MOLECULAR ANALYSIS OF INSECT STAGE

Trypanosoma brucei

Hachemi Zeraia

A thesis submitted for the Degree of Doctor of Philosophy University of Edinburgh – 2005

Declaration

I hereby declare that the research described within this thesis is my own composition, that the thesis is my own work and also certify that it has never been submitted for any other degree or professional qualification

University of Edinburgh
April 2005

Dedication

In memory of my Grandmothers, E. Djaballah and D.Djaballah

ACKNOWLEDGEMENTS

This research was funded by the Algerian government; work was carried out at the Centre of Tropical Veterinary Medicine, University of Edinburgh.

I would like to thank my family for their constant encouragement and belief in me, and of course Vicky. Also a special thanks to Mr Harchaoui who was instrumental in me coming to the United Kingdom

I would like to thank my supervisor Dr S.C. Welburn and Professor I. Mauldin for their constructive advice, discussion and guidance throughout this work. I must give a special mention to Kim Picozzi, Karine Delroux, Simon Young and Ewan MacLeod for their expertise in the laboratory, in addition to their help and encouragement during this project. I would also like to acknowledge the support of everyone else in the laboratory. Finally I would like to thank Ana, Goerge and Grigori.

Table of Contents **Page**

Declaration ii

Dedication iii

Acknowledgements........ iv

Table of Contents........ v

List of Figures........ xiii

List of Tables........ xv

Abbreviations........ xvi

Abstract xix

CHAPTER I
INTRODUCTION

1.1 Introduction 2

1.2 *T. brucei* life cycle. 5

1.3 Human African Trypanosomiasis........ 10

1.3.1 Clinical signs 10

1.3.2 Treatment........ 11

1.3.3 Animal Reservoir........ 11

1.3.4 Tsetse stage........ 12

1.4 RNA editing and trans-splicing........ 13

1.5 Cyclic nucleotides in eukaryotic cells 15

1.6 Reasoning for studying *T. brucei* procyclics........ 19

1.7 Aims of this project 19

CHAPTER 2
MATERIAL AND METHODS

2.1 Commonly used buffers 22

2.1.1 8-Br-cGMP Stock solution 22

2.1.2 Cunningham's medium 22

2.1.3 DNA loading buffer (6x 22

2.1.4 EthidiumBromide 23

2.1.5 Luria-Bertani (LB) medium 23

2.1.6 LB agar 23

2.1.7 PBS (1x) 23

2.1.8 RNA loading buffer (6x) 23

2.1.9 SOC medium 23

2.1.10 TAE (10x) 24

2.1.11 TBE(10x) 24

2.1.2 TE x) 24

2.2 Bacterial strains, ligation vector and vector details 25

2.3 Primers sequences 25

2.3.1 Primers sequences used in the PCR-Select™ cDNA Subtraction library 25

2.3.2 Anchored random primers used for low/high stringency PCR 26

2.4 Parasite Material 26

2.4.1 Bloodstream form preparation 26

2.4.2 Tsetse fly Infection 26

2.4.3 Preparation and maintenance of procyclic culture 27

2.4.4 Parasite Count ... 27
2.4.5 Treatment with 8-Br-cGMP ... 27
2.5 RNA methodologies ... 28
2.5.1 Total RNA extraction ... 28
2.5.2 DN*ase* Treatment of RNA ... 28
2.5.3 Poly A^{+} RNA isolation ... 29
2.5.4 RNA quantification ... 30
2.5.5 RNA Gel Electrophoresis ... 30
2.6 DNA ... 30
2.6.1 Gel electrophoresis ... 30
2.6.2 Gel extraction and purification ... 31
2.6.3 Quantification ... 32
2.7 Cloning and recombinant bacterial culturing ... 32
2.7.1 Ligation ... 32
2.7.2 Transformation ... 32
2.7.3 Overnight bacterial cultures and plasmid purification ... 33
2.7.3.1 Bacterial culture ... 33
2.7.3.2 Glycerol stock ... 33
2.7.3.3 Plasmid DNA Purification ... 33
2.8 DNA sequencing and sequence analysis ... 34
2.8.1 Preparation of DNA sequences ... 34
2.8.2 Blast Search ... 34

2.9 Characterisation of differentially expressed proteins on ProteinChipArrays 35

2.9.1 Method 35

2.9.2 Peak Detection 36

CHAPTER 3
INTRODUCTION

3.1 Introduction 38

3.1.1 Establishment 38

3.1.2 Maturation 39

3.1.3 Investigation Process 41

3.2 Material and methods 42

3.2.1 Dissection and collection of infected tsetse fly guts 42

3.2.2 Differential Display (DD) Methods 43

3.2.2.1 1cDNA Synthesis 43

3.2.2.1.1 First strand cDNA synthesis for DD 43

3.2.2.1.2 Double stranded cDNA synthesis 43

3.2.2.2 Rapid Amplification of Differential Expressed Sequences Poly chain reaction (RADES-PCR)- 45

3.2.2.3 Low/high Stringincy PCR 46

3.3 Result 47

3.3.1 Isolation of Total RNA from *Glossina morsitans morsitans* infected guts 47

3.3.2 Enrichment for *T.b. brucei* double stranded cDNA (dscDNA) synthesis from RNA of infected gut of *G.m. morsitans* using miniexon and oligodT primer sequence 48

3.3.3. RADES-PCR reactions were performed as described in section 3.2.2.2 52

3.3.3.1 Re-amplification and cloning of identified RADES-PCR product 54

3.3.3.2 Blast Search results of differentially expressed transcripts between group (++) and group (+-) 54

3.3.4 Amplification of differentially expressed sequences between group (++) and group (+-) using low / high stringency –PCR 57

3.4 Discussion 60

3.5 Future Work 62

CHAPTER 4
IDENTIFICATION OF GENES DIFFERENTIALLY EXPRESSED IN RESPONSE TO TREATMENT OF *T. BRUCEI* PROCYCLIC 8-BR-CGMP

4.1 Introduction 64

4.2 Material and Methods 64

4.2.1 Culture treatment 64

4.2.2 RNA extraction 65

4.2.2.1 Total and pol A^+ RNA extraction from TP and UP 65

4.2.3 Subtracted cDNA Library 65

4.2.3.1 First-Strand cDNA synthesis on poly A^+ mRNA extracted from treated and un-treated procyclic forms 65

4.2.3.2 Second-Strand cDNA Synthesis 66

4.2.3.3 Double strand cDNA purification 66

4.2.3.4 Rsa I digestion and purification of TP and UP dscDNAs 67

4.2.3.4.1 Introduction 67

4.2.3.4.2 Method 67

4.2.3.5 Ligation Adpators1 and 2R on dscDNA TP: 67

4.2.3.6 Ligation Efficiency Analysis 69

4.2.3.7 Hybridisations 71

4.2.3.7.1 Introduction 71

4.2.3.8 PCR Amplification 72

4.2.3.8.1 Introduction 72

4.2.3.8.2 Methods 75

4.2.3.8.2.1 First round PCR 75

4.2.3.8.2.2 Second round PCR 75

4.2.3.9 Semi-quantitative PCR analysis of subtraction efficiency 76

4.2.3.10 Construction of the subtractive library 76

4.2.3.11 PCR screening of recombinant clones 76

4.2.3.11.1 Introduction 76

4.2.3.11.2 Method 77

4.2.3.11.3 Real-time reverse-transcriptase Polymerase Chain Reaction [Real-time RT- PCR] 77

4.3 Results 78

4.3.1 Total and poly A^+ RNA extracted *T b brucei* treated procyclic forms with 8-Br-Cgmp 78

4.3.2 cDNA synthesis and *RSA* I digestion 79

4.3.3 Ligation Efficienc 80

4.3.4 Selective DNA amplification of differentially expressed transcripts in the Subtracted Treated Procyclic (STP) cDNA 82

4.3.6 Construction of the STP cDNA library 84

4.3.7 PCR screening of recombinant cloneS 85

4.3.8 Sequence analysis of differentially expressed cDNAs from 8 Br-cGMP treated *T. b. brucei* 85

4.3.8.1 Blast search homology 86

4.3.9 Relative Quantification of arginase transcript during treatment of *T.b.brucei* with 8-br-cGMP using real-time reverse transcriptase PCR (real time RT-PCR) 90

4.3.9.1 Confirmation of the purity of Arginase and GAPDH PCR products 92

4.3.9.2 Data analysis 93

4.4 Discussion 95

CHAPTER 5
PROTEOMIC ANALYSIS OF *T.B. BRUCEI* PROCYCLIC FORMS AFTER 8-BR CGMP TREATMENT

5.1 Introduction 103

5.2 Material and Methods 105

5.2.1 Culture treatment 105

5.2.1 Sample preparation 105

5.2.2 Surface-Enhanced Laser Desorption/Ionization Time-of-Flight Mass Spectrophotometer (SELDI-TOF- MS) 105

5.3 Results 106

5.3.1 SELDI-TOF-MS Proteins Analysis 106

5.3.1.1 External calibration 106

5.3.1.2 Optimisation and SELDI-TOF MS analysis 108

5.3.1.2.1 Optimisation of Protein profiling using WCX2 108

5.3.1.2.2 Optimisation of Protein profiling using SAX2 109

5.3.1.2.3 Optimisation of Protein profiling using H 50 111

5.3.1.3 Summary of SELTI-TOF MS optimisation 113

5.3.1 Identification of proteins differentially expressed during treatment of *T.b. brucei* with 8-Br-Cgmp 115

5.3.2.1 Protein design 115

5.3.2.2 Identification of differentially expressed proteins using H 50 ProteinChip.. 115

5.3.2.3 Identification of differentially expressed proteins using SAX 2 ProteinChip array surface 115

5.3.2.3.1 Data analysis 115

5.3.2.3.2 Data normalisation 116

5.3.2.4 Data base results 118

5.4 Discussion 119

5.5 Future Work 123

CHAPTER 6
GENERAL DISCUSSION

6.1 Objectives for this project 125

6.2 Experimental approach 125

6.1.1 RADES-PCR 125

6.1.2 Low/High Stringency PCR 126

6.1.3 Subtractive cDNA library 126

6.1.4 Real Time PCR 126

6.3 Future work 128

REFERENCES 131

APPENDIX I 149

APPENDIX II 152

APPENDIX III 203

LIST OF FIGURES

Figure 1.1 Morphology of Trypanosomes 3

Figure 1.2 Phylogeny of Salivarian trypanosomes 4

Figure 1.3 *T. brucei* life cycle. 6

Figure 1.4 The effects of 8-Br-cGMP on midgut infection rates of *T. b. brucei* in *G. m. morsitans*The effects on midgut infection rates of *T. b. brucei* BUT 135 in *G. m. morsitans* when the blood meal was supplemented with 8-Br-cGMP (Courtesy of Ewan MacLeod) 19

