Pfnek-3(2)	771	TGATTTATTTCAAGTTGGATGTATTTTATATGAAC TTGCCACTTTGTCATCACCC	825
Pfnek-3(4)	771	TGATTTATTTCAAGTTGGATGTATTTTATATGAAC TTGCCACTTTGTCATCACCC	825
Pfnek-3	805	TTTTGTGCTACAAATATCAATGATATGATATCTCTTTTGGAGATAAAAAATTATA	859
Pfnek-3(2)	826	TTTTGTGCTACAAATATCAATGATATGATATCTCTTTTGGAGATAAAAAATTATA	880
Pfnek-3(4)	826	TTTTGTGCTACAAATATCAATGATATGATATCTCTTTTGGAGATAAAAAATTATA	880
Pfnek-3	860	AAAGCTATATTATCAAAAAATATATCTTCTATATATTCACAAAAATTAGTAAATGT	914
Pfnek-3(2)	881	AAAGCTATATTATCAAAAAATATATCTTCTATATATTCACAAAAATTAGTAAATGT	935
Pfnek-3(4)	881	AAAGCTATATTATCAAAAAATATATCTTCTATATATTCACAAAAATTAGTAAATGT	935
Pfnek-3	915	TATTTCTAAACTATTGTCTTTTAAATACATTAGAAAGATTGGAAGTAATTACTAAT	969
Pfnek-3(2)	936	TATTTCTAAACTATTGTCTTTTAAATACATTAGAAAGATTGGAAGTAATTACTAAT	990
Pfnek-3(4)	936	TATTTCTAAACTATTGTCTTTTAAATACATTAGAAAGATTGGAAGTAATTACTAAT	990
Pfnek-3	970	TATAACTTAGAAACGACACAACACGTACATTTAGAAAATGCAAATAAAATAATAT	1024
Pfnek-3(2)	991	TATAACTTAGAAACGACACAACACGTACATTTAGAAAATGCAAATAAAATAATAT	1045
Pfnek-3(4)	991	TATAACTTAGAAACGACACAACACGTACATTTAGAAAATGCAAATAAAATAATAT	1045

Shaded sequence = identical sequence; Blue sequence = vector sequence

Sequence	ELIDLYSNFEISWKIDIWMVGCILYLLLFKIHPFQNNNFLSIINASFTIPYCTKYSKRRII	268
Q9VMY8_DROME	EMLDTWSNNPIGPKVDI WALGCILYFLCYRKHPYEDGGKLRIINANYMLPP-DPQYQCFS	295
Q5RJ35_BRARE	EMIDLYSNYPIN EKQDIWALGCILYLLCFKQHPFEEGAKLQIVNGKYSIPQNDTKYTVFH	289
Q5RDC0_PONPY	EIIDLYSNFPIGEKQDIWALGCILYLLCFRQHPFEDGAKLRIVNGKYSIPPHDTQYTVFH	289
GAK_HUMAN	EIIDLYSNFPIGEKQDIWALGCILYLLCFRQHPFEDGAKLRIVNGKYSIPPHDTQYTVFH	289
Q6P1I8_MOUSE	EIVDLYSNFPIGEKQDIWALGCILYLLCFRQHPFEDGAKLRIVNGKYSIPVNDTRYTVFH	289
GAK_RAT	EIVDLYSNFPIGEKQDIWALGCILYLLCFRQHPFEDGAKLRIVNGKYSIPVNDTRYTVFH	289
	* * * * *	

```

Sequence      SILLMTLNKNPQKRIDTSTLLFILENYADLKWFVHIPAD----- 308
Q9VMY8_DROME  DIIRGCLKVNPFFERLDITMVLEGLAALAEETHNWSLKGPLD----- 335
Q5RJ35_BRARE  QLIRSMKINPEERLSINELVNQLQEIAAARNVNPKSPITERSDARVGVCFPPDACLCAC 349
Q5RDC0_PONPY  SLIRAMLQVNPEERLSIAEVVHQLQEIAAARNVNPKSPIT----- 329
GAK_HUMAN     SLIRAMLQVNPEERLSIAEVVHQLQEIAAARNVNPKSPIT----- 329
Q6P1I8_MOUSE  DLIRAMLKVNPEERLSIAEVVRQLQEIAAARNVNPKAPIT----- 329
GAK_RAT       DLIRGMLKVNPEERLSIAEVVRQLQEIAAARNVNPKAPIT----- 329
               .:.      *  **  :*.      :.  *      *      :.  :  *

Sequence      -----IKKRVNQIFEKMNELNMNNK 328
Q9VMY8_DROME  -----LHIMPIESLPAESPLKSSAYSEVYTEPLSAA 366
Q5RJ35_BRARE  VHWTVDPMSSLSSYYCLRSVYDGYESLQELLEQNGGFNN--SAQLPSQINNHN--SGVY 405
Q5RDC0_PONPY  -----ELLEQNGGYGNAALSRGPPPPVGPAGSGYSGL 362
GAK_HUMAN     -----ELLEQNGGYGSATLSRGPPPPVGPAGSGYSGL 362
Q6P1I8_MOUSE  -----ELLEQNGGYGNSGSPRAQPPCGGTVNS--SGVL 360
GAK_RAT       -----ELLEQNGGYGNSGSPRAQPPSGGPVNS--SGVL 360
               .

Sequence      HTDLIEVSHEKLIDVKIS-SQKYETNLPKSNNNFLKFFT-----SNLSGDIFFKK 378
Q9VMY8_DROME  IPSSYNG-HGSLSSLRGGAGTLLKNIKDTSTKVMQTMQQLARNDLISHITSRIIVMP 425
Q5RJ35_BRARE  DTDPTMG--GGFLDILKGGTERLFTNLKDTSSKVIQSVAS-YAKGDLDSYITSRLAVMS 462
Q5RDC0_PONPY  ALAEYDQPYGGFLDILRGGTERLFTNLKDTSSKVIQSVAN-YAKGDLDSYITSRIAVMS 421
GAK_HUMAN     ALAEYDQPYGGFLDILRGGTERLFTNLKDTSSKVIQSVAN-YAKGDLDSYITSRIAVMS 421
Q6P1I8_MOUSE  ALAEYDQPYGGFLDILRGGTERLFTNLKDTSSKVIQSVAN-YAKGDLDSYITSRIAVMS 419
GAK_RAT       ALAEYDQPYGGFLDILRGGTERLFTNLKDTSSKVIQSVAN-YAKGDLDSYITSRIAVMS 419
               :.  .  :      *.  :.  :.  :.  :.  :      *  :.  :

Sequence      IEKEETKKEIK--QVGNKQVSETEKE-PKDKLENNPDGNKLQ----NDKDSGKLQNDKD 430
Q9VMY8_DROME  CPSDGFESTYKTNNIEDVRLSLESRFVQPKLSIYNFGQRTPEPLPPVVRTVEAGSVYGCP 485
Q5RJ35_BRARE  FPAEGVESAIK-NNIEDVRLFLDSRH-AGHYAVYNLSKRSYR---PARFHNRVSECGWQ 516
Q5RDC0_PONPY  FPAEGVESALK-NSIEDVRLFLDSKH-PGHYAVYNLSPRTYR---PSRFHNRVSECGWA 475
GAK_HUMAN     FPAEGVESALK-NNIEDVRLFLDSKH-PGHYAVYNLSPRTYR---PSRFHNRVSECGWA 475
Q6P1I8_MOUSE  FPAEGVESAIK-NNIEDVRMFLDAKH-PGHYAVYNLSPRIYR---ASKFHNRVTECGWA 473
GAK_RAT       FPAEGVESAIK-NNIEDVRLFLDAKH-PGHYAVYNLSPRIYR---ASKFHNRVTECGWA 473
               :  :.  *  :.  :.  :.  :.  :      *  .  .

Sequence      SGKLQNEKDSGKLQNDKDSGKLQNDKDSGKLQNDKDSGKLQNDKVSGLQNDKDSGKLQN 490
Q9VMY8_DROME  QAHAPNLQGLFTVSADMYNFLNADPKSVVIVQTGDSGGCTAATVICALLMYADLLREPED 545
Q5RJ35_BRARE  PRRAPTLLKNLYSICKNMHWLQKQDQRNICIVHCLD-GRAASAVVVCFLCFCLFSTAEA 575
Q5RDC0_PONPY  ARRAPHLHTLYNICRNMHAWLRQDHKNVCVVHCMD-GRAASAVAVCSFLCFCLFSTAEA 534
GAK_HUMAN     ARRAPHLHTLYNICRNMHAWLRQDHKNVCVVHCMD-GRAASAVAVCSFLCFCLFSTAEA 534
Q6P1I8_MOUSE  VRRAPHLHSLYTLCRSMHAWLRREDHRNVCVVHCMD-GRAASAVAVCAFLCFCLFSTAEA 532
GAK_RAT       VRRAPHLHSLYTLCRSMHAWLRREDHRNVCVVHCMD-GRAASAVAVCAFLCFCLFSTAEA 532
               :  :  :.  .      :.  :.  :.  :.  :      :.  *  :

Sequence      DKVSGKLQNDRDGNKLQNNHIIIDYTDNKMDNDKKNKSDIEISENITSSGDKQNSVPDQSEN 550
Q9VMY8_DROME  AVQVFAVKRHTINLRPSEFRYLYYFGDILRPTLLPHY---KNTTLVLSLSCQPVPRMTKA 602
Q5RJ35_BRARE  AVYMFSMKRCPPGISASHKRYIEYMCMMAEPIPHS---KPITIRSVAMTPVPLFNKQ 632
Q5RDC0_PONPY  AVYMFSMKRCPPGIWPSHKRYIEYMCMMAEPIPHS---KPILVRVAVMTPVPLFSKQ 591
GAK_HUMAN     AVYMFSMKRCPPGIWPSHKRYIEYMCMMAEPIPHS---KPILVRVAVMTPVPLFSKQ 591
Q6P1I8_MOUSE  AVYMFSMKRCPPGIWPSHKRYIEYVCDMVAEEPITPHS---KPMLVKSVVMTVPVPLFSKQ 589
GAK_RAT       AVYMFSMKRCPPGIWPSHKRYIEYVCDMVAEEPITPHS---KPMLVKSVVMTVPVPLFSKQ 589
               :.  .  .  :.  :.  *  :.  :      :  .  .  **  :.

Sequence      QSKSHSYNLIKIQSK---DTLIIDHVD RDPDKTNTDFSFQDTLCEFEKKCMNNS---- 603
Q9VMY8_DROME  RDGCRIMYEVYCNGNLLLSTLQDYEKMRLYQAGPGKIVLPINLTACGDVTVVLFHARKGM 662
Q5RJ35_BRARE  RNGCRPFCEVYVGDERVATTSQEYDRMRDFKMEDGRAEIPLNITVQGDVLIIVYHARSTL 692
Q5RDC0_PONPY  RSGCRPFCEVYVGDERVATTSQEYDKMRDFKIEDGKAVIPLGVTVQGDVLIIVYHARSTL 651
GAK_HUMAN     RSGCRPFCEVYVGDERVATTSQEYDKMRDFKIEDGKAVIPLGVTVQGDVLIIVYHARSTL 651
Q6P1I8_MOUSE  RNGCRPFCEVYVGEERVTTTSQEYDRMKFEKIEDGKAVIPLGVTVQGDVLIIVYHARATL 649
GAK_RAT       RNGCRPFCEVYVGEERVTTTSQEYDRMKFEKIEDGKAVIPLGITVQGDVLIIVYHARSTL 649
               :.  :.  :.  :  :      *  .  :.  :  .  :.  :.  :.  :

```

Sequence -----SNYSNDKKVAGDIKK---DDCSVNIQVVNNNDVCHNDMKDVNNECTNLLNLDVFE 655
Q9VMY8_DROME -----VRPQGLKICQFQFNTGFIPEPETLITFTNQDLDDLPDPEQVTPRFCVSLSLAVTD 717
Q5RJ35_BRARE GGRLLQAKMASMKMFQIQFHTGFVPRNASTVKFAKYDLDACDIQEKPDLFQVHVNVVEVEP 752
Q5RDC0_PONPY GGRLLQAKMASMKMFQIQFHTGFVPRNATTVKFAKYDLDACDIQEKPDLFQVNVLEVEVEP 711
GAK_HUMAN GGRLLQAKMASMKMFQIQFHTGFVPRNATTVKFAKYDLDACDIQEKPDLFQVNVLEVEVEP 711
Q6P1I8_MOUSE GGRLLQAKMASMKMFQIQFHTGFVPRNATTVKFAKYDLDACDIQEKPDLFQVNVLEVEVEP 709
GAK_RAT GGRLLQAKMASMKMFQIQFHTGFVPRNATTVKFAKYDLDACDIQEKPDLFQVNVLEVEVEP 709

. . * : . . . : * : . . . *

Sequence YENETNDSIKKINDGDFIKQSGCDFIKQNG-----GDFI-KPNDNSNCIFADINFNNDIN- 708
Q9VMY8_DROME SESPPSHKPPWMPAKPK-RSPAALFSSDLEYAEMLDNFVTKPSTRSSPPPARKSDRVSS 776
Q5RJ35_BRARE QDRPSNKTPPWENFATKGLNPKILFSSRDEQQEILTKEFG-KPELPRQPGSTACYASDPQ 811
Q5RDC0_PONPY RDRPSREAPPWENSSMRGLNPKILFSSREEQQDILSKFG-KPELPRQPGSTAQYDAGAG- 769
GAK_HUMAN RDRPSREAPPWENSSMRGLNPKILFSSREEQQDILSKFG-KPELPRQPGSTAQYDAGAG- 769
Q6P1I8_MOUSE RDRPSREAPPWENTSRLGLNPKILFSNREEQQDILSKFG-KPELPRQPGSTAQYDAEAG- 767
GAK_RAT RDRPSRDVPPWENTSRLGLNPKILFSNREEQQDILSKFG-KPELPRQPGSTAQYDAEAG- 767

: * . . * * . .