Figure 3.1 Schematic representation of Rapid Amplification of Differentially Expressed Sequence and low/high stringency PCR technique used during the investigation of differentially expressed genes in gut group (++) and (+-) 44

Figure 3.2 *Glossina T.b .brucei* strain BUT 135 infected midgut RNA gel electrophoresis. 47

Figure 3.3 Gel electrophoresis of dscDNA obtained with total RNA extracted from guts group (++) and gut group (+-) 49

Figure 3.4 Gel electrophoresis of dscDNA obtained with total RNA extracted from control guts (group --) 50

Figure 3.5 Gel electrophoresis of the RADES-PCR products. 51

Figure 3.5.1 Gel electrophoresis of low/High stringency DD-PCR products obtained on cDNA prepared on RNA extracted from gut group (++) and gut group (+-) 57

Figure 4.3 PCR amplification of type e hybrid molecules. 74

Figure 4.5 Gel electrophoresis of TP cDNA obtained from 2μg of TP poly A^{+}RNA 80

Figure 4.6 ligation efficiency of Adaptor1 and 2R on *RSA* I digested dscDNA TP 81

Figure 4.7 Gel electrophoresis of 2 round PCR products obtained from subtracted treated procyclic from (STP) and unsubtracted treated procyclic form samples.......... 83

Figure 4.8 Semi-quantitative PCR analysis of hybridization efficiency of subtracted treated procyclic sample. 84

Figure 4.9 PCR screening of 29 recombinant bacterial colonies using nested primer 1 and 2R 85

Figure 4.10 Derivative melting curve plots of Arginase and GAPDH amplicons ... 92

Figure 4.11 Calculation of real-time efficiency rates of arginase and GAPDH amplification 94

Figure 5.1 External Calibration of PBS II Reader 107

Figure 5.2 SELDI-TOF MS data obtained with WCX 2 ProteinChip array loaded with a *T.b. brucei* protein extract at pH 4.0 and at pH 6.0. 109

Figure 5.3 SELDI-TOF MS data obtained with a SAX 2 ProteinChip array loaded with *T b brucei* protein extract 110

Figure 5.4 SELDI-TOF MS data obtained with H 50 array surface with *T.b. brucei* protein extract. 112

Figure 5.5 SELDI-TOF MS data from different array surfaces loaded with *T b brucei* protein extracts. 113

Figure 5.6 SELDI-TOF MS data obtained on SAX with proteins extract from treated and untreated *T.brucei.* 117

Figure 5.7 Biomarker identification wizards used over the protein region between 24-35 kDa. 118

LIST OF TABLES

Table 3.1 1 Amplification steps used for the synthesis dsc DNA from group (++, +-, --) 45

Table 3.2 Amplification steps used for RADES 45

Table 3.3 Amplification steps used for low/high PCR 46

Table 4. 1 PCR amplification parameters used to asses the ligation efficiency of Adaptors 1and 2R on TP sample 70

Table 4. 2 PCR parameters used for the selective amplification of hybrids e 75

Table 4. 3 PCR amplification parameters used for hybrids e enrichment 75

Table 4. 4 PCR amplification used during the semi-quantifcation analysis of the substraction efficiency 76

Table 4. 5 PCR parameters used for bacterial screening 77

Table 4. 6 PCR amplification parameters used for Real-time RT-PCR 78

Table 4. 7 The C_T obtained from TP and UP of arginase and GAPDH amplicons 93

Table 5.1 Comparison of MW of protein standards provided by manufacturer and MW of protein standards obtained with PBS II 106

Table 5.2 Peak proteins detected with 3 different proteinChips surface 114

Table 6.1 1 Merits of methods used in this thesis 128

ABBREVIATIONS

aa	Amino acid
ADK	Adenylate kinase.
ATP	Adenosine triphosphate
bp, kbp	Base pairs, kilo base pairs
BSA	Bovine serum albumin
cAMP	Cyclic adenosine mono phosphate
cDNA	Complementary deoxyribonucleic acid
cGMP	Cyclic guanosine mono phosphate
CK1	Casein Kinase 1
dATP	Deoxyadenosine 5'-triphosphate
dCTP	Deoxycytidine5'-triphosphate
dGTP	Deoxyguanosine 5'triphosphate
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxynucloetide 5' triphosphate
dscDNA	Double stranded complementary deoxyribonucleic acid
DTT	Dithiothreitol
dTTP	Deoxythymidine 5'triphosphate
dUTP	Deoxyuracile 5'triphosphate
EAM	Enhancing Absorbing Molecule
EDTA	Ethylene diamine tetra-acetic acid
ES	Ectoperitrophic space
EtBr	Ethidium bromide
FBS	Foetal bovine serum
g	Relative centrifugal force
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase
GC	Guanyl cyclase
IPTG	Isopropyl-ß-D-thio-galactopyranoside

kDa	**kilodalton**
LB	**Luria Bertoni broth**
MALDI-ToF	**Matrix Assisted Laser Desorption /Ionization- Time of Flight**
MS	**Mass Spectrometry**
MW	**Molecular weight**
nm- μm-mm	**Nanometre-micrometre-millimetre**
NO	**Nitric oxide**
OD	**Optical density**
PBS	**Phosphate buffered saline**
PCD	**Programmed cell death**
PCR	**Polymerase chain reaction**
PDE	**Phosphodiesterase**
PKA	**Protein Kinase A**
PKG	**Protein Kinase G**
Poly A	**Polyadenylic acid**
PM	**Peritrophic membrane**
RADES	**Rapid amplification of differentially expressed sequences**
RNA	**Ribonucleic acid**
rRNA, tRNA	**Ribosomal ribonucleic acid, transfer ribonucleic acid**
rpm	**Revolutions per minute**
RSU	**Regulatory sub unit**
RT-PCR	**Reverse transcription-polymerase chain reaction**
s-min-h	**Second-minute-hour**
SAGE	**Serial analysis of gene expression**
SELDI-ToF	**Surface Enhanced Laser Desorption/Ionisation-Time of Flight**
SL	**Spliced leader**
sscDNA	**Single stranded complementary deoxyribonucleic acid**
STP	**Subtracted treated procyclics**
TBS	**Tris-buffered saline**
TP	**Treated procyclics**

TXN	**Trypanodoxine peroxidase**
mU - U	**MilliUnit - Unit**
μA- mA	**Microampere - milliAmper**
μg – mg - g	**Microgram – milligram -gram**
μl - ml - l	**Microlitre - millilitre - litre**
μM - mM- M	**Micromolar – millimolar - molar**
UP	**Untreated procyclics**
UTP	**Unsubtracted treated procyclics**
v/v	**Volume/volume**
V	**Volt**
VAT	**Variant antigen type**
VSG	**Variable surface glycoprotein**
w/v	**Weight/volume**
W	**Watt**
WHO	**World Health Organisation**
X-Gal	**5-Bromo-4-chloro-3-indolyl-β-D-galactopyranoside**

ABSTRACT

This project used both a genomic and proteomic approach to investigate the molecular and biochemical changes that occur in the insect stage of *Trypanosoma brucei.* The project had three main aims, the first of which was the application of an already established approach RADES-PCR to investigate genes expressed during the maturation of established mid-gut infections. The method consisted of two steps, 1) reverse transcription of first strand DNA using oligo (dT) strech and 2) enrichment of parasite material over host material using the splice leader (SL) the miniexon. It was found that the high prevalence of artefacts and the presence of bacterial material from symbionts in the tsetse midgut hampered this investigation. The application of a low and high stringency approach using random anchored primers did not improve the situation. This feature of *T. brucei* maturation could therefore not be investigated further by this approach.

Our second aim was to investigate the effect that 8 Br-cGMP has on *T. brucei* procyclic forms. It had been observed in our laboratory that the addition of this chemical to an infected bloodmeal increased the level of *T. brucei* establishment within the fly population to 100%. An enriched subtractive cDNA library was made to study the effect that this compound has on *T. brucei*, this library provided a full set of the differentially expressed genes resulting from this 8 Br-cGMP treatment. Analysis of over 1000 recombinant clones highlighted the involvement of the polyamine and trypanothione synthesis pathway, in particular the potential role of the antioxidant response.

Our third aim was to investigate the protein profile of *T. brucei* in response to this treatment. Protein profiles were characterised using surface enhanced laser desorption/ionization time of flight mass spectrophotometry (SELDI-TOF-MS) where proteins are separated according to their biochemical properties. This approach revealed a number of potential differences that resulted in response to the treatment. Therefore there is a response at the proteomic level that may reflect those genes identified by the subtractive library.

Chapter 1

Introduction

1.1 Introduction

Kinetoplastids are a diverse group of organisms that includes obligatory parasites that infect all classes of vertebrates, plants and some insects (Maslov *et al.,* 2001). Their influence is most profoundly felt in sub-Saharan Africa where over 10 million km^2 (around one third of the continents land mass) are affected. Both human and animal diseases in Africa are caused by parasitic kinetoplastid trypanosomes.