Sequence ---LNNDINLNNDIN--FNNDINLNNDIDFFDVIQNHDLKKEDVLNINRNE----- 756
Q9VMY8_DROME PLVLDPDVTETAHELP---SPMPEPSPPIDLLNLNQPSDPAADPLTSAKPST----- 826
Q5RJ35_BRARE SENSPDDPQTESDSPSPQSTDTNANNFFQTLDE----- 845
Q5RDC0_PONPY ---SPEAEPTDSDSP--PSSSADASRFLHTLDWQEEKEAETGAENASPKENESALMEDRD 824
GAK_HUMAN ---SPEAEPTDSDSP--PSSSADASRFLHTLDWQEEKEAETGAENASSKESALMEDRD 824
Q6P1I8_MOUSE ---SPEAEITESDSP--QSSSTDTNHFLHTLDWQEEKEPETGLDNTSPKESQSVLIA DGD 822
GAK_RAT ---SPEAEITESDSP--QSSSTDTNHFLHTLDWQEEKDPETGV DNTSPKESQSNLIA DGD 822

: : . : . : . :

Sequence ----- 833
Q9VMY8_DROME -----DASFDLL 833
Q5RJ35_BRARE ----- 884
Q5RDC0_PONPY ESEVSDDEGGSPISSEGGEPRADEPPGLAAGLVQQDLVFEVETPAVLPEPVPQEDGVDLL 884
GAK_HUMAN ESEVSDDEGGSPISSEGGEPRADEPPGLAAGLVQQDLVFEVETPAVLPEPVPQEDGVDLL 884
Q6P1I8_MOUSE GSEVSDDEEASFPSEERKPGAGEDTPRLAAGTKQQDLIFDVGM LAAPQEPVQPEEGVDLL 882
GAK_RAT GSEVSDDEEASCPSEERKPGAGEDTPRLAAGTRQQDLIFDVGM LAAPQEPVQPEEGVDLL 882

Sequence ----- 857
Q9VMY8_DROME GAFGD-----DDSTGIG-----SAPIPDILPPPP 857
Q5RJ35_BRARE ----- 943
Q5RDC0_PONPY GLHSEVCAGPAVPPQACRAPSSNTHLLSCLLGPPEAASQGPPEDLLG-EAPLLLASPAPP 943
GAK_HUMAN GLHSEVAGPAVPPQACRAPSSNTDLLSCLLGPPEAASQGPPEDLLS-EDPLLLASPAPP 943
Q6P1I8_MOUSE GLHSEGDLRPAAPLQACGVPSNTDLLSCLLEPSDAAQVGGPPGDLGGEAPLLLASPVSP 942
GAK_RAT GLHSEGDLRPAAPLQASGVQSSNTDLLSSLEPSDASQVGGPPGDLGGETPLLLASPVSL 942

Sequence -----HDKNAITCSN----- 766
Q9VMY8_DROME LQQP-----QPKINNDPFDIFG-SVDQGGVPSMKPSGFPFVG--SLPTFVAQPAATTNP 909
Q5RJ35_BRARE ----- 1003
Q5RDC0_PONPY LSVQSTPRGGPPAVADPFGLLPSSGNSQPCSNPDLFGFELNSDSVSIPPSFPSAHSAP 1003
GAK_HUMAN LSVQSTPRGGPPAAADPFGLLPSSGNSQPCSNPDLFGFELNSDSVTVPSPFSAHSAP 1003
Q6P1I8_MOUSE LGLQNNLQGVKVPDTPDFDQFLLSSNSDTQPCSKPDLFGFELNSDSVASSTAFPSTHSAP 1002
GAK_RAT LGVQSNLQGVKVPDTPDFDQFLLPSSSDTQPCSKPDLFGFELNSDSVASSTAFPSTHSAP 1002

Sequence -----EYYNLSLFYDP-----WNNNTK----- 783
Q9VMY8_DROME SP-----THPARASADPFANIADLATGLNLN----FNRSTLSGKS----- 945
Q5RJ35_BRARE ----- 1063
Q5RDC0_PONPY PPSCSADLLHLGNLPGEPKMTASSSNPDLLGGWAAWTETAASAVAPTPAETGPLFSPGG 1063
GAK_HUMAN PPSCSADFLHLGDLPGEPKMTASSSNPDLLGGWAAWTETAASAVAPTPAETGPLFSPGG 1063
Q6P1I8_MOUSE PPSCSTAFHLHLGDLPAEPKVIASSSHPDLLGGWDTWADTATPGPASIPVPEGLTFSSAG 1062
GAK_RAT PPSCSTAFHLHLGDLPAEPKVIASSSHPDLLGGWDTWAEALPGPASMPVPEGLTFSSAG 1062

Appendix C2 Clustal alignment of PfcGAK with cGAK and BIKE kinases

[illegible]


```

Sequence -----
GAK_HUMAN      SKQRSGCRPFCEVYVGDERVASTSQEYDKMRDFKIEDGKAVIPLGVTVQGDVLIVIIYHAR 648
GAK_RAT        SKQRNGCRPFCEVYVGEERVTTTSQEYDRMKFEKIEDGKAVIPLGITVQGDVLTIIYHAR 646
BMP2K_HUMAN    AQHQPSQQQASPEYL-----TSPQEFS-----PALVSYTSSLPAQVGTIMDSSY 594
BMP2K_MOUSE    ARHQQPAQQVSPEYL-----TSPQEFS-----PALVSYASSLPAQVGTIVDSSY 581

```

```

Sequence -----
GAK_HUMAN      ---GNKLQNDKDSGK-LQNDKDSGKLQNEKDSGKLQNDKDSGKLQNDKDSGKLQNDKDSG 468
GAK_RAT        STLGGRLQAKMASMKMFQIQFHTGFVPRNATTVKFAKYDLDACDIQEKYPDLFQVNLEVE 708
BMP2K_HUMAN    S--ANRSVADKEAIAFTNQKINISNPPDMSGWNPFGE DNFSKLT EEELLDR EFDLLRSNR 652
BMP2K_MOUSE    G--ANRSVAEKEAVANFTNQKITSHPPDMSGWNPFGE DNFSKLT EEELLDR EFDLLRSNR 639
               . . . . . : : : . : : . : : . : : . : : . : : . : : . : :

```

```

Sequence -----
GAK_HUMAN      KLQNDKVS-----GKLQNDKDS 485
GAK_RAT        VEPRDRPSREAPPWENSSMRGLNPKILFSSREEQQDILSKFGKPELPRQPGSTAQYDAGA 768
GAK_RAT        VEPRDRPSRDVPPWENTSLRGLNPKILFSNREEQQDILSKFGKPELPRQPGSTAQYDAEA 766
BMP2K_HUMAN    LEERASSDKNVDSL-----APHNHP--PEDPFGSVPFISHS 687
BMP2K_MOUSE    LGASTPSDKTVDLPP-----APHSRP--PEEPFASVPFISHS 674
               . . . . . : : : . : : . : : . : : . : : . : :

```

```

Sequence -----
GAK_HUMAN      GKLQNDKVSGLQNDNRDGNKLQNNHIIIDYDNKMDND-----KKNKSDEI 530
GAK_HUMAN      GSPEAETDSDSPSSSADASRFLHTLDWQEEKEAETGAENASSKESESALMEDRDESEV 828
GAK_RAT        GSPEAEITESDSPSSSNTDNHFLHTLDWQEEKPETGVDNTSPKESQSNLIADGDGSEV 826
BMP2K_HUMAN    GSPEKKAEHS-SINQENG TANPIKNGKTS PASKDQRTGKKTS---VQGGVQKGNDESES 742
BMP2K_MOUSE    GSPEKKTTEH-SPNQKSITANLTKNGGSSPLCKDQ RAGKKTS ENPVIRGGVQKGHDDSES 733
               * . . . . . : : : . : : . : : . : : . : : . : : *

```

```

Sequence -----
GAK_HUMAN      SENITSSGDKQNSVPDQSENQSKSHSYNLIKIQSKDTLIIDHVDRDFPSDK----- 581
GAK_HUMAN      SDEGGSPISSEGQEPRADEPPG LAAGLVQDQLVFEVETPAVLPEPVPQEDGVDLLGLHS 888
GAK_RAT        SDEEEASCPSSEERKPGAGEDTPRLAAGTRQQDLIFDVGM LAAPQEPVQPEEGVDLLGLHS 886
BMP2K_HUMAN    DFESDPPSPKSSEEEEEQDDEEVLQGEQGD FNDDDEPENLGH RPLLMDESD-----E 794
BMP2K_MOUSE    DFESDPPSPKSSEEEQEDED--AQGEHGD FNDDDEPENLGH RPLLMDESD-----E 783
               . : . . . . . : : : . : : . : : . : : . : : . : :

```

```

Sequence -----
GAK_HUMAN      -TNTDFSQDTLCEFEKKCMNNSSNYSNDKKVAG-----DIKKDDCSVNIQVNNNNNDVC 635
GAK_HUMAN      EVGAGPAVPPQACKAPSSNTDLLSCLLGPPEAASQGGPEDLLS-EDPLLLASAPPLSVQ 947
GAK_RAT        EGDLRPAAPLQASGVQSSNTDLLSSLLEPSDASQVGGPDLLGGETPLLLASPVSLLG VQ 946
BMP2K_HUMAN    EEEEKHSSDSYEQAKAKYS DMSSVYRDRSGSGPTQDLNTILLTSAQLSSDVAVETPKQE 854
BMP2K_MOUSE    EEDDKHSSDSECEQAKTKRGDTSSLRRDKPGVAP---DTALLTPAR--SPADALTPSQE 837
               : . : . . . : : * . : : . : : . : : . : :

```

```

Sequence -----
GAK_HUMAN      HNDMKDVNECTNLLNLDVFEYENETNDSIKKINDGDFIKQSGCDFIKQNGGDFIKPNDSN 695
GAK_HUMAN      STPRGGPPAAADPFGLLPSSGNNSQPCSNPDLFGEFLNSDSVTVPSPFSAHSAPP-PS 1006
GAK_RAT        SNLQGKVPDTPDPFDQFLLPSSSDTQPCSKPDLFGEFLNSDSVASSTAFPSTHSAPP-PS 1005
BMP2K_HUMAN    FDFVGAVPFFAVRAQQPQ-QEKNE-KNLPQHRFPAAGLEQEEFDVFTKAPFSKKVNV-QE 911
BMP2K_MOUSE    FDFVGAVPFFAAPAPQSL-QH RGDGKNLSQHAFP---EQEDFDVFTKAPFNKKVSV-QD 891
               . . . . . : : : . : : . : : . : : . : : . : :

```

```

Sequence -----
GAK_HUMAN      CIFADINFNNDINLNNDINLNNDIN----FNNDINLNN-----DIDFFDVIQNHD 741
GAK_HUMAN      CSADFLHLGDLPGEPSKMTASSSNPDLLGGWAAWTETAASAVAPTATEGPLFSPGGQPA 1066
GAK_RAT        CSTAFLHLGDLPAEPNKVIASSSHPDLLGGWDTWAETALPGPAMPVPEGLTFSSAGHPA 1065
BMP2K_HUMAN    CHAVGPEAHTIPGYPKSVDFVGSTP-----FQPFLTSTSKSES-----NEDLFLGVPFDE 961
BMP2K_MOUSE    WPAVGPDARPLPARPRSDIFGSTP-----FQPFVSASKSES-----KEDVFLGVPFEE 941
               . . . . . : : . . : : : : : . : : . : : . : :

```

```

Sequence -----
GAK_HUMAN      LKKKEDVLNINRN-----NEHDKNAITCSN 766
GAK_HUMAN      PCGSQASWTKSQNPDPFADLGD LSSGLQGS PAGFPPGGFIPKTATTPKGSSSWQTSRPPA 1126
GAK_RAT        PPGPNPSQTKSQNPDPFADLSD LSSSLQGLPAGLPAGSFVGT SATTHKSNSSWQTT RPTA 1125
BMP2K_HUMAN    ITGSQQQ-KVKQR-----SLQKLSSRQRRTKQDMS-----KSNGKRHHGTPTS 1003
BMP2K_MOUSE    ITGSQQQKVKQR-----SLQKLSSRQRRTKQDVS-----KSNGKRHHGTPTS 984
               : . . . . : : : . : : . : : . : : . : : . : :

```


	o--*o--+-ooh-	oohrDok+-Nooo	oko+DFGo+	g+--	
Sequence	LHQILNGLNIHQN	GIIHGD LKSTNIFIKDN-	KIKIGDFGISSEQS----	-SNNNLGTLN	234
NEK9_XENLA	LFQIVSAVSCIHKA	GILHRDIKTLNIFLTKA	NLIKLG DYGLAKQLSSEYS	MAETCVGTLY	198
NEK9_HUMAN	LFQIVSAVSCIHKA	GILHRDIKTLNIFLTKA	NLIKLG DYGLAKKLNSEYS	MAETLVGTPY	216
NEK9_MOUSE	LFQIVSAVSCIHKA	GILHRDIKTLNIFLTKA	NLIKLG DYGLAKKLNSEYS	MAETLVGTPY	216
NEK1_MOUSE	FVQICLALKHVHDR	KILHRDIKSNIFLT KD	GTVQLGDFGIARVLNSTVE	LARTCIGTPY	168
Q5RFZ0_BRARE	FVQICLALKHVHDR	KILHRDIKSNIFLT KD	GTVQLGDFGIARVLNSTVE	LARTCIGTPY	168
NEK1_HUMAN	FVQICLALKHVHDR	KILHRDIKSNIFLT KD	GTVQLGDFGIARVLNSTVE	LARTCIGTPY	168

	--APEoo	Doo+oGoooo-o-----po	o o	
Sequence	CLSYESI KFKKTN	KLSDLFQVGCILYELATLSSPFCA	TNINDMISLFEDKNYKSYIIKNI	294
NEK9_XENLA	YMSPEICQGVKYS	FKSDIWAVGCVLYELLTLRTFDA	TNPLNLCVKIVQGNW--AVGLD	255
NEK9_HUMAN	YMSPELCQGVKYN	FKSDIWAVGCVIFELLTLKRTFDA	TNPLNLCVKIVQGIR---AMEVD	273
NEK9_MOUSE	YMSPELCQGVKYN	FKSDIWAVGCVIFELLTLKRTFDA	TNPLNLCVKIVQGIR---AMEVD	273
NEK1_MOUSE	YLSPEICENKPYN	NKSDIWALGCVLYELCTLKHAFAE	GNMKNLVLKIIISG----SFPPV	223
Q5RFZ0_BRARE	YLSPEICENKPYN	NKSDIWALGCVLYEMCTLKHAFAE	GNMKNLVLKIIIRG-----SYPPV	223
NEK1_HUMAN	YLSPEICENKPYN	NKSDIWALGCVLYELCTLKHAFAE	GSNMKNLVLKIIISG-----SFPPV	223

	0 00 00	R	
Sequence	SSI YSQKLVNVISKLLSLNTLER	LEVITN YNLETTQHVHLENANKIISCITVQ-----	347
NEK9_XENLA	NTV YTQELIEVVHACLEQDPEKRPTADEIL	ERPILSWRRRDMEEEKVSMNLRSNKKPRTGT	315
NEK9_HUMAN	SSQ YSLELIQMVHSCLDQDPEQRPTADELL	DRPLLKRKRREMEEKVTLNAPTKRPRSST	333
NEK9_MOUSE	SSQ YSLELIQLVHACLDQDPEQRPAAADALL	DLPLLRTRRREMEEKVTLNAPTKRPRSST	333
NEK1_MOUSE	SPH YSYDLRSLLSQLFKRNPRDRPSVNSIL	EKGFI AKRIEKF LSPQLIAEEFCLKTLSKF	283
Q5RFZ0_BRARE	SIH YSPDLRSLLAQLFKRNPRERPSVSTIL	DKPFLARRIHKF LSPQLIAQEF SHSIHLQP	283
NEK1_HUMAN	SLH YSYDLRSLSVSLFKRNPRDRPSVNSIL	EKGFI AKRIEKF LSPQLIAEEFCLKTFSKF	283

Sequence	VTEAPIAVVTSRSSEVYVWGGGKTPQKLDVFKGGWRARQVCAGDAHFAVVTVEKELYTW	375
NEK9_XENLA	VTEAPIAVVTSRTSEVYVWGGGKSTPQKLDVIKSGCSARQVCAGNTHFAVVTVEKELYTW	393
NEK9_HUMAN	VTEAPIAVVTSRTSEVYVWGGGKSTPQKLDVIKSGCSARQVCAGNTHFAVVTVEKELYTW	393
NEK9_MOUSE	GP---QPLPGKRP---ASGQGVSSFVPAQKITKPAAKYGVPLTYKKYGDKKLLEKKPPP	336
NEK1_MOUSE	KMSVAHAAPAKRP---APG--PIPITSAQKITKPAAKYGVPLTVRRPS-----	326
Q5RFZ0_BRARE	GS---QPIPAKRP---ASGQNSISVMPAQKITKPAAKYGIPLAYKKYGDKKLHEKKPLQ	336
NEK1_HUMAN		

Sequence	VNMQGGSKLHGQLGHGDRASYRQPKHVEKLQGKSVQKVSCGSDFTVCITDEGQLYSFGSD	435
NEK9_XENLA	VNMQGGTKLHGQLGHGDKASYRQPKHVEKLQGKAIHQVSCGDDFTVCVTDEGQLYAFGSD	453
NEK9_HUMAN	VNMQGGTKLHGQLGHGDKASYRQPKHVEKLQGKAIHQVSCGDDFTVCVTDEGQLYAFGSD	453
NEK9_MOUSE	KHKQAHQIPVKKMNSGEERRKKMSEEAARKRRLEFIEKEKKQKDQIR-FLKAEQMKRQEKE	395
NEK1_MOUSE	-----EVARQVPDAAKKPVKLKQ-----	344
Q5RFZ0_BRARE	KHKQAHQTPEKRVNTGEERRKISEEAARKRRLEFIEKEKKQKDQIISLMKAEQMKRQEKE	396
NEK1_HUMAN		

Sequence	YYGCLGVNQSAGSEVLEPLLVDFFLNEPVDQVSCGDSHIMALT RSKSVYSWGCGEYGRLG	495
NEK9_XENLA	YYGCMGVDKVAGPEVLEPMQLNFFLSNPVEQVSCGDNHVVLT RKNKEVYSWGCGEYGRLG	513
NEK9_HUMAN	YYGCMGVDKVSGPEVLEPMQLNFFLSNPVEQVSCGDNHVVLT RKNKEVYSWGCGEYGRLG	513
NEK9_MOUSE	RLERINRAREQG--WRNVLRAGSGSEVKASFFGIGGAVSPSPCSPRGQYEHYHAIFDQM	453
NEK1_MOUSE	-----	
Q5RFZ0_BRARE	RLERINRAREQG--WRNVLSAGSGSEVKAPFLGSGGTIAPSSFSSRGQYEHYHAIFDQM	454
NEK1_HUMAN		

Sequence	LDSEDDVYSPQKVEVQ-RGLCIVNVCCGSDGSFLLTLTGK-----	534
NEK9_XENLA	LDSEEDYYTPQKVDVP-KALIIVAVQCGCDGTFLLTQSGK-----	552
NEK9_HUMAN	LDSEEDYYTPQKVDVP-KALIIVAVQCGCDGTFLLTQSGK-----	552
NEK9_MOUSE	RLRAEDNEARWKGGIYGRWLPER-----	476
NEK1_MOUSE	-----	
Q5RFZ0_BRARE	QQRAEDNEAKWKREIYGRGLPERGILPVRPGFPYGAAGHHHFPDADDIRKTLKRLKAVS	514
NEK1_HUMAN		

Sequence	-----	
NEK9_XENLA	-----VLACGLNEHNKLGLNQYTAGIINHEAFQEVPTTSLTLAKQLSFYKIRISISPGRT	589
NEK9_HUMAN	-----VLACGLNEFNKLGLNQCMGSIINHEAYHEVPTTSTFLAKQLSFYKIRTIAPGKT	607
NEK9_MOUSE	-----VLACGLNEFNKLGLNQCMGSIINHEAYHEVPTTSTFLAKQLSFYKIRTIAPGKT	607
NEK1_MOUSE	-----QKGHLAVERANQVEEFLQRKREAMQNKARAEHGMVYLARLRQIRLQNFNERQQ	529
Q5RFZ0_BRARE	-----	
NEK1_HUMAN	KQANANRQKGQLAVERAKQVEEFLQRKREAMQNKARAEHGMVYLARLRQIRLQNFNERQQ	574
Sequence	-----	
NEK9_XENLA	HTAAIDERGRLLTFGCSNKGQQLGVGDYRKHLGINLLGGPLGGKQVIRVSCGDEFTTAATA	649
NEK9_HUMAN	HTAAIDERGRLLTFGCSNKGQQLGVGNKKRLGINLLGGPLGGKQVIRVSCGDEFTIAATD	667
NEK9_MOUSE	HTAAIDERGRLLTFGCSNKGQQLGVGNKKRLGINLLGGPLGGKQVIRVSCGDEFTIAATD	667
NEK1_MOUSE	IKAKLRGENKEADGTKGQEATEETDMRLKKMES-LKAQTNARA AVLKEQLERKRKEAYER	588
Q5RFZ0_BRARE	-----	
NEK1_HUMAN	IKAKLRGEKKEANHSEGGQEGSEADMRRKKIES-LKAHANARA AVLKEQLERKRKEAYER	633
Sequence	-----	
NEK9_XENLA	DNHIFAWGNGGNGRLAMTPNERPQGS DICTSWPRPIFGSLHHVT-DLSCRGWHTILIVEK	708
NEK9_HUMAN	DNHIFAWGNGGNGRLAMTPTERPHGSDICTSWPRPIFGSLHHVP-DLSCRGWHTILIVEK	726
NEK9_MOUSE	DNHIFAWGNGGNGRLAMTPTERPHGSDICTSWPRPIFGSLHHVP-DLSCRGWHTILIVEK	726
NEK1_MOUSE	EKKVWEEHLVAR-VKSSDVPLPLELLETTGGSPSKQKQVPVISVTSALKEVGLDGS LDTQ	647
Q5RFZ0_BRARE	-----	
NEK1_HUMAN	EKKVWEEHLVAKGVKSSDVSPPLGQHETGGSPSKQQMRSVISVTSALKEVGV DSSLTDTR	693
Sequence	-----	
NEK9_XENLA	VLNSKTI RSNSSGLS IGT LAQSCSSGG-----SGS REEDSDRESLTS DPSRGFR	757
NEK9_HUMAN	VLNSKTI RSNSSGLS IGT VFQSSSPGG-----GGGGGGGEEEDSQQES ETPDPSGGFR	779
NEK9_MOUSE	VLNSKTI RSNSSGLS IGT VVQSSSPGGGIGGGGGGGGGGEEEDSQQES ETPDPSGGFR	786
NEK1_MOUSE	E--EEMEKSNSA ISSKREILRRLNENL-----KAQEDEKEKQHHS GSCETVGHK	694
Q5RFZ0_BRARE	-----	
NEK1_HUMAN	ETSEEMQKTNNAISSKREILRRLNENL-----KAQEDEKKGQNLSDTFEINVHE	742
Sequence	-----	
NEK9_XENLA	GTIEAEP-ETGPFNTTENME-----SSSCP SWLRQLEEA EFIPMPDTPNFVSAESSQN-	810
NEK9_HUMAN	GTMEADRGM EGLISPT EAMGNSNGASSSCP GWLRKELENA EFIPMPDSPSPLSAAFSESE	839
NEK9_MOUSE	GTMEADRGM EGLISPT EAVGNSCGASSSCP GWLRKELENA EFIPMPDSPAPLSAAFSQSE	846
NEK1_MOUSE	DEREYETENAISSDRKKWEMGGQLVIPLDAVTLDTSF SATEKHTVGEVIKLD SNGSPRKV	754
Q5RFZ0_BRARE	-----	
NEK1_HUMAN	DAKEHEKEKSVSSDRKKWEAGGQLVIPLDELTLDTSFSTTERHTVGEVIKLG PNGSPRRA	802
Sequence	-----	
NEK9_XENLA	-GTTSMQDVR-----ETPPD-----ALEKP-----SFQACT	835
NEK9_HUMAN	KDTLPYEELQGLKVASEAPLE-----HKPQVEASSPRLNPAVTCAGKGTPLTPPACA	891
NEK9_MOUSE	KDTLPYEELQGLKVASEVPPE-----PQRAAGAWPPRLDPAVPCVGK--ALTSAACA	896
NEK1_MOUSE	WGKNPTDSVLKILGEAELQL-----	774
Q5RFZ0_BRARE	-----	
NEK1_HUMAN	WGKSPTDSVLKILGEAELQLQTELLNTTIRSEISPEGEKYKPLITGEKKVQCISHEINP	862
Sequence	-----	
NEK9_XENLA	CSTLQEEVRKLQDLVSTYRHEQELLCRENSRLALEVQELNGKLQSAMQMNQVVNKHGEAI	895
NEK9_HUMAN	CSSLQVEVERLQGLVLKCLAEQQKLQQENLQIFTQLKLNKKLEGGQQ----VGMHSGKT	947
NEK9_MOUSE	CSALQVEVDRLQALVLKCL EEQQKLQQENLQMSTQLKLNKKLEGGQQ----VGMHSRG T	952
NEK1_MOUSE	-----	
Q5RFZ0_BRARE	-----	
NEK1_HUMAN	SAIVDSPVETKSPFSEASPQMSLKLEGNLEEPDDLETEILQEPSGTNKDES L PCTITDV	922
Sequence	-----	
NEK9_XENLA	HQIQKQLSLHWKSSPPSSGNPEMSREDSDAESWCFLGT EACRSSSGTGM-----	944
NEK9_HUMAN	QTAKEEMEMDPKP-----DLDSWCLLGT DSCRPSL-----	979
NEK9_MOUSE	QTAKEEMEMDPKP-----DLDSWCLLGT DSCRPSL-----	984
NEK1_MOUSE	-----	
Q5RFZ0_BRARE	-----	
NEK1_HUMAN	WISEEKETKETQSADRI TIQENEVSEDGV SSTVDQLSDIHI EPGTND SQHSKCDVDKSVQ	982

```

Sequence -----
NEK9_XENLA -----
NEK9_HUMAN -----
NEK9_MOUSE -----
NEK1_MOUSE -----
Q5RFZ0_BRARE -----
NEK1_HUMAN PEPFFH KVVVHSEHLNLVPQVQSVQCSPEESFAFRSHSHLPPK NKNKNSLLI GLSTGLFDA 1042

```

Appendix C5 Clustal alignment of human cGAK excluding PfcGAK and including bovine auxilin.