Kinetoplastids are a group of protozoa characterised by the presence of a specific organelle, the kinetoplast (Figure 1.1), which contains the mitochondrial DNA (Vickermann and Preston, 1976). Kinetoplastids are subdivided into two suborders Bodonina and Trypanosomatina (Vickermann and Preston, 1976, Lom 1976), the first suborder is formed by two families, Bodonidae and Cryptobiidae (defined by the presence of a large kinetoplast and two flagella); the second is represented by a single family, Trypanosomatidae (single flagellum and a small kinetoplast). The species from the genera *Trypanosoma, Leishmania and Endotrypanum* have a digenic lifestyle, which involves cyclical transmission through an insect vector and vertebrate host (Vickermann, 1994, Maslov *et al.,* 1994, Lukes *et al.,* 1997).

The origin of the evolution of parasitism in kinotoplastids has developed around two main hypotheses. The first being the "Invertebrate first" theory proposed by Leger in 1904 (Cited in Maslov *et al.,* 2001), this claimed that parasitism in the Kinetoplastids group began in the digestive tract of insect and leeches from where the hemoparasites of vertebrates evolved later. The second hypothesis "Vertebrate first" was put forward by Minchin in 1908 (Cited in Maslov *et al*., 2001) and claimed that parasitism originated first in the vertebrate gut and hemoparasites of vertebrates evolved from their endoparasites. Phylogenetic analysis based on the sequence of the small subunit (*18S*) of ribosomal RNA has sustained both hypotheses (Maslov *et al.,* 2001).

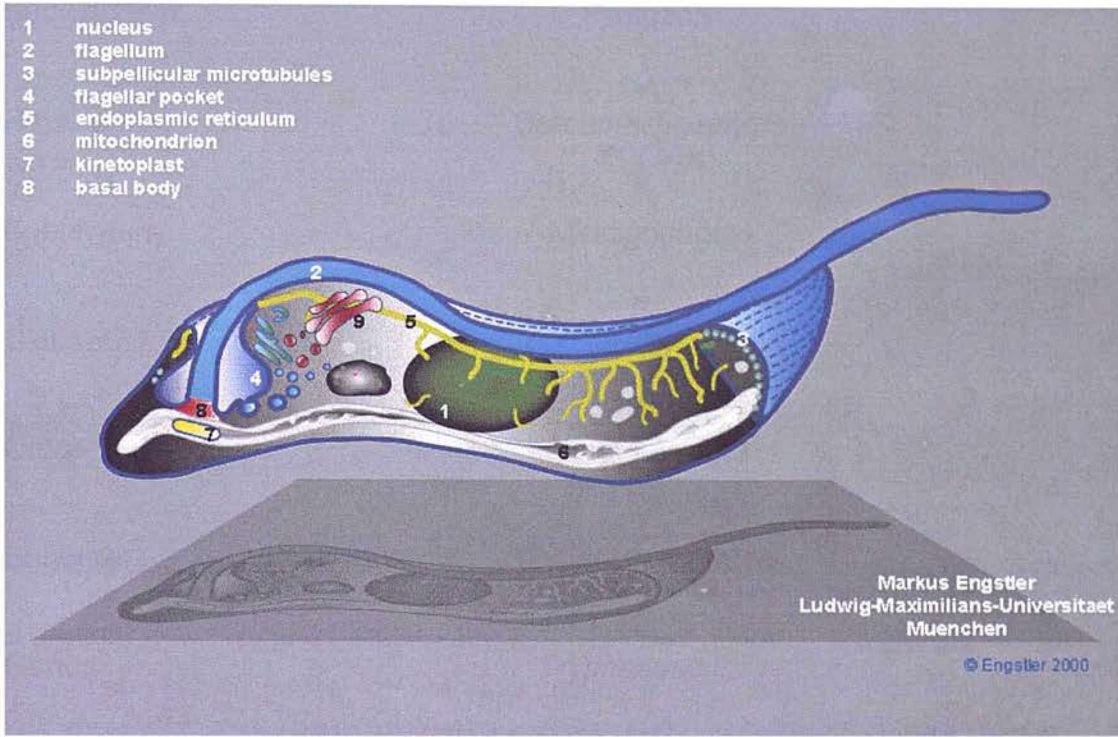


Figure 0.1 **Morphology of Trypanosomes**

The genus *Trypanosoma* is subdivided into two sections (Stercoraria and Salivaria), defined by the manner in which the infective form leaves their insect host. Stercoraria are excreted in the faeces of the insect, while Salivaria are delivered during the feeding of the insect on the vertebrate host. The Salivaria group contains four subgenera; *Dutonnella*, *Nannomonas*, *Trypanozoon* and *Pycnomonas* (Figure 1.2). These trypanosomes are pathogenic unicellular parasites that live in the blood and intracellular compartments of their host, causing disease recognized as nagana in livestock.

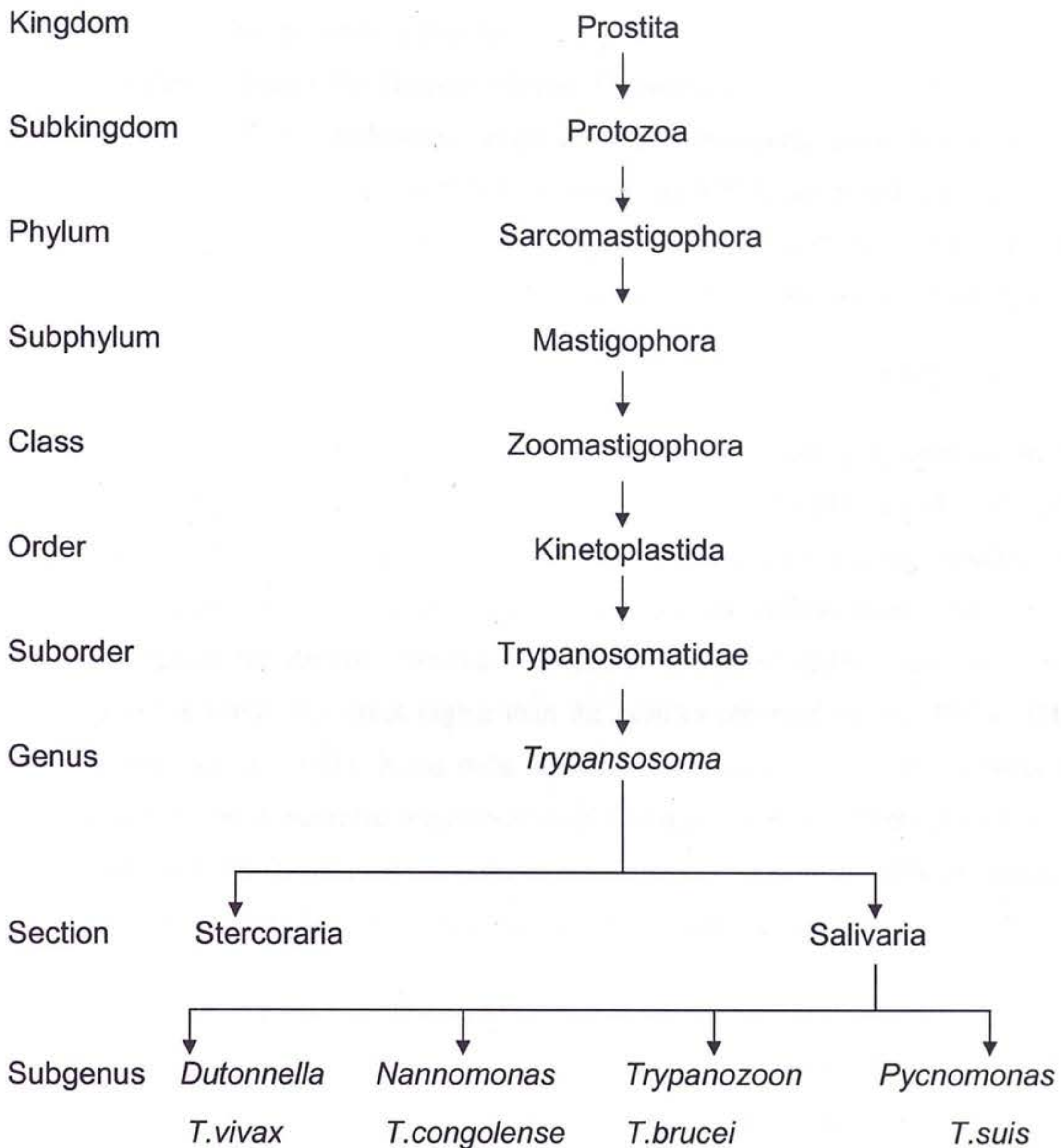


Figure 0.2 **Phylogeny of Salivarian trypanosomes**

The subgenus *Trypanozoon* is represented by the *brucei* group which compromises three members *T. brucei* s.l., *T. evansi* and *T. equiperdum*. *T. brucei* s.l. are transmitted *via* the bite of an infected tsetse fly (*Glossina* spp.) whereas *T. evansi* is transmitted by blood-sucking insects from the Tabanid family and *T. equiperdum* is sexually transmitted and restricted to equine hosts. Within the *T. brucei* s.l. species there are a further 3 sub-

species, *T. brucei brucei*, *T. b. gambiense* and *T.b. rhodesiense*. These parasites are asymptomatic within the animal host however the latter two members are zoonotic and are the causative agents for Human African Trypanosomiasis (HAT) also known as sleeping sickness. *T. b. gambiense* is responsible for the chronic form of HAT mainly found in West and Central Africa from Ethiopia to and NW Uganda and are transmitted by the *palpalis* group (*Glossina morsitans*). *T. b. rhodesiense* is the acute form of HAT found in East Africa and transmitted by tsetse flies of the *morsitans* group (e.g. *G. palpalis*).

Human African trypanosomiasis (HAT) is considered a re-emerging disease. In the 1960's the disease was controlled, however in the last few decades the World Health Organization (WHO) estimated that more than 60 million people are continuously exposed in more than 20 African countries; yet just 3-4 million people are actively protected against the disease. Between the years 1998-1999 45,000 new cases were reported to the WHO, ten times higher than the number recorded in the 1960's (WHO 1998, Cattand *et al.,* 2001). It has been estimated that some 45-50 million cattle are under risk of tsetse transmitted trypanosomiasis (Kristjanson *et al.*, 1999). As a disease, trypanosomiasis has great socio-economic consequences with a 20-40% reduction in livestock productivity due to forced grazing in areas of low tsetse risk.