Colour code for amino acids: **Red** = small hydrophobic amino acid residues (including aromatic residues); **Blue** = acidic amino acid residues; **Magenta** = basic amino acid residues; **Green** = hydroxyl, amine and glutamine residues, **[*]** = identical amino acids; **[:]** = conserved substitutions; **[.]** = semi-conserved substitutions. Sequence = human cGAK

```

Sequence -----MSLLQSALD FLAGPGLGGASGRDQSD FVGQTVELGELRLRVRRVLAEGG 50
Q5RDC0_PONPY -----MSLLQSALD FLAGPGLGGASGRDHSDFVGQTVELGDLRLRVRRVLAEGG 50
Q6P1I8_MOUSE -----MSLLQSALD FLAGPGLGGAAAGRDQSD FVGQTVELGELRLRVRRVLAEGG 50
GAK_RAT -----MSLLQSALD FLAGPGLGGAAAGRDQSD FVGQTVELGELRLRVRRVLAEGG 50
Q9VMY8_DROME MGEFFKSLN LNYFSSSEVNGAAGNGAPEALGGRLDND FVGQVVEVAGHRLRIKCVIAEGG 60
Q5T614_HUMAN -----
AUXI_BOVIN -----

```

```

Sequence FAFVYEAQDV GSGREYALKRLLSNEEEKNRAIIQEVCFMKKLSGHPNIVQFCSAASIGKE 110
Q5RDC0_PONPY FAFVYEAQDV GSGREYALKRLLSNEEEKNRAIIQEVCFMKKLSGHPNIVQFCSAASIGKE 110
Q6P1I8_MOUSE FAFVYEAQDL GSGREYALKRLLSNEEEKNRAIIQEVCFMKKLSGHPNIVQFCSAASIGKE 110
GAK_RAT FAFVYEAQDL GSGREYALKRLLSNEEEKNRAIIQEVCFMKKLSGHPNIVQFCSAASIGKE 110
Q9VMY8_DROME YAFVYVAQDV QTGT EYALKRLIGADMQASTAINEINIHKQLSGHENIVAFV GSSYTAPS 120
Q5T614_HUMAN -----
AUXI_BOVIN -----

```

```

Sequence ESDTGQAEFLLL TELCKG-QLVEFLKKMESRGPLSCDTVLKIFYQTCRAVQHMHQKPP I 169
Q5RDC0_PONPY ESDTGQAEFLLL TELCKG-QLVEFLKKMESRGPLSCDTVLKIFYQTCRAVQHMHQKPP I 169
Q6P1I8_MOUSE ESDTGQAEFLLL TELCKG-QLVEFLKRVCKGPLSCDSILKIFYQTCRAVQHMHQKPP I 169
GAK_RAT ESDTGQAEFLLL TELCKG-QLVEFLRRVECKGPLSCDSILKIFYQTCRAVQHMHQKPP I 169
Q9VMY8_DROME TQLG--AQYLLL TELCKGGSGLVDCFR--TNNAPFNPTCVLRIFYQMARAVASLHSQSPPI 176
Q5T614_HUMAN -----
AUXI_BOVIN -----

```

```

Sequence IHRDLKVENLLLSNQGTIKLCDFGSATTISHYPDYSWSAQRRALVEEEEITRNTTPMYRTP 229
Q5RDC0_PONPY IHRDLKVENLLLSNQGTIKLCDFGSAT AISHYPDYSWSAQRRALVEEEEITRNTTPMYRTP 229
Q6P1I8_MOUSE IHRDLKVENLLLSNQGTIKLCDFGSATTISHYPDYSWSAQKRAMVEEEEITRNTTPMYRTP 229
GAK_RAT IHRDLKVENLLLSNQGTIKLCDFGSATTISHYPDYSWSAQKRAMVEEEEITRNTTPMYRTP 229
Q9VMY8_DROME AHRDIKIENFLIGNDKQIKLCDFGSASTEVLSPTFEWSANQSMLEQLNTVTTTPMYRSP 236
Q5T614_HUMAN -----
AUXI_BOVIN -----

```

```

Sequence EIIDLYSNFPIGEKQDIWALGCILYLLCFRQHPPFEDGAKLRIVNGKYSIPPHDTQYTVFH 289
Q5RDC0_PONPY EIIDLYSNFPIGEKQDIWALGCILYLLCFRQHPPFEDGAKLRIVNGKYSIPPHDTQYTVFH 289
Q6P1I8_MOUSE EIVDLYSNFPIGEKQDIWALGCILYLLCFRQHPPFEDGAKLRIVNGKYSIPVNDTRYTVFH 289
GAK_RAT EIVDLYSNFPIGEKQDIWALGCILYLLCFRQHPPFEDGAKLRIVNGKYSIPVNDTRYTVFH 289
Q9VMY8_DROME EMLDTWSNNPIGPKVDIWALGCILYFLCYRKHPYEDGGKLRIINANYMLPP-DPQYQCF S 295
Q5T614_HUMAN -----MSLLGSYRKKT SNDGYES----- 18
AUXI_BOVIN -----

```

Sequence	SLIRAMLQVNPEERLSIAEVVHQLQEIAAARNVNPKSPITELLEQNGGYGSATLSRGPPP	349
Q5RDC0_PONPY	SLIRAMLQVNPEERLSIAEVVHQLQEIAAARNVNPKSPITELLEQNGGYGNAALS R G P P P	349
Q6P1I8_MOUSE	DLIRAMLKVNPEERLSIAEVVRQLQEIAAARNVNPKAPITELLEQNGGYGNSGSPSRAQPP	349
GAK_RAT	DLIRGMLKVNPEERLSIAEVVRQLQEIAAARNVNPKAPITELLEQNGGYGNSGSPSRAQPP	349
Q9VMY8_DROME	DIIRGCLKVNPFERLDITMVLEGLAALAETHNWSLKGPLDLHIMPIESLPAESPLKSSAY	355
Q5T614_HUMAN	---LQLVDSNGDL SAGSGGVGGKQ R V N A G A A A R S P A R Q P P D R A S T M D S S G A S S P D M E P S Y	75
AUX1_BOVIN	---LQLVDSNGDL SAGSGGVGGKQ R V N A G A A A R S P A R Q P P D R A S T M D S S G A S S P D M E P S Y	15

Sequence		
Q5RDC0_PONPY	DISYITSRIAVMSFPAEGVESALK-NNIEDVRLFLDSKHP-GHYAVYNLSPRTYRPSRFH	466
Q6P1I8_MOUSE	DISYITSRIAVMSFPAEGVESALK-NSIEDVRLFLDSKHP-GHYAVYNLSPRTYRPSRFH	466
GAK_RAT	DISYITSRIAVMSFPAEGVESAIK-NNIEDVRMFLDAKHP-GHYAVYNLSPRIYRASKFH	464
Q9VMY8_DROME	DISYITSRIAVMSFPAEGVESAIK-NNIEDVRLFLDAKHP-GHYAVYNLSPRIYRASKFH	464
Q5T614_HUMAN	DISHITSRIILVMPCPSDGFESTYKTNNIEDVRLSLESRFVPQKLSIYNFGQETPEPLPP	472
AUX1_BOVIN	DFTYVTSRIIVMSFPLNDVIGFR-NQVDDIRSFLDSRHL-DHYTVYNLSPKSYRTAKFH	179
	DFTYVTSRIIVMSFPLDSVDITGER-NQVDDIRSFLDSRHL-DHYTVYNLSPKSYRTAKFH	119

Sequence	FLCFCRLFFSTAEAAVYMFSMKRCPPIWPSHKRYIEYMCDMVAEEPITPHSKPILVRAVV	581
Q5RDC0_PONPY	FLCFCRLFFSTAEAAVYMFSMKRCPPIWPSHKRYIEYMCDMVAEEPITPHSKPILVRAVV	581
Q6P1I8_MOUSE	FLCFCRLFFSTAEAAVYMFSMKRCPPIWPSHKRYIEYVCDMVAEEPITPHSKPMLVKSVV	579
GAK_RAT	FLCFCRLFFSTAEAAVYMFSMKRCPPIWPSHKRYIEYVCDMVAEEPITPHSKPMLVKSVV	579
Q9VMY8_DROME	LLMYADLLREPEDAVQVFAVRHTINLRPSEFRYLFFGDIRLPPTLLPHYKNTTLVLSLS	592
Q5T614_HUMAN	MFIFCNLYSTPGPAIRLLYAKRPGIGLSPSHRRYLYGMCDDLADKPYRPHFKPLTIKSTIT	294
AUX1_BOVIN	MEFECNLYSTPGPAVRLLYAKRPGIGLSPSHRRYLYGMCDDIADKPYRPHFKPLTIKSTIT	233

Sequence	QVNLEVEVEPRDRPSREAPPWENSSMRGLNPKILFSSREEQQDILSKFGKPELPRQPGST	761
Q5RDC0_PONPY	QVNLEVEVEPRDRPSREAPPWENSSMRGLNPKILFSSREEQQDILSKFGKPELPRQPGST	761
Q6P1I8_MOUSE	QVNLEVEVEPRDRPSREAPPWENTSLRGLNPKILFSSNREEQQDILSKFGKPELPRQPGST	759
GAK_RAT	QVNLEVEVEPRDRPSRDVPPWENTSLRGLNPKILFSSNREEQQDILSKFGKPELPRQPGST	759
Q9VMY8_DROME	CVSLSLAVTDSSEPPSHKPPWMPAKPK-RSPAALFSSDLEYAEMLDNFVTKPSTRSSPPP	766
Q5T614_HUMAN	QVTLDVELQPHDKVIDLTPPWEHYCTKDVNPSILFSSHQEHQDTLALGGQAPIDIPPDP	474
AUXI_BOVIN	QVTLDVELQPHDKVMELTPPWEHYCTKDVNPSILFSSHQEHQDTLVLGQAPIDIPPDP	414

* * * * *

Sequence	AQYDAGAGSPEAEPTDSDSPSSSADASRFLHTLDWQEEKEAETGAENASSKESESALME	821
Q5RDC0_PONPY	AQYDAGAGSPEAEPTDSDSPSSSADASRFLHTLDWQEEKEAETGAENASSPKENESALME	821
Q6P1I8_MOUSE	AQYDAEAGSPEAEITESDSPSSSADTNHFLHTLDWQEEKEPETGLDNTSPKESQSVLIA	819
GAK_RAT	AQYDAEAGSPEAEITESDSPSSSADTNHFLHTLDWQEEKEPETGVDNTSPKESQSNLIA	819
Q9VMY8_DROME	-----	
Q5T614_HUMAN	RHYGQSG-----FFASLCWQDQKS-----EKSFCCEEDHAALVN	507
AUXI_BOVIN	RHFGQGG-----FFSTLCWQDQKS-----EKSFCCEEDHAALVN	447

Sequence	DRDESEVSDEGGSPISSEGGEPADPEPPGLAAGLVQQDLVFEVETPAVLPEPVPQEDGV	881
Q5RDC0_PONPY	DRDESEVSDEGGSPISSEGGEPADPEPPGLAAGLVQQDLVFEVETPAVLPEPVPQEDGV	881
Q6P1I8_MOUSE	DGDGSEVSDEEEASFSEERKPGAGEDTPRLAAGTKQQDLIFDVGMLAAPQEPVQPEEGV	879
GAK_RAT	DGDGSEVSDEEEASCPSERKPGAGEDTPRLAAGTKQQDLIFDVGMLAAPQEPVQPEEGV	879
Q9VMY8_DROME	---GARKSDRVSSPLVLPDVTETAHELPSMP-----EPSPP---I	801
Q5T614_HUMAN	Q--ESEQSDELLTLSSPHGNANGDKPHGVKKPSKKQQ-----EPAAPPPED-V	554
AUXI_BOVIN	Q--ESEQSDELLTLSSPHGNANGDKPHAARKPSKKQQ-----EPAAPPPED-V	494

* * *

Sequence	DLLGLHSEVGAGPAVPPQACKAPSSNTDLLSCLLGPPEAASQGGPEDLLS-EDPLLLASP	940
Q5RDC0_PONPY	DLLGLHSEVCAGPAVPPQACRAPSSNTHLLSCLLGPPEAASQGGPEDLLG-EAPLLLASP	940
Q6P1I8_MOUSE	DLLGLHSEGDLRPAAPLQACGVPSNTDLLSCLLEPSDAAQVGGPDLLGGEPALLASP	939
GAK_RAT	DLLGLHSEGDLRPAAPLQASGVQSSNTDLLSCLLEPSDAAQVGGPDLLGGEPALLASP	939
Q9VMY8_DROME	DLLNLNQPPSDVPAADPLTSAPKS--TDAFDLLGAFGDDSTGIG-----SAPIPDILP	854
Q5T614_HUMAN	DLLGLEGSAMNSFSPP---AAPPTNSELLSDFGSGGAAGPTQAG-QSGVEDVFHPSGP	610
AUXI_BOVIN	DLLGLEGSAMNSKFSSP---AAPPTNSELLSDFGSGGAAGPVQSG-QSGVDDVFHPSGP	550

* * * * *

Sequence	APPLSVQS---TPRGPPAAADPFGLLPSSGNNSQPCSNPDFGFEFLNSDSVTVPPSFP	997
Q5RDC0_PONPY	APPLSVQS---TPRGPPAVADPFGLLPSSGSNSQPCSNPDFGFEFLNSDSVIPPSP	997
Q6P1I8_MOUSE	VSPLGLQN---NLQGKVPDTPDPFDQLLSSNSDTQPCSKPDFGFEFLNSDVSASTAFP	996
GAK_RAT	VSLLGVQS---NLQGKVPDTPDPFDQLLSSSDTQPCSKPDFGFEFLNSDVSASTAFP	996
Q9VMY8_DROME	PPPLQQPQ-----PKINNDPFDIFGSDVQGGVPSMKPSGFPFVGS LPTFVAQPA	905
Q5T614_HUMAN	ASTQSTPRRSATSTASPTLRVGEATDFPFGAPSKPSGQ-DLLGSFLNTSSASSDPFLQ	669
AUXI_BOVIN	TSTQSTPRRSATSTASPTLRVGEATDFPFGAPSKPSGQ-DLLGSFLNTASASSDPFLQ	609

* * * * *

Sequence	SAHSAPPPSCSADFLHLGDLPGEPSKMTASSNPDLLGGWAAWTETAASAVAPTPATEGP	1057
Q5RDC0_PONPY	SAHSAPPPSCSADLLHLGNLPGEPSKMTASSNPDLLGGWAAWTETAASAVAPTPATEGP	1057
Q6P1I8_MOUSE	STHSAPPPSCSTAFHLHLGDLPAEPSKVIASSSHPDLLGGWDTWADTATPGASIPVPEGT	1056
GAK_RAT	STHSAPPPSCSTAFHLHLGDLPAEPNKVIASSSHPDLLGGWDTWAETALPGASMPVPEGT	1056
Q9VMY8_DROME	TTNPSPT-----HPARASADPFAN-IADLATGLNLNFRSTLS-GKSPVNTSP	951
Q5T614_HUMAN	PTRSPSP-----TVHASSTPAVNIQPDVSGGDWHAKPGGFGMGSKSAATSP	716
AUXI_BOVIN	PTRSPSP-----TVHASSTPAVNIQPDVSGAWDWHHTKPGGFGMGSKSAATSP	656

* * * * *

Sequence	LFSPGGQPAPCGSQASWTKSQNPDPFADLGLDSSGLQGSFAGFPP---GGFIPKTATTAK	1114
Q5RDC0_PONPY	LFSPGGQPAPCGSQASWTKSQNPDPFADLGLDSSGLQGSFAGFPP---GGFVPKTATTAK	1114
Q6P1I8_MOUSE	LFSSAGHPAPPGPNPSQTKSQNPDPFADLSDLSSSLQGLPAGLPA---GGFVGAPAPTQK	1113
GAK_RAT	LFSSAGHPAPPGPNPSQTKSQNPDPFADLSDLSSSLQGLPAGLPA---GSFVGTSATTHK	1113
Q9VMY8_DROME	QPTQFSSPT-----HKPSPSSQPQATFMHTPP---TPQTLPS-----TSSIRT	991
Q5T614_HUMAN	TGSSHGTP-----HQSKPQTLDPFADLGLTGGSSSFASKPTTPTGLGGGFPPLLSSPQKA	770
AUXI_BOVIN	TGSSHGTP-----HQNKPQTLDPFADLGLTGGSSSFASKPTPTGLGGGFPPLLSSPQKA	710

* * * * *

Appendix D1 Blastn analysis of cGAK nanomeric repeat region (first 10 pairwise alignments)

>EM_INV:AE014851 AE014851.1 Plasmodium falciparum 3D7 chromosome 12, section 8
of 9 of the complete sequence.
Length = 251,762

Plus Strand HSPs:

Score = 1215 (188.3 bits), Expect = 1.1e-46, P = 1.1e-46
Identities = 243/243 (100%), Positives = 243/243 (100%), Strand = Plus / Plus

Query: 1 AAAGATAGTGGTAAATTACAAAATGATAAAGATAGTGGTAAATTACAAAATGAAAAAGAT 60
|||||
Sbjct: 201570 AAAGATAGTGGTAAATTACAAAATGATAAAGATAGTGGTAAATTACAAAATGAAAAAGAT 201629

Query: 61 AGTGGTAAATTACAAAATGATAAAGATAGTGGTAAATTACAAAATGATAAAGATAGCGGT 120
|||||
Sbjct: 201630 AGTGGTAAATTACAAAATGATAAAGATAGTGGTAAATTACAAAATGATAAAGATAGCGGT 201689

Query: 121 AAATTACAAAATGATAAAGATAGTGGTAAATTACAAAATGATAAAGTTAGTGGTAAATTA 180
|||||
Sbjct: 201690 AAATTACAAAATGATAAAGATAGTGGTAAATTACAAAATGATAAAGTTAGTGGTAAATTA 201749

Query: 181 CAAAATGATAAAGATAGTGGTAAATTACAAAATGATAAAGTTAGTGGTAAATTACAAAAT 240
|||||
Sbjct: 201750 CAAAATGATAAAGATAGTGGTAAATTACAAAATGATAAAGTTAGTGGTAAATTACAAAAT 201809

Query: 241 GAT 243
|||
Sbjct: 201810 GAT 201812

>EM_INV:AE014845 AE014845.1 Plasmodium falciparum 3D7 chromosome 12, section 2
of 9 of the complete sequence.
Length = 250,531

Plus Strand HSPs:

Score = 523 (84.5 bits), Expect = 2.0e-15, P = 2.0e-15
Identities = 177/242 (73%), Positives = 177/242 (73%), Strand = Plus / Plus

Query: 11 GTAAAT-TACAAAATGATAAAGATAGTGGTAAATTA-CAAAATGAAAAAGATAGTGGTAA 68
| ||||| ||||| | ||| | ||||| ||||| | ||| | |||||
Sbjct: 41116 GAAAAATACAAAATGATGATAATAAAGA-AAATAAACAAAATGATGACAATAA-GTTAA 41173

Query: 69 AT-TACAAAATGATAAAGATAGTGGTAAAT-TACAAAATGATAAAGATAGCGGTAAAT-T 125
|| ||||| ||||| | ||||| ||||| ||||| | ||| | |||||
Sbjct: 41174 ATATACAAAATGATGATAATAAAGA-AAATATACAAAATGATGATAATAAAG-TAAATAT 41231

Query: 126 ACAAATGATAAAGATAGTGGTAAAT-TACAAAATGATAAAGTTAGTGGTAAAT-TACAA 183
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Sbjct: 41232 ACAAATGATGATAATAAAG-TAAATATACAAAATGATGATAATAAAGA-AAATATACAA 41289

Query: 184 AATGATAAAGATAGTGGTAAAT-TACAAAATGATAAAGTTAGTGGTAAAT-TACAAAATG 241
||||| | ||| | ||||| ||||| ||||| ||||| ||||| ||||| |||||
Sbjct: 41290 AATGATGATAATAAAG-TAAATATACAAAATGATGATAATAAAG-TAAATATACAAAATG 41347

Query: 242 AT 243
||
Sbjct: 41348 AT 41349

>EM_INV:AE014832 AE014832.1 Plasmodium falciparum 3D7 chromosome 10 section 4
of 7 of the complete sequence.
Length = 258,658

Score = 251 (43.7 bits), Expect = 0.0038, P = 0.0038
Identities = 125/196 (63%), Positives = 125/196 (63%), Strand = Plus / Plus

```

Query:      13 AAATTACAAATGATAAAGATAGTGGTAAATTACAAATGAAAAAGA-TAGTGGTAAATT  71
             ||| ||| |||| | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Sbjct: 81650 AAAATA-AAAAGAAAAAAATAAAAA-AAAGAAAAAAATTAAAAAAATAAAGATAA-TG 81706

Query:      72 ACAAA-ATGATAAAGATAG-TGGTAAATTACAAATGATAAAGATAGCGGTAAATTACAA 129
             | ||| ||||| || | ||| ||| | | ||||| ||||| ||| ||| ||| |||
Sbjct: 81707 ATAAAGATGATAATAA-AGATGATAATAAAGATGATGATAAAGATGATGATAAG-ATGA 81764

Query:     130 AATGATAAAGATAGTGGTAAATTACAAATGATAAAGTTAGTGGTAAATTACAA-AATGA 188
             --||| ||||| || |||| | |||| | |||| | |||| | |||| | ||||
Sbjct: 81765 --TGCTAAAGATGATGCTAAAG---ATGATACTAAAGATGATGATAA-TAATAATAATTA 81818

Query:     189 TAAAGATAGTGGTAAA 204
             ||| ||| ||| |||
Sbjct: 81819 TAACAATAATATTTAAA 81834

```

Plus Strand HSPs:

```

Score = 399 (65.9 bits), Expect = 7.8e-10, P = 7.8e-10
Identities = 163/240 (67%), Positives = 163/240 (67%), Strand = Plus / Plus

Query:      1 AAAGATAGTGGTAAATTACAA-AATGATAAAGATAGTGGTAAATTACAAA-ATGAA-AAA 57
             |||||  |||  |  |||  |||||  |||  |||  |||  |||  |||
Sbjct: 122112 AAAGATAAAGATAAAGAAAGATAAAGATAAAGATAAAGATAAAG-ATAAAGATAAAGAAA 122170

Query:      58 GATAGTGGTAAATTACAAAATGATAAAGATAGTGGTAAATTACAA-AATGATAAAGATAG 116
             |||  |  |||  |||  |||||  |||  |||  |||  |||  |||
Sbjct: 122171 GATAAAGATGAAG-ATAAA--GATAAAGATAAAGATAAAGAAAGATAAAGATAAAGATAA 122227

Query:     117 CGGTAAATTACAAAATGATAAAGATAGTGGTAAAT-TACAAA-ATGATAAAGTTAGTGGT 174
             |||  |||  |||||  |||  |||||  |||  |||  |||  |||
Sbjct: 122228 AGATAGATAAAGAAA-GATAAAGATAA-GATAAAGATAAAGATAAGATAAAGATAAAGAT 122285

Query:     175 AAAT-TACA-A-AATGATAAAGATAGTGGTAAATTACAAAATGATAAAGTTAGTGGTAAA 231
             |||  |  |  |||  |||||  |||  |||||  |||  |||
Sbjct: 122286 AAAGATGAAGATAAAGATGAAGATAAAGTTAAAG-ATAAA--GATAAAGATAAAGATAA 122342

```

Plus Strand HSPs:

Score = 396 (65.5 bits), Expect = 1.1e-09, P = 1.1e-09
Identities = 168/252 (66%), Positives = 168/252 (66%), Strand = Plus / Plus

Query: 1 AAAGATAGTGGTAAATTAC-AAAATGATAAAGATAGTGGTAAATTAC-AAAATGAAAAAG 58
 ||| ||| || | || ||||||||| ||| || | || |||||||||
Sbjct: 248338 AAACATAAAA-TATATGACGAAAATGATAAACATAAAA-TATATGACGAAAATGATAAAA 248395

Query: 59 ATAGTGGTAAATTAC-AAAATGATAAAGATAGTGGTAAATTAC-AAAATGATAAAGATAG 116
 ||| || | || ||||||||| ||| || | || |||||||||
Sbjct: 248396 ATAAAA-TATATGACGAAAATGATAAACATAAAA-TATATGATGAAAAGGATAAACATAA 248453

Query: 117 CGGTAAATTAC-AAAATGATAAAGATAGTGGTAAATTAC-AAAATGATAAAGTTAGTGGT 174
 || | || | || ||||||||| ||| || | || |||||||||
Sbjct: 248454 AA-TATATGACGAAAATGATAAAAAATAAAA-TATATGATGAAAAGGATAAACATAAAA-T 248510

Query: 175 AAATTAC-AAAATGATAAAGATAGTGGTAAATTAC-AAAATGATAAAGTTAGTGGTAAAT 232
 | || | || || || || || || | || | || || || || ||
Sbjct: 248511 ATGTGATGAAAATAATAAAAAATAA-TATATGATGAAAAGGATAAACATAAAA-TATAT 248568

Query: 233 TAC-AAAATGAT 243
 | || || || || ||
Sbjct: 248569 GATGAAAATGAT 248580

>EM_INV:AY898648 AY898648.1 Plasmodium falciparum tryptophan-rich antigen
pseudogene, complete sequence.
Length = 2543

Plus Strand HSPs:

Score = 393 (65.0 bits), Expect = 1.1e-09, P = 1.1e-09
Identities = 163/234 (69%), Positives = 163/234 (69%), Strand = Plus / Plus

```
Query:      1 AAAGATAGTGGTAAATTACAAAAT-G--ATAAAGATAGTGGTAAAT-TACAA--AATGAA 54
             ||||| ||| ||| ||| ||| ||||| ||||| ||| ||| ||| |||
Sbjct:    705 AAAGATAGTGATAAAGAAAAAGATAGCGATAAAGATAGTGATAAAGGTAAAGTAAAGGA 764

Query:     55 AAAGATAGTGGTAAATTACAAAATGATAAAGATAGTGGTAAATTACAAAATGATAAAGAT 114
             ||||| ||| ||| ||| ||||| ||||| ||| ||| ||| |||
Sbjct:    765 AAAGATAGCGATAAAGGT--AAA-GATAAAGATAGCGATAAAGGT-ATA--GATAGCGAT 818

Query:    115 AGCGGTAAAT-TA-CAA-AATGATAAAGATAGTGGTAAAT-TACA-A-AATGATAAAGTT 168
             | |||| | ||| ||| ||| ||||| ||| ||||| ||| ||| ||| |||
Sbjct:    819 AAAGGTATAGATAGCGATAAAGGTAAAGATAGCGATAAATGTAAAGATAAAGATAGCGAT 878

Query:    169 AGTGGTAAAT-TA-CAA-AATGATAAAGATAGTGGTAAAT-TACAAAATGATAA 218
             | |||| | ||| ||| ||| ||||| ||| ||||| ||| ||| ||| |||
Sbjct:    879 AAATGTAAAGATAACAATAAAGATAAAGATAGCGATAAATGTA-AAGATAACAA 931
```

>EM_INV:CR382401 CR382401.1 Plasmodium falciparum chromosome 6, complete
sequence; segment 4/5
Length = 337,203

Plus Strand HSPs:

Score = 296 (50.5 bits), Expect = 3.5e-05, P = 3.5e-05
Identities = 164/256 (64%), Positives = 164/256 (64%), Strand = Plus / Plus

```
Query:      1 AAAGATAGTGGTAAATTA-CAAAATGATAAAGATAGTGGT-AAATT-ACAAA-ATGAAAA 56
             ||| ||| ||| ||| ||||| ||||| ||| ||| ||| ||| ||| |||
Sbjct: 26520 AAAAAACAATTAATTAATCAAAATGATAAATATAGAGAACAAATTGATAAATATAAAAA 26579

Query:     57 AGATAGTGGTAAATTACAAAATGATAAAGATAGTGGTAAATTACAAAATGATAAAGATAG 116
             | |||| ||||| ||||| ||| ||| ||| ||||| ||||| ||| |||
Sbjct: 26580 TCA-ACAAGAAAAT-ATAAAACACAACAA-ACCAGTAATAAAAAAATATTATACA-AC 26635

Query:    117 CGGTAAATTACAA-AATGA-TAAAGATAGTGGTAAATTACAAA-AT-GATAAAG-TTAGT 171
             | |||| | ||| ||||| ||||| ||| ||| ||||| ||||| ||| |||
Sbjct: 26636 TTGAAAATGAAATCAATGAATTAA-ATA-TGA-AAATTAATAATATAGATAAAAAATTA-T 26691

Query:    172 GGTAA-ATTACAAAATGATAAAGATAGTGGTAAATTACAAAAT---GATAAAGTTAGTG 226
             | |||| ||||| ||| ||||| ||||| ||| ||||| ||||| ||| |||
Sbjct: 26692 AATCTTATTAAAAAA-GAAAAAGATAATT-TAAATCAAGAAATTAACGATAAA-TCAGAA 26748

Query:    227 GTAAATTACAAAATGA 242
             | ||||| ||| |||
Sbjct: 26749 G-AAATTAATATATTA 26763
```