1.2 *T. brucei* life cycle

During feeding on a mammalian host, trypanosome-infected tsetse flies can inoculate infective metacyclic trypanosomes into the host, inducing a localized skin reaction (chancre). Injected trypanosomes multiply within the chancre for several days before passing initially into the lymphatic system and then into the blood where they transform into dividing long slender forms. In order to evade the host immune response invading trypanosomes express a unique surface coat consisting of a single variant surface glycoprotein (VSG). VSGs envelop the entire surface (including the flagellum) of bloodstream and metacyclic trypanosomes (Vickermann and Preston, 1976). To evade host immunity metacyclic forms express metacyclic VSGs (MVSGs) for as long as 7

days post inoculation, selecting from a repertoire of 10-20 different VSG genes (Leonardo *et al.*, 1984; Turner *et al.*, 1988).

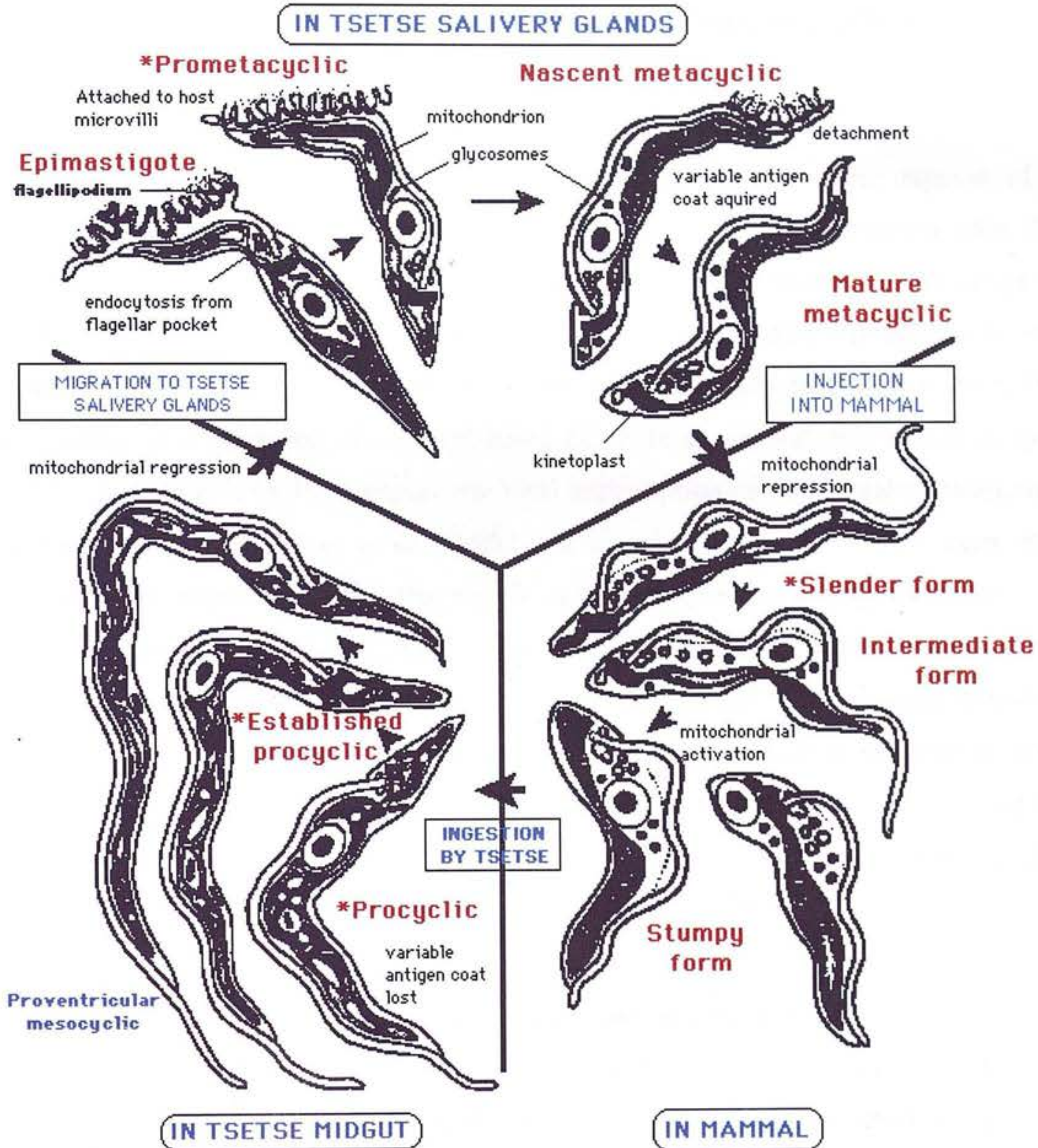


Figure 0.3 ***T. brucei*** **life cycle.**

As metacyclic trypanosomes switch into long slender forms, the repertoire of available bloodstream VSGs increases up to 1000 (Van der Ploeg *et al.*, 1982). During the ascending parasitaemia predominated by long slender dividing forms, host immune responses ultimately recognise the actively expressed VSG and clear the majority of the parasites. Within this population, some trypanosomes activate a different VSG type, survive and consequently generating a new wave of parasitaemia.

Active VSGs are expressed from localised sites near the telomeric regions of the trypanosome chromosomes and associated with specific gene expression sites (ES), polycistronic units stretching from the VSG promoter to telomeric repeats (Pays and Nolan 1998; Pedram and Donelson 1999). These telomeric-linked bloodstream ES units span 45-60 kb and are preceded by 20-50 kb of 50 bp repeats that are not transcribed. Beginning at a promoter the bloodstream ES possess several ES-associated genes (ESAGs), 5-20 kb of a 70 bp repeat, the VSG transcription unit and finally the telomeric repeat sequence (Donelson *et al.*, 2003). Several VSG expression sites exist in the trypanosome genome though only one is actively expressed at any one time. The switching mechanism of VSG genes in and out of the ES may be performed by different molecular mechanisms such as duplicative transposition, the most common hypothesis for VSG switching where an expressed telomeric VSG sequence is replaced by a new VSG gene (Robinson *et al.*, 1999). It has been speculated that several factors could be at the origin of VSG switching, a candidate being RAD51, an enzyme involved in DNA break repair (McCulloch and Barry, 1999; Donelson *et al.*, 2003).

During the ascending parasitaemia in the mammalian host, a small portion of proliferative long slender forms transform into a non-proliferative short stumpy form adapted to living and developing in the tsetse fly. The transformation of long slender (LS) to short stumpy (SS) forms is controlled by the size of the trypanosome population, mediated by parasite density and triggered by the release of a stumpy induction factor (SIF) through the c-AMP transduction pathway (Reuner *et al.*, 1997; Vassella *et al.*,

1997). The transition from the SL to SS form is irreversible and as the pre-adapted form for the insect stage, the SS forms die unless taken up during feeding by tsetse flies on the infected host.

Several factors exist that can induce the differentiation of mammalian bloodstream trypanosomes to insect forms *in vitro*. The presences of cis-aconitate, citrate or a drop in temperature (37°C to 27°C) have been found to trigger the switch (Hunt *et al.*, 1994). Upon uptake by the tsetse fly in a blood meal, differentiation of bloodstream forms to insect forms (procyclics) sees the shedding of VSG coat, thought to be undertaken by a specific metalloendoprotease (**Bangs *et al.*, 1997). The VSG coat is replaced by an insect form specific surface coat molecule known as procyclin** (Roditi *et al.*, 1987; Mowatt and Clayton, 1987). Procyclin is thought to play several important roles during the transformation of trypanosomes in the tsetse fly, primarily in the protection of the parasites against midgut proteases (Acosta-Serrano *et al.*, 2001), though it has also been suggested that procyclin plays a role in parasite differentiation and survival within the tsetse gut (Pearson *et al.*, 2000). Procyclin was also demonstrated to have a physical role in preventing adherence of the parasites to the midgut epithelium, aiding the transition of procyclic trypanosomes in completion of the life-cycle within the fly (Roditi *et al.*, 1990). The deletion of genes encoding for procyclin revealed that the surface coat protein aids the establishment of heavy parasite infection in the fly midgut (Ruepp *et al.*, 1997). Two forms of procyclin exist, classified according to internal repeats, EP procyclin (30 tandem repeats of glutamic acid-proline) and GPEET procyclin (five or six pentapeptide repeats-gly–pro–glu–glu–thr). Procyclin is attached to the membrane by a complex glycosylphosphatidylinositol (GPI) anchor (Ferguson *et al.*, 1993; Bütikofer *et al.*, 1997; Treumann *et al.*, 1997). Initially both procyclin molecules are expressed simultaneously in the tsetse gut, however the expression of GPEET procyclin ceases as the procyclic parasites establish themselves within the midgut (Vassela *et al.*, 2000). The classical morphological event characterising the transition from bloodstream into procyclic insect form is the repositioning of the trypanosome kinetoplast from an extreme posterior position in the bloodstream form into

a position midway between the nucleus and posterior in procyclic forms (Vickermann, 1985). During the mammalian stage, bloodstream form trypanosomes use glucose as a primary energy source since it is abundant in the host, however as they make their transition into fly midgut forms, glucose becomes scarce and proline, an abundant amino acid in the fly is as utilised the main source of energy (Bursell *et al.*, 1973).