>EM_MUS:AC144552 AC144552.3 Mus musculus BAC clone RP24-271M6 from chromosome
Y, complete sequence.
Length = 166,988

Plus Strand HSPs:

Score = 365 (60.8 bits), Expect = 2.7e-08, P = 2.7e-08
Identities = 161/241 (66%), Positives = 161/241 (66%), Strand = Plus / Plus

```
Query:      1 AAAGATAGTGGTAAATTACAAAATGATAAAGATAGTGGTAAATTACAA-AATGAAAAAGA 59
             ||||| ||| |||| | ||| ||||| ||| ||| ||| ||| |||||
Sbjct: 114004 AAAGATAAAGATAAAG-ATAAA--GATAAAGAAAGATAAAGATGAAGATAAAGATAAAGA 114060

Query:     60 TAGTGGTAAATTACAA-AATGATAAAGATAGTGGTAAATTACAAAATGATAAAGATAGCG 118
             || | |||| | | ||| ||||| ||| ||| ||| ||||| ||||| |||
Sbjct: 114061 TAAAGATAAAGAAAGATAAAGATAAAGATAAAGATAGATAAAGAAA-GATAAAGATAA-G 114118

Query:    119 GTAAAT-TACAAA-ATGATAAAGATAGTGGTAAAT-TACA-A-AATGATAAAGTTAGTGG 173
             |||| | ||| ||| ||||| ||||| ||| ||| ||| ||||| ||| |||
Sbjct: 114119 ATAAAGATAAAGATAAAGATAAAGATAAAGATAAAGATGAAGATAAAGATAAAGATGAAGA 114178
```

Query: 174 TAAATTACAAAATGATAAAGATAGTGGTAAATTACAAA-ATG-ATAA-AGTTAGTGGTAA 230
 ||||| | ||| ||||| ||||| | ||||| | ||| ||||| ||| ||||| ||
 Sbjct: 114179 TAAAGTT--AAA-GATAAAGATAAAGATAAAG-ATAAAGATGCATGTGAGTG-GTGGCAA 114233

Query: 231 A 231
 |
 Sbjct: 114234 A 114234

>EM_OV:BX005461 BX005461.8 Zebrafish DNA sequence from clone DKEY-192L18 in
 linkage group 24
 Length = 162,216

Minus Strand HSPs:

Score = 359 (59.9 bits), Expect = 5.0e-08, P = 5.0e-08
 Identities = 155/235 (65%), Positives = 155/235 (65%), Strand = Minus / Plus

Query: 241 CATTTTGTAAATTTACCACTAACTTTAT-CATTTTGTGTA-ATTTACC-ACT-ATCTTTATCA 186
 ||||| | ||| | | ||||| ||||| | ||||| | | || ||||| ||
 Sbjct: 43534 CATTTT-TATTTTATATTTA-TTTATTCATTTT-TATATTTATTTATTCATTTTCA-CA 43589

Query: 185 TTTTGTAAATTTACCACTAACTTTAT-CATTTTGTAAATTTACCACTATCTTTATCATTTTG 127
 ||| | |||| | | | ||||| ||||| ||||| | || | |||||
 Sbjct: 43590 TTTATTTATTTTCATATTTA-TTTATACATTTTGTAAATTTATTTAT-TCATATTGATTAAT 43647

Query: 126 TAATTTACCGCTATCTTTATCATTTTGTAAATTTACCACTATCTTTATCATTTTGTAAATTT 67
 |||| | | | | ||||| |||| | |||| | | ||||| | || |||||
 Sbjct: 43648 TAATGTATCTATTTATTTAT-ATTTATTTATTCATTTTACATTTATT-TATT-TATTTTC 43704

Query: 66 ACCACTATCTTTTTCATTTTGTAAATTTACCACTATCTTTATCATTTTGTAAATTTA 12
 | ||| || ||||| ||||| ||||| | | || ||||| ||||| |
 Sbjct: 43705 ATATTTAT-TTATTCATTTTGTAAATTTATTTATTTATTCAT-ATTTATTAATTA 43757

>EM_INV:PFA929352 AL929352.1 Plasmodium falciparum strain 3D7, chromosome 5,
 segment 2/4
 Length = 313,050

Plus Strand HSPs:

Score = 354 (59.2 bits), Expect = 8.4e-08, P = 8.4e-08
 Identities = 158/234 (67%), Positives = 158/234 (67%), Strand = Plus / Plus

Query: 1 AAAGATAGTGGTAAATT--ACAAAA-TGATAAAGATAGTGGTAAATTACAA-AATGAAAA 56
 || ||| || |||| | |||| ||||| || | |||| | || ||||| ||
 Sbjct: 142762 AATAAATGATAAAAAACGATAAAAAATGATAAAAAATGATAA-AAATGATAAGAATGATAA 142820

Query: 57 AGATAGTGGTAAAT-TAC-AA-AATGATAAAGATAGTGGTAA-ATTACAA--AATGATAA 110
 | || ||||| || || || | || |||| | || || || ||||| |||||
 Sbjct: 142821 AAATGATGGTAAAAAATAATAAAGGTAATGATAAAAAATAATAATAGAGGTAATGATAA 142880

Query: 111 AGATAGCGGTAAATTACAAAATGATAAAGATAGTGGTAAATTACAAAATGATAAAGTTAG 170
 | ||| | |||| | |||| || |||| | |||| | || ||||| |||||
 Sbjct: 142881 AAATAATGATAAA--A-AAAATAATAGAGATAATGATAAA--A-ATAATGATAAAAAATAA 142934

Query: 171 TGGTAAATTACAAAATGATAAAGATAGTGGTAAATTACAAAATGATAAA-GTTA 223
 || |||| | ||||| ||||| || || || || || || || || || ||
 Sbjct: 142935 TGATAAA--A-AAAATAATAAAGATAATAATAA-TAAAAAGAAAAAAATGGTA 142984

Appendix D2 Blastp of a single nanomeric repeat of PfcGAK (first 10 pairwise alignments)

>gi|15927015|ref|NP_374548.1| hypothetical protein SA1267 [Staphylococcus aureus
 subsp. aureus N315]Length=6713

Score = 25.7 bits(53), Expect = 205 Identities = 7/7 (100%), Positives = 7/7 (100%),
 Gaps = 0/7 (0%)

Query 1 KLQNDKD 7
KLQNDKD
Sbjct 4608 KLQNDKD 4614

>gi|15924424|ref|NP_371958.1| hypothetical protein ebhA [Staphylococcus aureus subsp. aureus Mu50] Length=6713

Score = 25.7 bits(53), Expect = 205 Identities = 7/7 (100%), Positives = 7/7 (100%), Gaps = 0/7 (0%)

Query 1 KLQNDKD 7
KLQNDKD
Sbjct 4608 KLQNDKD 4614

>gi|57651379|ref|YP_186319.1|Cell wall associated fibronectin-binding protein [Staphylococcus aureus subsp. aureus COL]Length=10498

Score = 25.7 bits(53), Expect = 205 Identities = 7/7 (100%), Positives = 7/7 (100%), Gaps = 0/7 (0%)

Query 1 KLQNDKD 7
KLQNDKD
Sbjct 8519 KLQNDKD 8525

>gi|49483625|ref|YP_040849.1| very large surface anchored protein [Staphylococcus aureus subsp.aureus MRSA252]Length=10746

Score = 25.7 bits(53), Expect = 205, Identities = 7/7 (100%),Positives = 7/7 (100%), Gaps = 0/7 (0%)

Query 1 KLQNDKD 7
KLQNDKD
Sbjct 8641 KLQNDKD 8647

>gi|21283053|ref|NP_646141.1|hypothetical protein MW1324 [Staphylococcus aureus subsp. Aureus MW2] Length=9904

Score = 25.7 bits(53), Expect = 205, Identities = 7/7(100%), Positives = 7/7(100%), Gaps = 0/7 (0%)

Query 1 KLQNDKD 7
KLQNDKD
Sbjct 7876 KLQNDKD 7882

>gi|87202726|gb|ABD30536.1|conserved hypothetical protein [Staphylococcus aureus subsp aureus NCTC 8325]Length=9535

Score = 25.7 bits(53), Expect = 205, Identities = 7/7(100%), Positives = 7/7 (100%), Gaps = 0/7 (0%)

Query 1 KLQNDKD 7
KLQNDKD
Sbjct 7430 KLQNDKD 7436

>gi|70880228|gb|EAN93339.1|mucin-associated surface protein (MASP),putative Trypanosoma cruzi] Length=266

Score = 25.7 bits(53), Expect = 205, Identities = 7/7 (100%), Positives = 7/7 (100%), Gaps = 0/7 (0%)

Query 1 KLQNDKD 7
KLQNDKD
Sbjct 44 KLQNDKD 50

>gi|82656556|emb|CAI80978.1| truncated cell surface fibronectin-binding protein
[Staphylococcus aureus RF122]Length=1916

Score = 25.7 bits(53), Expect = 205, Identities = 7/7 (100%), Positives = 7/7
(100%), Gaps = 0/7 (0%)

Query 1 KLQNDKD 7
KLQNDKD
Sbjct 13 KLQNDKD 19

>gi|90428787|gb|EAS54227.1| YSIRK Gram-positive signal peptide:YD repeat
[Staphylococcus aureus subsp. aureus JH1]Length=10624

Score = 25.7 bits(53), Expect = 205, Identities = 7/7 (100%), Positives = 7/7
(100%), Gaps = 0/7 (0%)

Query 1 KLQNDKD 7
KLQNDKD
Sbjct 8519 KLQNDKD 8525

>gi|15235160|ref|NP_193708.1| hydrolase, hydrolyzing O-glycosyl compounds
[Arabidopsis thaliana]Length=332

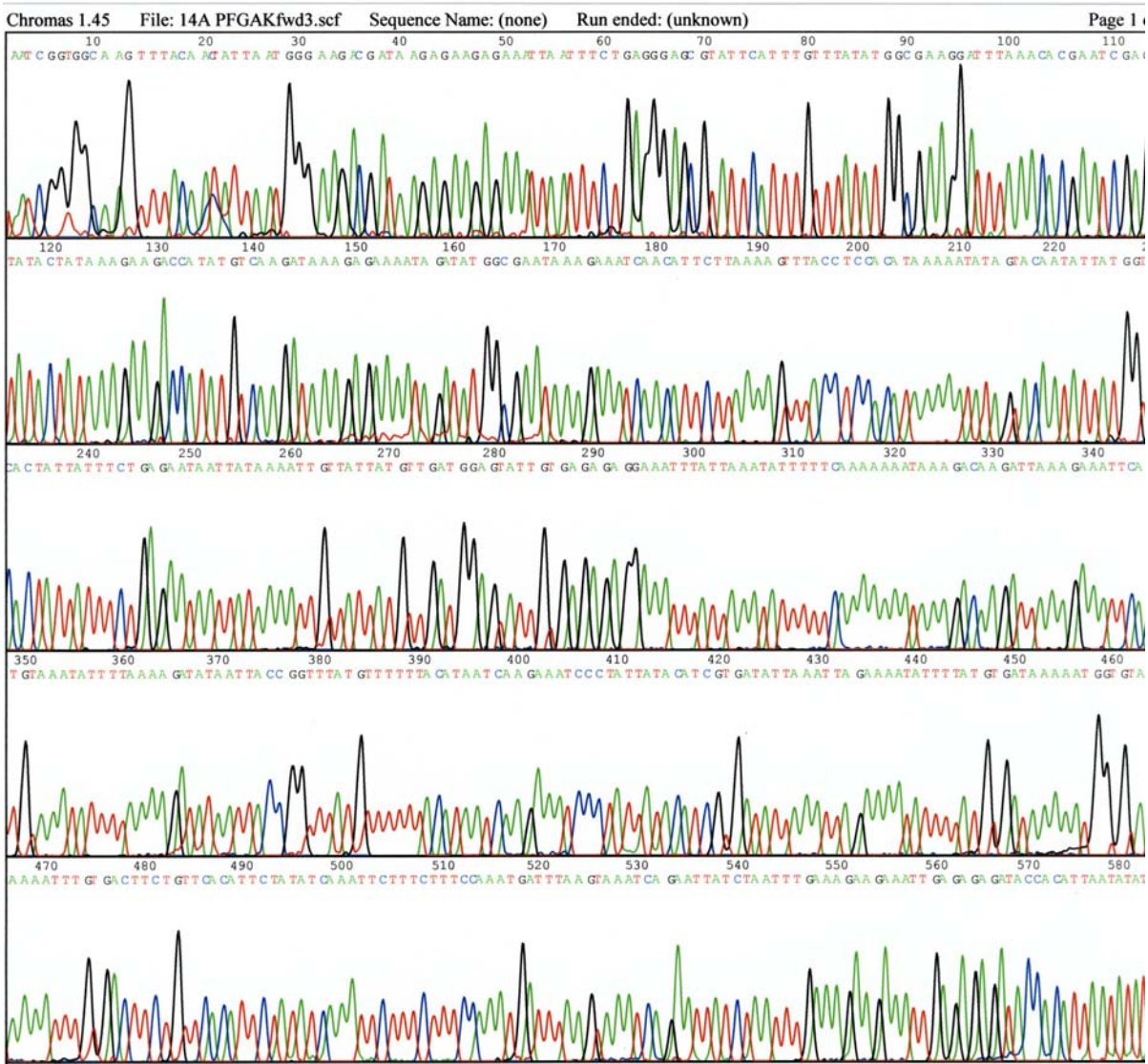
Score = 25.2 bits(52), Expect = 275, Identities = 7/8 (87%), Positives = 8/8 (100%),
Gaps = 0/8 (0%)

Query 2 LQNDKDSG 9
LQNDKD+G
Sbjct 258 LQNDKDTG 265

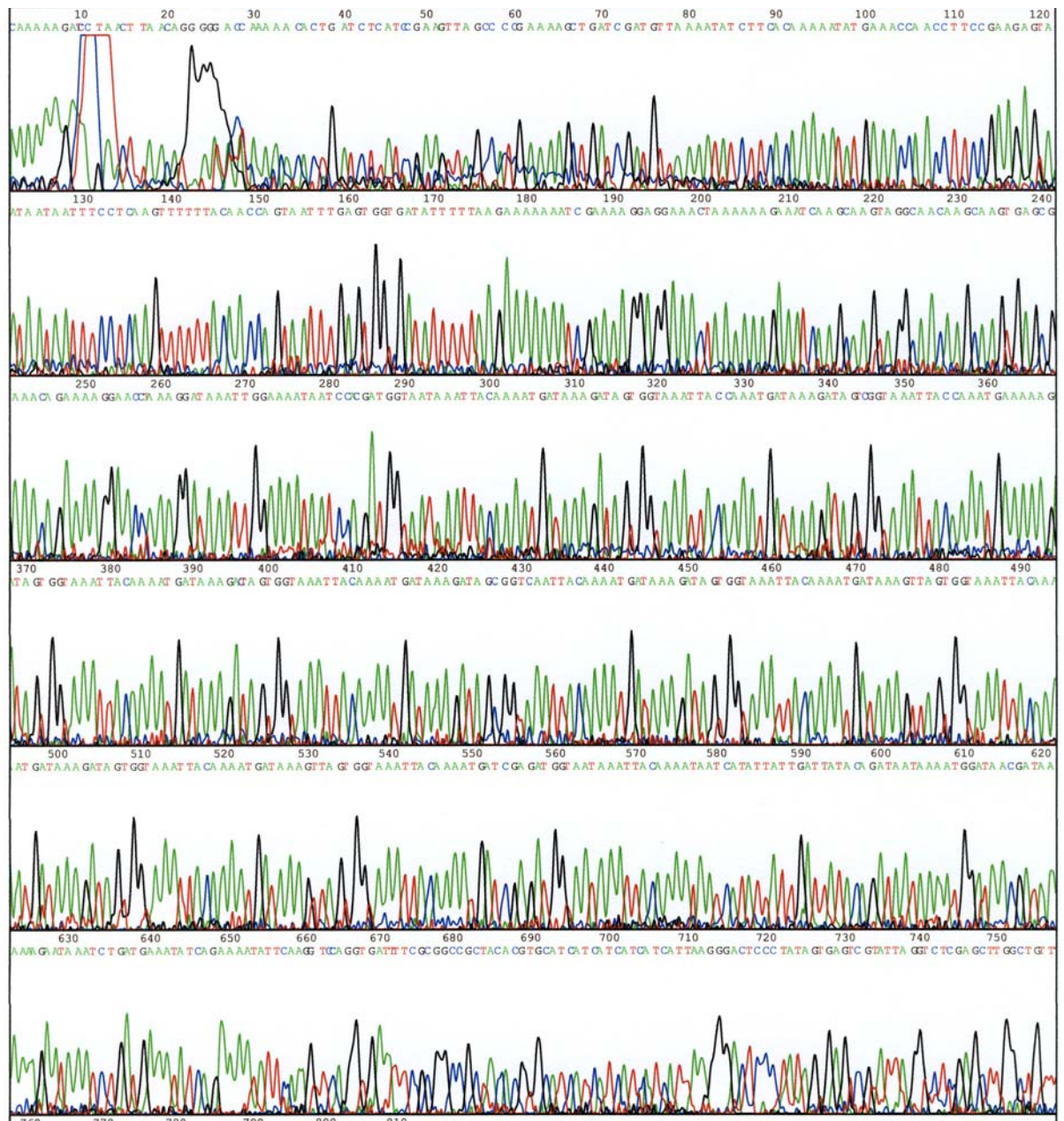
Appendix E Sequencing chromatograms

Appendix E1 PfcGAK sequencing chromatograms

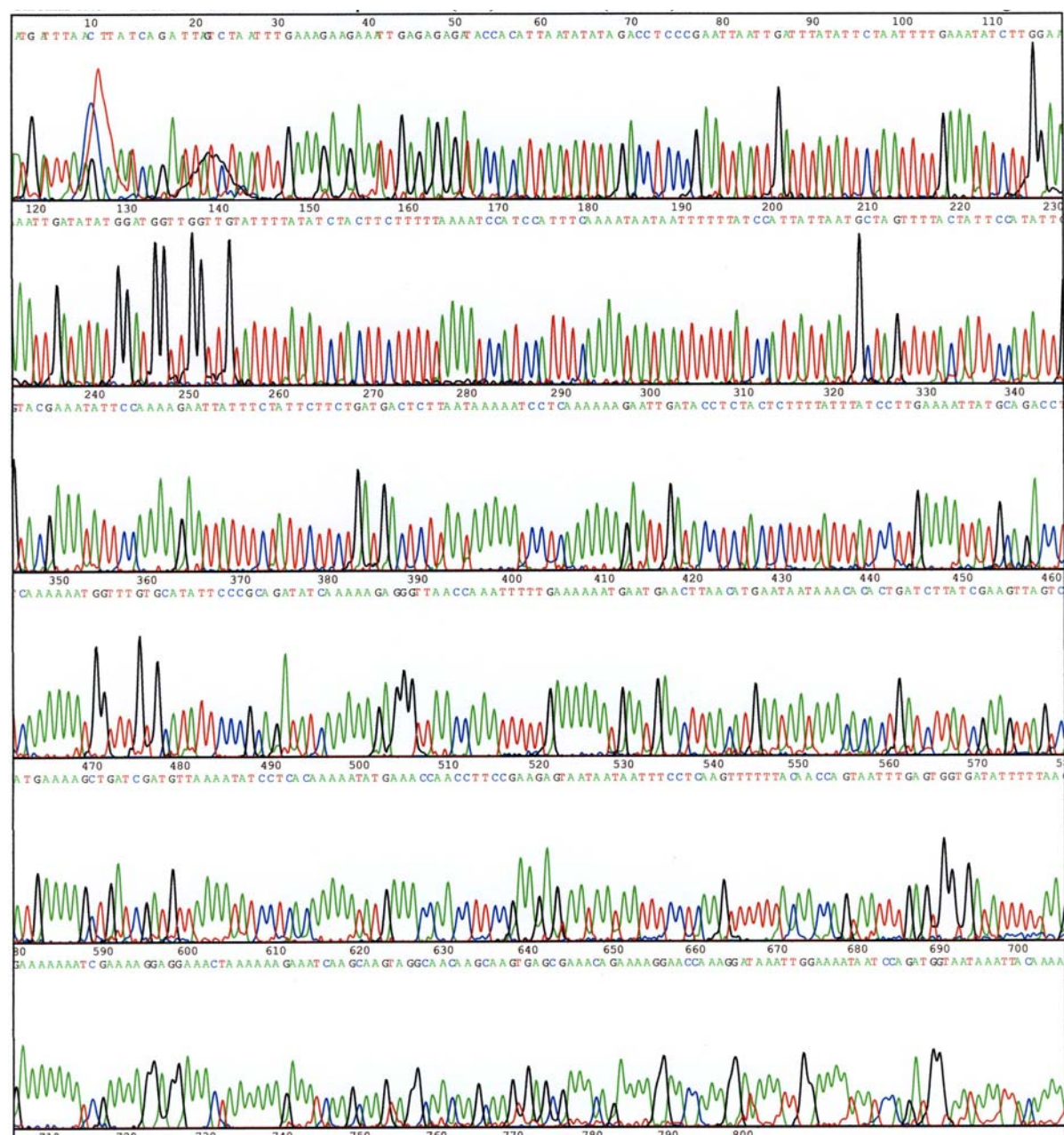
PfcGAK14: forward sequence



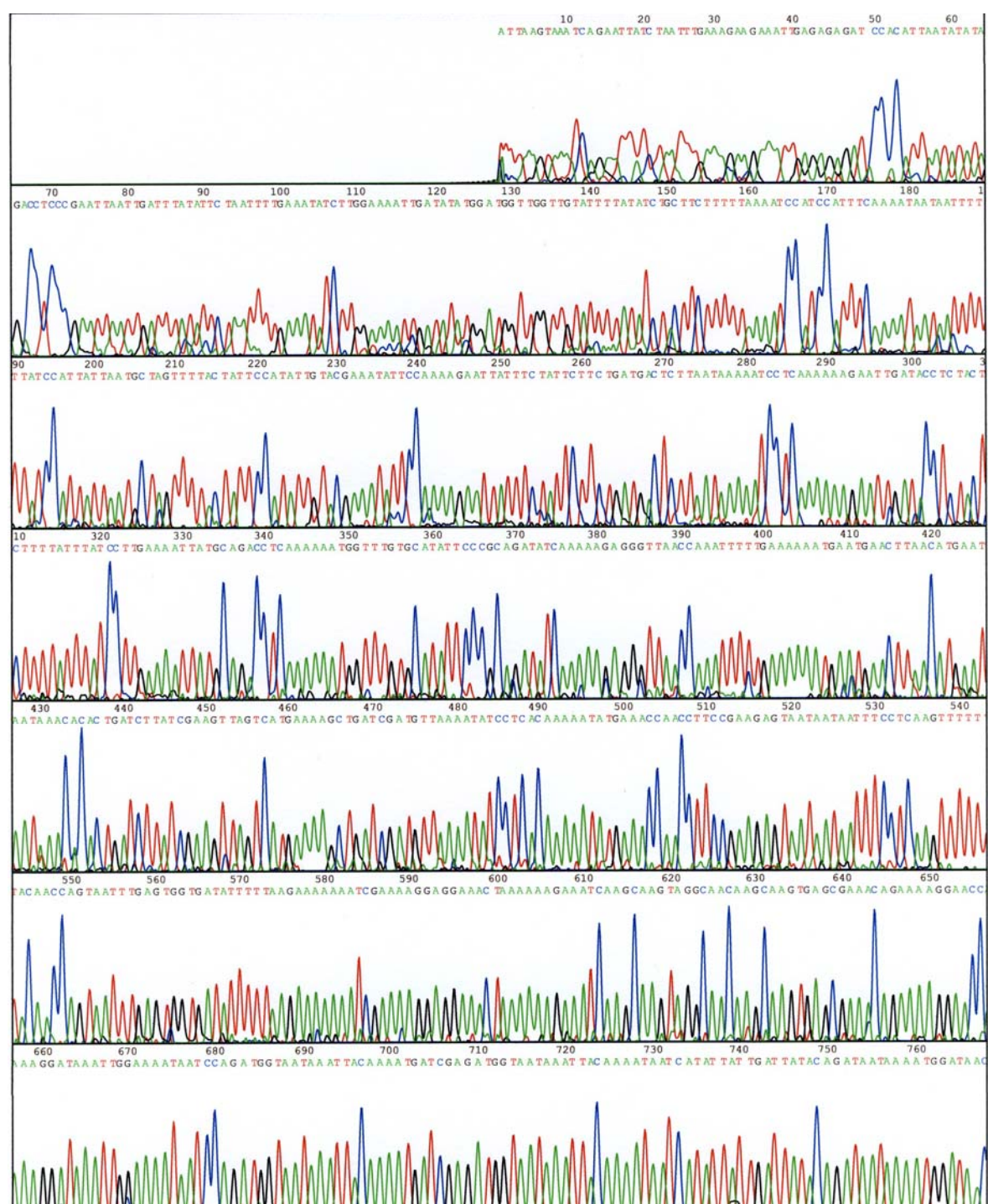
PfcGAK 14 : Internal forward 2 sequence



PfcGAK24 forward 1 sequence

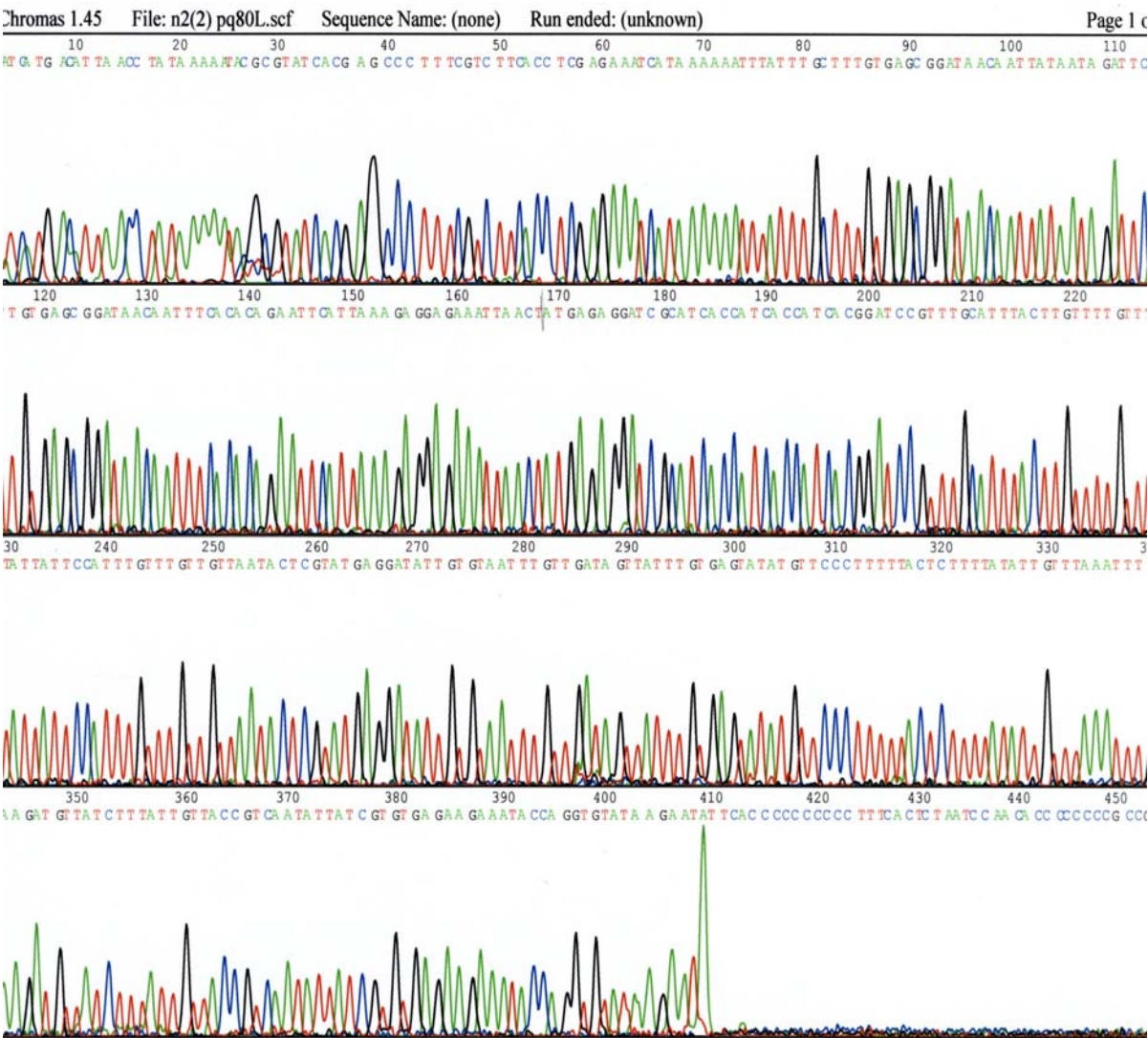


PfcGAK24 pHB6 reverse sequence (presented in forward orientation)