Several days after uptake of an infective blood meal by the fly, parasites and digested blood are confined in the lumen of the gut bordered by the peritrophic membrane (PM). Six days post-infection, parasites are found mainly in the anterior portion of the midgut where they take refuge in the ectoperitrophic space (ES), located between the PM and the tsetse gut wall. Two routes of migration are possible by which trypanosomes reach the ES, either by an unlikely direct penetration through the PM (Evans *et al.*, 1983; Ellis *et al.*, 1977) or more likely through the opening at the posterior end of the midgut. Between nine and fortheen day's post-infection trypanosomes are present in the proventriculus, in the middle portion of the midgut and in the ES (Gibson and Bailey, 2003). During colonisation of the ES, procyclic trypanosomes migrate toward the anterior portion of the midgut near the proventricule, becoming more elongated and transforming into the mesocyclic forms that represent the final stage of established trypanosomes in the tsetse midgut. These parasite forms are non-dividing and are halted in the G0/G1 phase of the cell cycle, before regaining dividing status in response to extracellular factors present in the anterior portion of the midgut. In the final stage of the trypanosome life cycle in tsetse, mesocyclic forms migrate from the oesophagus to the mouthparts, then through the salivary conducts to reach the salivary glands. During their migration, mesocyclic forms undergo complex differentiation to become epimastigotes. Upon reaching the salivary glands, epimastigotes attach to the salivary gland epithelium using outgrowths from the flagellum called flagellipodia. Epimastigotes then differentiate into pre-metacyclics that have the ability to proliferate and later differentiate to non-proliferative mammalian infective forms (Tetley and Vickermann, 1985). These mature metacyclics have a VSG surface coat and are pre-adapted to the relatively rich glucose environment in the mammalian bloodstream.

1.3 Human African Trypanosomiasis

Human African trypanosomiasis is a fatal progressive disease in the absence of treatment. The disease occurs in two stages, the first haemolymphatic stage is identified by the presence of parasites within the blood, the second by their migration into the central nervous system (CNS). Of the two parasites that cause this disease, *T.b. gambiense* has a pathognomonic symptomatic stage followed by chronic meningoencephalitis with death occuring after several years.

1.3.1 Clinical signs

At the site of a tsetse fly bite a chancre appears 4-14 days after inoculation of parasites, a region of swelling in the subcutaneous tissue that appears as a hard red nodule. It has been noticed that the chancre is more common in the case of the acute form of sleeping sickness. The inoculated parasites migrate toward the lymphatic and blood circulation, their presence causing intermittent episodes of high fever in response to waves of high parasitaemia. The spleen increases in size due to anaemia and the heart, the organ at most risk, becomes damaged and stressed (Apted, 1970). Skin lesions can also appear in fair skinned patients characterised by clear centre erythematous localised at the shoulders and thighs. Infections such as conjunctivitis might also appear, these are believed to be secondary infections due to suppression of the patient's immune system. (Apted, 1970). The lymph nodules close to the inoculation sites also become swollen; in *T .b. gambiense* the posterior cervical lymph nodes at the back of the neck are also affected (Winterbottom's sign) whereas in *T. b. rhodesiense* infection occurs in the glands under the jaw, the armpit or in the groin. In the second stage of infection trypanosomes migrate into the CNS, the presence of parasites here results in a number of neuropathological signs. It is due to the symptoms experienced at this stage of the disease that gave the name of "sleeping sickness". In the case of the *T. b. gambiense* infection, sleeping pattern becomes disturbed, in *T.b. rhodesiense* infection on the other hand sleeping disturbance is less severe.

Etiopathology of the disease is thought to include (i) the breakdown of remnants of inactivated trypanosomes, (ii) substances releases by the active parasites, (iii) or substances produced by the host in the presence of the invading parasites. Mammalian forms of trypanosomes have been demonstrated to produce prostaglandin (Kubata *et al.*, 2000), this chemical is recognised as an endogenous sleep inducing substance (Hayaishi *et al.*, 1995).

1.3.2 Treatment

Development of a satisfactory chemotherapy against HAT has been challenging. Drugs such as Atoxyl, tryparsamide, suramin or pentamidine have been around for almost 60 years. The early stage of disease is easy to treat; parasites within the blood are killed by pentamidine and suramin; however suramin is nephrotoxic while pentamidine is directly toxic. The late stage of disease requires a more aggressive treatment in order for drugs to cross the blood brain barrier and reach those parasites within the CNS; to achieve this highly toxic arsenical compound is recommended.

The only drug developed against HAT over the last 40 years is Eflornithine (DL-α-difluoromethylornithine), which has been successfully used against the second stage of *T.b. gambiense* infection, though this drug is less effective against *T.b. rhodesiense* (Iten *et al.,* 1995). Overall, the treatment of HAT is still limited and calls for more research and treatment protocols to be implemented.

1.3.3 Animal Reservoir

Of the human infective trypanosomes *T. b. rhodesiense* has been known to have an animal reservoir for many years, since Heisch (1958) demonstrated that trypanosomes isolated from a Bushbuck were infective to humans. Since then it has been shown that a large number of vertebrate species are capable of maintaining this parasite. Recent evidence has also highlighted the role of the animal reservoir in the spread of this disease. For example the movement of infected cattle in Uganda has resulted in the spread of *T.b. rhodesiense* into previously unaffected areas (Fèvre *et al.*, 2001).

Although primarily an anthroponosis disease there is growing evidence, based on isoenzyme and DNA, that gambiense sleeping sickness might also have an animal reservoir (Gibson *et al.*, 1980; Mehlitz *et al.*, 1981, 1982; Paindovoine *et al.*, 1986), though the epidemiological importance of these findings is still to be established.

1.3.4 Tsetse stage

The life cycles of trypansomes in the insect stage is divided into 2 phases, (i) the establishment of a midgut population (procyclics) and (ii) the maturation of mammalian infective parasites (metacyclics). The majority of trypanosomes infections however fail to establish midgut infections and even fewer proceed to maturation (Welburn and Maudlin, 1997). Fly susceptibility and refractoriness to the establishment of trypanosomes was determined to be maternally inherited and could involve maternally inherited symbionts (Maudlin 1991). Lectins, glycoproteins detected in the tsetse midgut (Grubhoffer *et al.*, 1997), could play a major role in the fly susceptibility to trypanosome establishment since 100% midgut infection rate was obtained when flies where fed with lectin-inhibitory sugars (Maudlin and Welburn, 1987). It was suggested that the lectins kill incoming *T. brucei* parasites; though the maternally inherited symbionts, through their chitinolytic activity could producing the necessary lectin-inhibitory sugars creating tsetse susceptibility to trypanosome infections (Welburn and Maudlin, 1999).

It has been found that fly immunity plays a role in the clearance of the establishment of trypanosome infections when flies were injected with lipopolysaccharides (LPS) or *E.coli* (Hao *et al.*, 2001). It has also been suggested that the invading parasites do not initially trigger a tsetse immune response but do so when they start to differentiate to procyclic forms and their number increase in the ectoperitrophic space of the midgut. However the inefficiency of such a response to clear the established procyclics was proposed to be due to the resistance of the trypanosomes to immune peptides or the inefficiency for such immune peptides to reach the trypanosomes (Lehane *et al.*, 2004). It has been found that successful and unsuccessful establishing trypanosome infections induce transcriptional changes of two fly peroxidases (Lehane *et al.*, 2003) as a high level of reactive oxygen species are encountered in the fly as a result of immune

responses during the trypanosome infection (Zhengrong Hao and Serap Aksoy, unpublished observations). Several genes were found to be up-regulated in the tsetse fly during the trypanosome infection of the midgut and such genes were identified to have a complement like or proteinase inhibitory (Lehane *et al.*, 2003).

Maturation of established procyclics is thought to be dictated by either extracellular conditions / extracellular signals and only occurs to receptive parasites (mesocyclics) within a certain time period after the gut infection (Welburn and Maudlin, 1999; Van den Abbeele *et al.*, 1999). The time window has been defined to be 8-11 days following the initial infection (Dale *et al.*, 1995).

Maturation of salivary gland infections in established parasite populations is significantly influenced by the fly sex where salivary gland infections are higher in males than in females flies due to an X-linked gene that kills or prevents the maturation of trypanosomes (Milligan *et al.*, 1995). A second lectin believed to be from the haemolymph has been suggested to play a role in the maturation process of trypanosomes within the tsetse fly (Welburn and Maudlin, 1990). Maturation of the trypanosomes could also be related to parasite genotype (Maudlin *et al.*, 1986).

1.4 RNA editing and trans-splicing

RNA editing is a process occurring in the kinetoplast and is considered fundamental in the expression of the mitochondrial electron transport machinery and oxidative phosphorylation (Madison *et al.*, 2002). RNA editing consists of the addition and the deletion of uridylate residues within trypanosome mitochondrial mRNAs (Halbig *et al.*, 2004). Kinetoplast DNA is formed of maxicircles (50 copies of 25-27 kb) coding for the 9S and 12S ribosomal RNAs and 17 further open reading frames (ORFs), 11 genes involved in energy transduction (NADH dehydrogenase, cytochrome oxidase complexes I, II and III, ATPase 6 and cytochrome *b*) and minicircles (5000-10000 copies of 1 kb) encoding for small guide RNA (gRNAs) (Ferguson *et al.*, 1994) involved in mitochondrial mRNA editing by forming duplexes downstream of the pre-edited mRNA

(Sturn and Simpson, 1990). The process of mRNA editing starts with gRNAs forming short recognition complexes with their complementary mRNAs, the complex is then recognised by endonucleases which cleave the first base at the 3' ends formed between the gRNA and mRNA. Uridylate residues are then added to or deleted from the cleaved pre-mRNA by uridyl transferase and a 3' U-specific-exonuclease respectively, the final step involving an RNA ligase which joins the two mRNA halves (Madison *et al.*, 2002).

Three classes of chromosomes have been defined in trypanosomes, the large megabase chromosomes (≥ 1 Mb), the intermediate chromosomes (> 100 Mb) and the abundant mini-chromosomes (30 – 100 kb). Minichromosomes constitute 20 % of the nuclear DNA and other than a 177 bp repeat sequence, telomeric repeats and one or two unexpressed VSGs (Weiden *et al.*, 1991) do not seem to have any coding regions or recognizable promoters and it was suggested that they act solely as a reservoir for unexpressed VSGs. The intermediate chromosomes represent 0.1- 2 % of the genome lack 177 bp repeat sequences, but contain telomeric expression sites of VSGs and common gene sequences. The megabase chromosomes are represented by 11 pairs of chromosomes representing 80 % of the trypanosome genome, and are also called housekeeping chromosomes (containing all the housekeeping genes) (Melville *et al.*, 2004).

In higher eukaryotes protein encoding genes are arranged as monocistronic units, a single transcription unit comprising each gene and consisting of introns and exons representing non-coding and protein coding regions respectively. Generally eukaryotic mRNA transcripts are capped by the addition of a mGppp cap at the 5' end with polyadenylation at a specific site within the 3' untranslated region (3'UTR) (Meyer *et al.*, 2004). The maturation of the mRNA is carried out by a process defined as *cis*-splicing, which consists of the removal of introns and the linkage of the coding protein regions (exons). Most protein encoding gene sequences in trypanosomatids lack introns and the characteristic *cis*-splicing process of mRNAs is replaced by a trans-splicing process from polycistronic units (Agabian, 1990). Protein coding genes in trypanosomes

are arranged as polycistronic units comprising several genes separated by intergenic regions and controlled by only one promoter at the 5' end (Graham, 1995). Because of the existence of one promoter for several genes, the regulation of transcription of stage specific individual genes is undertaken at the post-transcriptional level to create a suitable cellular concentration for the specific gene product (Graham, 1995; Vanhamme and Pays, 1995).

Individual trypanosome mRNAs have an RNA cap of 40 nucleotides (nt) called the splice leader (SL) or "miniexon", and 5'-end and 3' polyadenylation and cleaved from the precursor by trans-splicing (Ullu *et al.*, 1993; Mathews *et al.*, 1994).The spliced leader of the trypanosome mature RNAs is encoded by a 1.35 kb unit and is present in at least 200 copies carried mainly in the mega chromosomes and in *Leishmania* it has been suggested that regions close to the transcription site might be critical for promoter activity and RNA stability (Saito *et al.*, 1994).

1.5 Cyclic nucleotides in eukaryotic cells

The adenosine 3',5'-cyclic monophosphate (cAMP) second messenger pathway was brilliantly described by the Nobel Prize winner Sutherland and colleagues whereby an external signal such as neurotransmitters or mammalian hormones recognise and bind to a specific external ligands resulting in conformational change of the receptor. This in turn transmits a signal that stimulates adenylyl cyclase in the interior of the cell. The activation of adenylyl cyclases results in the synthesis of cAMP from ATP . Increase in the intracellular level of cAMP results in the activation of protein kinase A (PKA), which induces the activation of target proteins by phosphorylation. The signal is then switched off by hydrolysis of cAMP to AMP by phophodiesterases (Robinson *et al.*, 1971).

The notion of a second messenger pathway utilising cAMP has been modified since the initial description whereby two type of receptors have been described one stimulating the activation of adenylyl cyclase while the other inhibits it (Strader *et al.*, 1994). The

cAMP pathway generated signal is also regulated by G-proteins, which are heterodimeric GTP-binding proteins acting as intermediates between the receptors and adenylyl cyclase and are also involved in the regulation of adenylyl cyclase. Protein Kinase A consists of two regulatory (R) subunits and two catalytic (C) subunits. Binding of cAMP causes the dissociation of the catalytic subunits from the regulatory subunits allowing them to convert to their active form (Francis and Corban, 1994; McKnight *et al.*, 1988).

Adenylyl cyclase plays a role in host-parasite communication of trypanosomatids. In *T. brucei ESAG4* is a gene associated with the ES of VSGs and codes for a transmembrane adenylyl cyclase (Nolan *et al.*, 2004). Typically, adenylyl cyclases are more abundant in trypanosomes than in higher eukaryotes (Naula and Seeback, 2000). In *T. cruzi* epimastigotes it has been found that during the insect stage, activation of adenylyl cyclase is triggered by a peptide produced during the cleavage of alpha-globin in the triatoma hindgut (Fraidenraich *et al.*, 1993) and proteolysis of fibronectin induces elevation of the level of cyclic AMP in *T. cruzi* (Nolan *et al.*, 2004).

In *T. brucei* the differentiation of long slender bloodstream forms into non-dividing short stumpy forms is preceded by high levels of cAMP activity and it has been suggested that long slender forms secrete SIF (stumpy induction factor) which has a similar effect as cAMP in the induction of differentiation (Reuner *et al.*, 1997).

Regulation of the cAMP signal is switched off through the action of phosphodiesterases (PDE). Tow Members of the PDE family have been retrieved from the *T. brucei* genome (Seebeck *et al.*, 2000) all five genes having a cGMP-binding regulatory GAF domain in their N-termini. The GAF abbreviation comes from the names of the first three different classes of proteins identified to contain them: cGMP-specific and -regulated cyclic nucleotide phosphodiesterase, adenylyl cyclase, and *E. coli* transcription factor FhlA (Aravind and Ponting, 1997).

Effector molecules of cAMP such as protein kinases are also being investigated in *T. brucei*. A *T. brucei* protein was found to have compelling similarity to the catalytic subunit of protein kinase A (Klöckner, unpublished) and it is thought that cAMP and PKA in *T. brucei* might be involved in the avoidance of the host immune response (Seebeck *et al.,* 2000).

cGMP is synthesized by guanylyl cyclase through the conversion of cytosolic GTP in response to diverse signals such as nitric oxide. Two families of guanylyl cyclases exist, soluble guanyl cyclases (sGC) activated by nitric oxide (NO) and particulate guanylyl cylases activated by natriuretic peptides (Lucas *et al.*, 2000). sGC is a heterodimeric protein consisting of alpha and beta subunits and both are required for catalytic activity (Kamisaki *et al.*, 1986; Hartnek *et al.*, 1990; Buechler *et al.*, 1991).

cGMP dependant protein kinase represents the major mediator of cGMP. The response induced by the binding of the ligand on GC results in an elevation of intracellular cGMP, which induces the activation of protein kinase G. This results in the catalytic transfer of a gamma-phosphate from ATP onto a serine or threonine residue resulting in phosphorylation of the target proteins (Lucas, 2000). As with cyclic AMP, PDE switches off the cyclic GMP signal by hydrolysis (McAllister-lucas *et al.*, 1995).

Biochemical activity of PDE in *T. brucei* was initially reported by the group of Walter (1974, 1978) and although GAF domains exist in *T. brucei* PDEs, no effect of cGMP could be detected on PDE activity (Rasco *et al.*, 2002). A regulatory subunit of PKA from *T. brucei* (TbRSU) was found however to be regulated by cGMP rather than cAMP. Shalaby *et al.* (2001) suggested that the cGMP-regulated TbRSU (PKA-like) kinase might well represent the major effector of cGMP signalling in these organisms.

In *Trypanosoma cruzi* a NO-GC-cGMP pathway has been detected in epimastiogote forms (Paveto *et al.*, 1995) and it has been observed that the mobility of epimastigotes is increased during high levels of cGMP (Paveto *et al.*, 1997).

Gene knock-out studies of PKG on *Toxoplasma gondi* provided evidence that the kinase is essential for parasite growth and that the enzyme has three putative cGMP binding sites (Donald and Liberator, 2002). Although the biological function is unknown, cGMP dependent activity has also been detected in epimastigotes of *Leishmania* parasites (Géigel *et al.*, 2003).

It has been shown in our laboratory that the addition of 8-Br-cGMP (a cell permeable cGMP analogue) to the infected blood meal increases by 100 % the establishment of *T .b.brucei* within the tsetse (*G.m.morsitans*) midgut (Macleod *et al.*, in press 2005). The rate of infections was as high in males as in the female flies when compared to the control flies groups which were fed only on non-treated bloodmeal. (Figure 1.4)

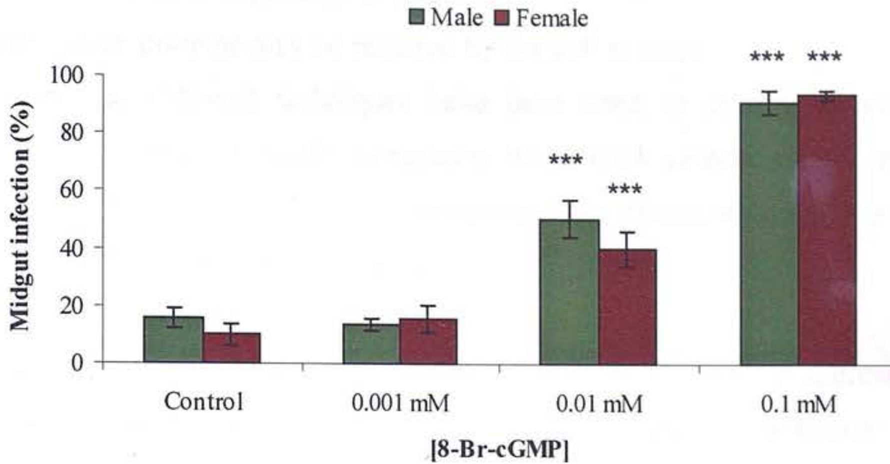


Figure 0.4

1.2.1 The effects of 8-Br-cGMP on midgut infection rates of *T. b. brucei* in *G. m. morsitans*
The effects on midgut infection rates of *T. b. brucei* BUT 135 in *G. m. morsitans* when the blood meal was supplemented with 8-Br-cGMP (Courtesy of Ewan MacLeod)

1.6 Reasoning for studying *T. brucei* procyclics

Possibly no other parasite has been studied so extensively as *T. brucei* (Burri, 2001) and with the near-completion of a genome sequencing project, analysis of the sequence of the trypanosome genome is now readily possible. *T. brucei* procyclic parasites can be readily grown *in vitro* and recent advances have meant that *in vivo* induction of infection within the tsetse fly vector can be initiated with 100% success.

1.7 Aims of this project

The digenic nature of the *T.brucei* life cycle means that these parasites must be capable of rapidly responding to changes within their environment in order to survive and complete transmission between the two hosts. This project focuses specifically on the tsetse vector stage of the life cycle and the rapid changes in gene expression *T. brucei* parasites may undergo.