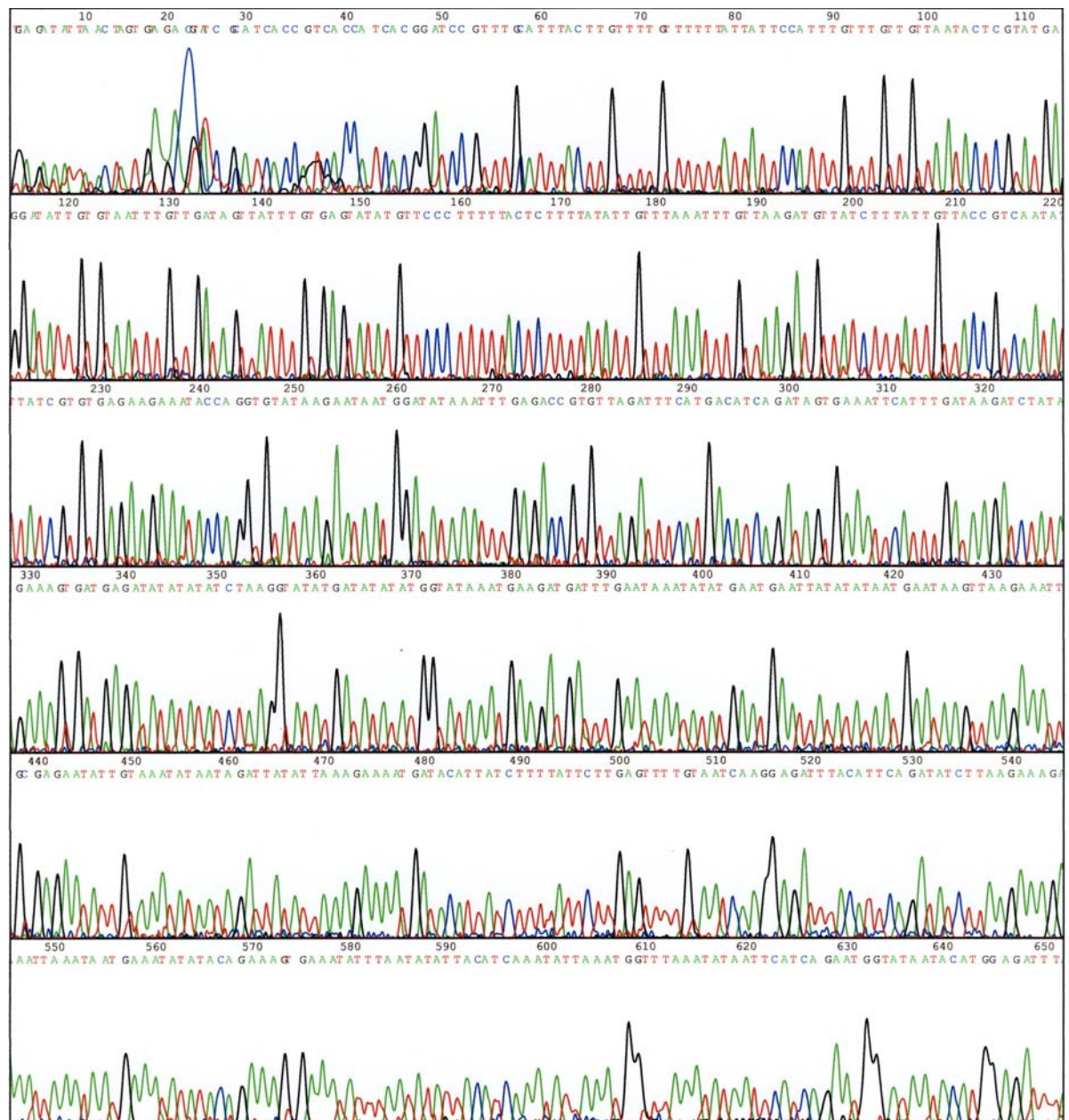


Appendix E2 Pfnek-3 sequencing chromatograms

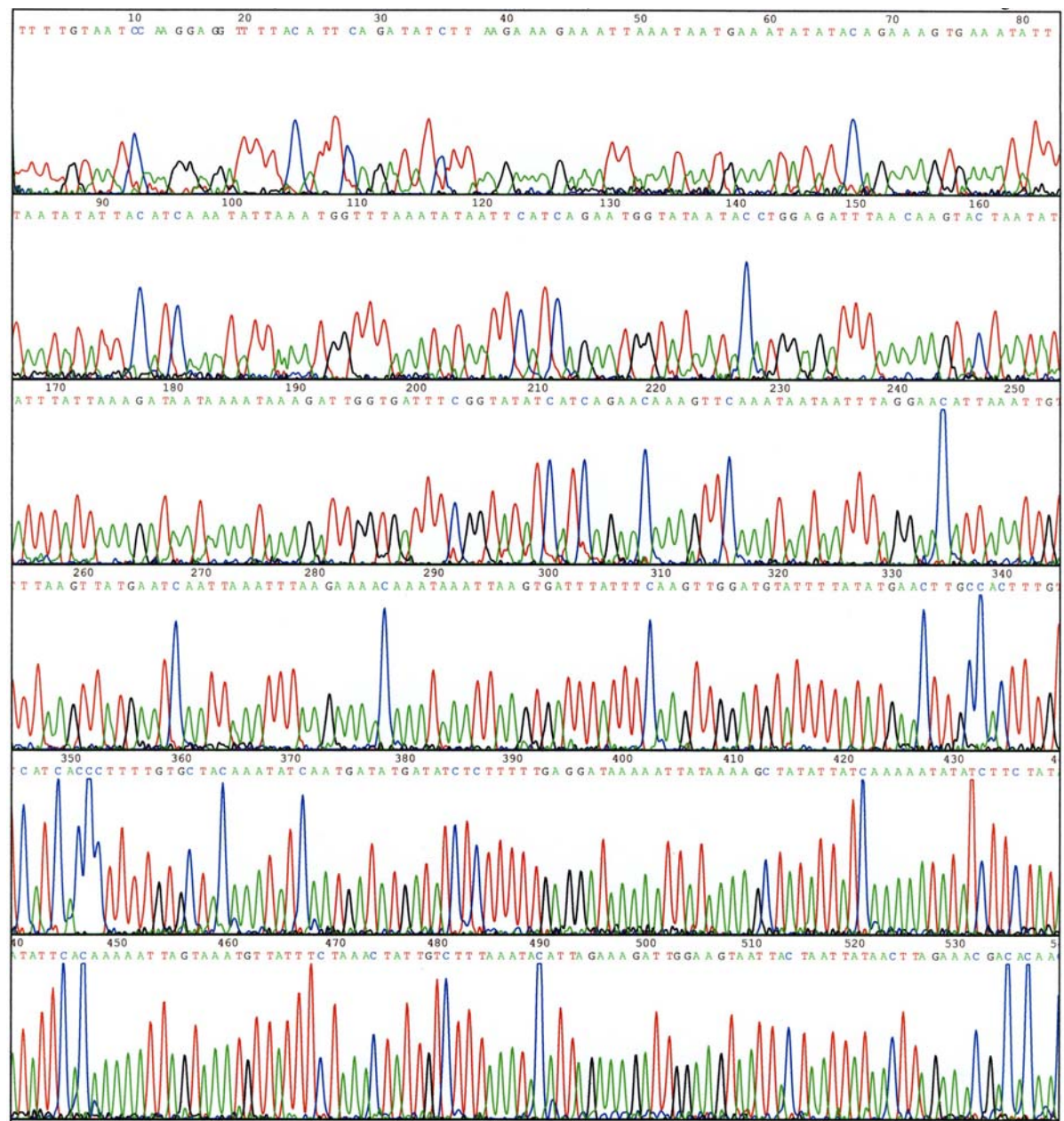
Pnek3 clone 2 forward, promoter primer



Pfnek-3 clone 2 : forward sequence

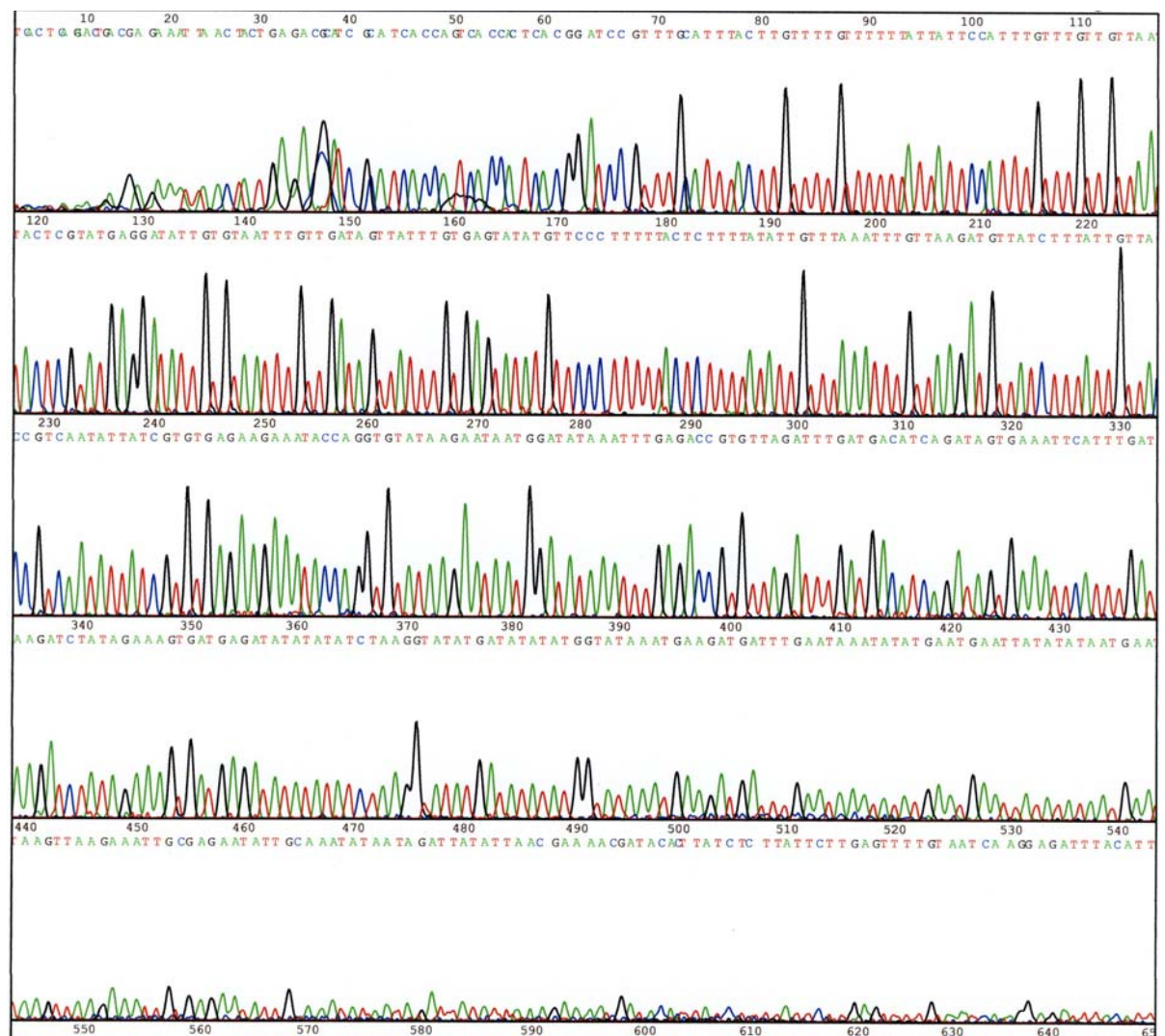


Pfnek-3 clone 2 : reverse sequence (presented in the forward orientation)



[illegible]

Pfnek-3 clone 4 : forward sequence



Pfnek-3 clone 4 : reverse sequence (presented in the forward orientation)