The analysis of gene expression is therefore based on the analysis of mRNA in order to identify which proteins may be required by the cell in response to a particular situation. A number of different techniques have been used to investigate changes in the expression of genes, typically comparing the mRNA present within one population against that of another. Differences are assumed to represent changes in gene expression in response to the stimulus investigated.

The aim of this thesis was to investigate the most appropriate approach to analyse changes in gene expression within procyclic *T. brucei* and to characterise such changes using the various approaches. Specifically, I have investigated genes differentially expressed in *T. brucei* from an established midgut population presenting a salivary gland infection and from an established midgut population with no salivary gland infection.

In addition I have examined the response of *in vitro* cultures of *T. brucei* to the cGMP analogue 8-Br-cGMP. Previous work carried out with cGMP treated *T. brucei* in which 8-Br-cGMP, an non-hydrolysable and cell-permeable analogue of cGMP, showed significantly increased levels of midgut infection in the vector fly when introduced with a trypanosome containing blood meal (McLeod *et al.*, in press) the addition of this chemical increasing the level of trypanosome establishment within the fly population to 100%. The specific effect of this compound on *T. brucei in vitro* was therefore investigated through the creation of an enriched subtractive cDNA library.

Recent advances in the field of proteomics have developed rapid throughput methods for the examination of protein expression profiles. These profiles reflect the nature of proteins separated according to their biochemical characteristics. Such a proteomic approach was employed in order to ascertain whether 8-Br-cGMP has any affect on the protein expression profile of these parasites *in vitro*.

Chapter 2

Material and Methods

2.1 Commonly used buffers

- All solutions were prepared with distilled water unless otherwise stated
- All chemicals were analytical grade, purchased from Sigma Chemical or BDH (British Drug House)

2.1.1 8-Br-cGMP Stock solution

10 mg of Cell-permeable 8-bromoguanosine 3', 5-cyclic monophosphate sodium salt monohydrate (Sigma-Aldrich) was reconstituted in 1 ml of deionised water, filter-sterilized using a 0.2 μm filter Minisart®(Sartorius), aliquoted in 100 μl and stored at -20°C

2.1.2 Cunningham's medium

Made to the recipe of Cunningham, (1977) by Gibco (UK).
0.53 g/l $NaH_2PO_4 .H_2O$, 3.04 g/l $MgCl_2 .6H_2O$, 3.70 g/l $MgSO_4.7H_2O$, 2.98 g/l KCl (anhydrous.), 0.15 g/l $CaCl_2 .2H_2O$, 0.70 g/l Glucose, 0.40 g/l D-Fructose, 0.40 g/l Sucrose, 0.67 g/l L-malic acid, 0.37 g/l alpha-ketoglutaric, 0.055 g/l Fumaric acid, 0.06 g/l Succinic acid, 0.522 g/l Cis-aconitic acid, 10 ml/l Pyruvate (100 mM), 2.0 g/l β-Alanine, 1.09 g/l DL-alanine, 0.44 g/l L-arginine, 0.24 g/l L-asparagine. H_2O, 0.11 g/l L-aspartic acid, 0.08 g/l L-cysteine, 0.03 g/l L-cystine, 0.25 g/l L-glutamic acid, 1.64 g/l L-glutamine, 0.12 g/l glycine, 0.16 g/l L-histidine, 0.09 g/l DL-isoleucine, 0.09 g/l L-leucine, 0.15 g/l L-lysine, 0.20 g/l DL-methionine, 0.20 g/l L-phenylalanine, 6.90 g/l L-proline, 0.20 g/l DL-serine, 0.27 g/l L-taurine, 0.10 g/l DL-threonine, 0.10 g/l L-tryptophane, 0.20 g/l L-tyrosine, 0.21 g/l DL-valine, 2 ml/l BME Vitamins (100X), 4 ml/l Phenol red.

To make up Cunningham's medium, 20% foetal calf serum (Sigma, UK) and 20 μg/ml gentamicin (Invitrogen, UK) were added and the solution filter sterilised before use.

2.1.3 DNA loading buffer (6x)

Ficoll	15 % (w/v)
Orange G	0.2 % (w/v)
Xylene cynol	0.1 % (w/v)

2.1.4 Ethidium Bromide

A stock of 50 mg per ml was made and diluted to give the working concentration of 50 μg per ml.

2.1.5 Luria-Bertani (LB) medium:

	g/litre
Bacto-tryptone	10
Bacto-yeast extract	5
NaCl	10

The compounds were dissolved in 950 ml deionized water; pH was adjusted to 7 with 5M NaOH. Prior to autoclaving the volume was adjusted to 1 litre

2.1.6 LB agar

LB medium as described above with addition of 15g/litre bacto-agar prior to autoclaving

2.1.7 PBS (1x):

NaCl	2.19 g
Na_2HPO_4	5.8 g
KH_2PO_4	2.4 g

Deionised water to 1 litre pH 7.5 and sterilised by autoclaving

2.1.8 RNA loading buffer (6x)

Bromophenol blue	0.25 % (w/v)
Xylen cyanol	0.25 % (w/v)
Glycerol	30 % (v/v)
SDS	1.2 % (w/v)
Sodium phosphate (pH6.8)	60 mM

2.1.9 SOC medium

Tryptone	2%
Yeast Extract	0.5%,
NaCl	10 mM

KCl	2.5 mM
$MgCl_2$	10 mM
$MgSO_4$	10 mM
Glucose	20 mM

2.1.10 TAE (10x)

Tris base	96.8 g
Sodium acetate	26.75 g
EDTA	14.8 g

Deionised water to two litres. The pH was adjusted to 8.0 with glacial acetic acid.

2.1.11 TBE (10x):

Tris base	121 g
Boric acid	61.8 g
EDTA	18.6 g

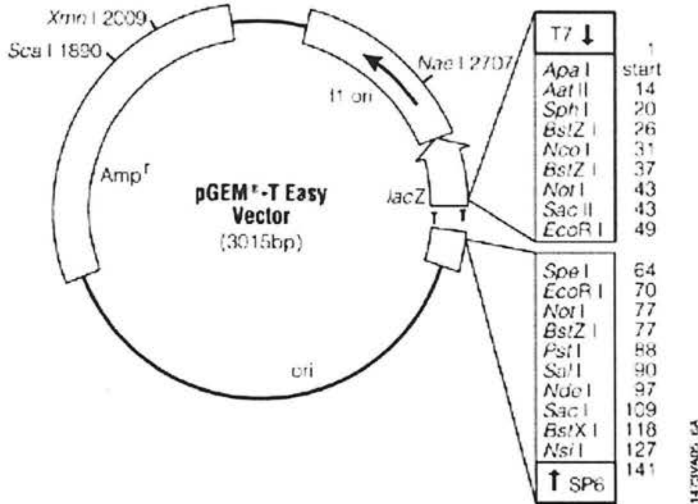
Deionised water to 1 litre, pH 8

2.1.2 TE (1x):

Tris base 1.2 g in 800 ml of deionised water, pH adjusted to 8.0 with 1M HCl, 2 ml 0.5M EDTA pH 8.0, deionised water to 1 litre, autoclaving

2.2 Bacterial strains, ligation vector and vector details

The plasmid vector pGEM®-T Easy was used to clone double-stranded DNA. This plasmid and its bacterial host strain JM109 were purchased from Promega, UK. The phenotype of *E coli* JM109 is: endA1, recA1, gyrA96, thi, hsdR17(r_{k-} m_{k+}), relA1, supE44, Δ(lac-proAB), [F', traD36, proAB, laql^{q}Z ΔM15]



2.3 Primers sequences

2.3.1 Primers sequences used in the PCR-Select™ cDNA Subtraction library :

cDNA synthesis primer-5'–TTTTGTACAAGCTT$_{30}$N$_1$N–3'(N= A G C T; N$_1$= A G or C)

Adaptor1-5'–CTAATACGACTCACTATAGGGCTCGAGCGGCCGCCCGGGCAGGT–3'

Nested PCR primer 1-5'–TCGAGCGGCCGCCCGGGCAGGT–3'

PCR Primer 1-5'–CTAATACGACTCACTATAGGGC-3'

Adaptor 2R -5'–CTAATACGACTCACTATAGGGCAGCGTGGTCGCGGCCGAGGT–3'

Nested PCR primer 2R-5'–AGCGTGGTCGCGGCCGAGGT–3'

SP6 RNA polymerase promoter-5'-ATTTAGGTGACACTATAGAAGNG-3'

2.3.2 Anchored random primers used for low/high stringency PCR

P_1-5'-ATTAACCCTCACTAAATGCTGGGGA-3'

P_2-5'-ATTAACCCTCACTAAATCGGTCATAG-3'

P_3-5'-ATTAACCCTCACTAAATGCTGGTGG-3'

P_4-5'-ATTAACCCTCACTAAATGCTGGTAG-3'

P_5-5'-ATTAACCCTCACTAAATGATCTGACTG-3'

P_6-5'-ATTAACCCTCACTAAATGCTGGGTG-3'

P_7-5'-ATTAACCCTCACTAAATGCTGTATG-3'

2.4 Parasite Material

2.4.1 Bloodstream form preparation

Trypanosoma brucei brucei Buteba 135 (But 135) strain was originally isolated from a cow in Buteba village – Tororo District, South East Uganda in 1990. This strain was used throughout the experiment.

Two TO (BKl:BLW) mice (B and K Ltd, UK) were immunosuppressed by intraperitoneal injection of 80 mg/kg cyclophosphamide (Pharmica and Upjohn Ltd) and infected from a frozen trypanosome stabilate (But 135) 24 h later by intraperitoneal injection. Regular parasiteamia checks were made by examination of blood by tail clipping of mice. When the parasitemia reached between 10^8 and 10^9 trypanosomes/ml the blood from the two mice was collected by cardiac puncture under anaesthesia and pooled. The infected blood was then injected into 10 TO mice which had been immunosuppressed with cyclophosphamide (80 mg/kg) 24 h previously. Mice were then checked regularly and when parasitaemia reached between 10^8 and 10^9 /ml the blood from the ten mice was collected as above by cardiac puncture, pooled then diluted 1:1 in phosphate-buffered saline-glucose 6:4 (Lanham and Godfrey, 1970) with 14% glycerol. The mixture was then allowed to equilibrate for 10 min then separated into 200 µl aliquots in 1 ml cryotubes (Nunc, Denmark) then frozen at -20°C overnight in polystyrene boxes before transfer to -70°C. This method yields a final dose of $1x10^6$ to $4x10^6$ trypansomes/ml when the aliquot is reconstituted in 5 ml of blood (Welburn and Maudlin, 1987).

2.4.2 Tsetse fly Infection

The parasite stabilates were re-suspended in defibrinated sheep blood (Welburn and Maudlin, 1987) and flies were infected as described by Maudlin (1997). One ml of PBS at a time was slowly added to the infective stabilates to make up a final volume of 4 ml. The mixture was centrifuged at 1500 x g for 10 min at 4°C, most of the supernatant fluid

was discarded and only 0.5 ml were left to resuspend the pellet containing the parasites. The viability viability of trypanosomes was checked under microscopy and the infective mixture was added to 4.5 ml of fresh defibrinated pig blood, constituting the infective feed.

Newly emerged *Glossina morsitans morsitans* flies were placed in netted cages and maintained at 25°C and approximately 70 % room humidity (insectarium) (Maudlin, 1997). Twenty four hours prior to infection, flies were immobilised by chilling briefly in a refrigerator and were transferred in a Petri dish placed on ice. Fly wings were removed using a pair of scissors, and the wingless flies were placed back in the insectarium. The infective feed was poured onto a sterile feeding plate and covered with a 15 cm diameter silicon membrane. The feed was evenly spread using a roller and was placed on a warmed heating plate at 37°C. The cages containing the wingless flies were the placed on a top of the membrane and covered with a black cloth. The flies were left to feed for 30 min.

2.4.3 Preparation and maintenance of procyclic culture

Infected flies were dissected 10 days post-infection for their midgut; midgut infection was confirmed by microscopy. Infected midguts were transferred to a universal tube containing 5 ml of Cunninghams medium. The procyclic parasites were cultured in 75 cm^2 EasY Flasks, at a constant temperature of 27°C.

2.4.4 Parasite Count

The concentration of trypansomes culture was assessed by microscopy using a haemocytometer of 0.2 mm depth (Weber, England).

2.4.5 Treatment with 8-Br-cGMP

Trypanosomes in the log phase of growth ($1x10^7$/ml) were treated with 8-Br-cGMP to a final concentration of 100 μM. The parasite cultures were pelleted by centrifugation at 14000 x g for 15 min at room temperature and stored at -70 °C. The same procedure was followed to prepare the controls where parasites were treated with the solvent that had been used to prepare the 8-Br-cGMP solution

2.5 RNA methodologies

Since RNases are present on all surfaces, pipettors, benchtops, glassware, and gel equipment, were decontaminated with a surface decontamination solution (RNase*Zap*™) RNase-free tips, tubes, and solutions were used and gloves were changed frequently

2.5.1 Total RNA extraction

Total RNA was extracted from trypanosomes cell cultures and tsetse infected guts using the RNeasy mini kit (Qiagen), according to the manufacturer's instructions.
All centrifugation steps were carried out at ≥ 8000 g at room temperature unless stated otherwise. The principle is based on the selective binding properties of a silica-gel based membrane where in the presence of high salt conditions; RNA binds onto the RNeasy silica-gel membrane.

Two milliliters of trypansomes cell culture or 10 infected guts were pelleted down by centrifiguation at 3000 g. The cell material was lysed in 350 µl of RLT lysis buffer containing 1% ß-mercaptoethanol by pipetting up and down for 5 min at room temperature. To shear high-molecular weight genomic DNA and other high-molecular-weight cellular components, an extra homogenization step was performed. Lysates were passed through QIAshredder spin column (Qiagen) by centrifuging for 2 min. The resulting homogenates were mixed by pipetting with 350 µl of 70 % ethanol. The mixture was then applied to an RNeasy mini column and centrifuged for 5 min to bind total RNA to the silica-gel membrane. The bound total RNA was washed in 700 µl of washing buffer RW1 by centrifugation for 5 min. Total RNA was then eluted in 50 µl of RNA*ase* free deionised water.

2.5.2 DN*ase* Treatment of RNA

RNA samples were treated with DN*ase* (DNA*free*™, Ambion®) to remove any DNA contamination. The procedure followed the manufacturer's instructions.

Five microlitres of 10x DN*ase* I buffer (100 mM Tris-HCl pH 7.5, 25 mM $MgCl_2$, 5 mM $CaCl_2$) and 3 µl of DN*ase* I (6 U) were mixed with 50 µl of total RNA sample , and the reaction mixture was incubated at 37°C for 30 min. The DN*ase* was deactivated by adding 10 µl of DN*ase* inactivation resin and incubated for 2 min at room temperature with occasional mixing. The mixture was then centrifuged at 10 000 g for 2 min to pellet the resin. The supernatant fluid containing the total RNA was transferred to a fresh tube and stored at -70°C until subsequent applications.

2.5.3 Poly A^+ RNA isolation

Poly [A]-enriched RNA was isolated from procyclic form trypanosome total RNA using the MicroPoly [A] Purist *mRNA Purification Kit* (Ambion®), the procedure followed the manufacturer's instructions.

The method is based on the isolation of mRNA from total RNA in a very low ionic-strength buffer using Oligo (dT) Cellulose to bind to the stretch of "adenosine" which is present at the 3' ends of eukaryotic messenger RNA (mRNA). The hybridization step is performed by a gentle rocking of the mixture for an hour at room temperature, followed by a washing step using a spin column. This column binds the hybridised mRNA while the non-specifically bound material and ribosomal RNA are washed away. In a final step the mRNA is eluted using a pre-warmed buffer. All centrifugation steps were performed at 4000 x g for 3 min at room temperature

Nuclease-free water was added to the RNA sample to a final volume of 250 µl then 250 µl of 2 x binding solution was added and mixed. This was follwed by adding 20 mg of Oligo (dT) cellulose to the mixture and incubated at 70°C for 5 min. The reaction mixture was then incubated at room temperature for 60 minutes with gentle rocking, followed by centrifugation. The RNA/Oligo (dT) mixture was pipetted on a spin column and the mixture was bound to the column through a centrifugation. The column was then washed twice with 500 µl Solution 1; this was then followed by 2 washes with 500 µl Solution 2. The poly A^+ RNA was then eluted twice in 100 µl of a pre-warm RNA storage solution. Poly A^+ RNA was then precipitated with 500 µl of absolute ethanol and 10 µl of 5 M ammonium acetate (Sambrook *et al.*, 1989). To help to precipitate RNA, 1 µl of glycogen was added to the mixture as it acts as a nucleic acid carrier (Hengene 1996). The solution was mixed by pipetting up and down and incubated at -70°C for 30 minutes. The precipitated RNA was then recovered by centrifugation at 14 000 g for 30 minutes at 4°C. The pellet was washed with 500 µl of 70 % (v/v) ethanol and centrifuged at 14 000 g for 10 minutes. The supernatant fluid was discarded and the pellet was air dried and dissolved in 50 µl of RN*ase* free water

2.5.4 RNA quantification

Total RNA and poly A^+ mRNA were quantified using RiboGreen RNA Quantitation Kit (Molecular Probes) the procedure followed the manufacturer's instructions. Presence of contaminants such as ethanol and proteins during RNA extraction could interfere with spectrophotometer readings, therefore quantification by sensitive fluorescent nucleic acid stains was carried out

Standard solutions were prepared as follows, ribosomal RNA was mixed to 1 x TE (16S and 23S rRNA from *E. coli*) to a final concentrations of 1µg/ml, 500 ng/ml, 100 ng/ml, and 20 ng/ml. 10 µl of RNA Ribogreen was added to 20 ml of 1 x TE to constitute a 200-fold dilution of RNA-Ribogreen stock solution. 2 ml of Ribogreen solution was added to each RNA standards mixed well and incubated in a dark place at room temperature for 5 min. The fluorescence intensity of each RNA standard solution was read at 520 nm using a fluorometre and a linear standard curve was calculated.

Total RNA and poly A^+ mRNA samples were subjected to the same procedure where the fluorescence readings of each sample was compared with the RNA standard curve and the poly A^+ RNA concentration of each RNA sample was calculated.

2.5.5 RNA Gel Electrophoresis

Gel electrophoresis of RNA was performed according to the protocol of Sambrook (1989). To remove any traces of exogenous RN*ase* that could affect the outcome of the experiment, gel tanks and gel combs were washed using RNA Zap® (Ambion) prior to usage.

0.5 µg of RNA was added to 1 x RNA loading buffer and the mixture was immediately denatured at 70°C for 5 minutes. The samples were then centrifuged briefly and placed on ice prior loading. The RNA samples were loaded onto a 2% Agarose gel (w/v) (multipurpose agarose, Sigma). RNA markers (Promega) were also loaded (0.281-6.583 kb) and the gel was run in a buffer consisting of 10 mM sodium Phosphate (pH 6.8) containing 1 µg/ml ethidium bromide. RNA samples were electrophoresed with a constant recirculation of the sodium phosphate buffer using a Varioperpex Peristaltic Pump (LKB Bromma 12000 Varioperrpex Peristaltic Pump). The RNA bands were visualised under a UV transluminator using Gel Doc 2000 (Bio-Rad).

2.6 DNA

2.6.1 Gel electrophoresis

Two grams of agarose where mixed to 100 ml of 1x TAE buffer. Ethidium bromide was added to a final concentration of 10ng/ml for DNA staining and the mixture was then heated briefly in the microwave to dissolve. The gel mixture was allowed to cool for 5

minutes at room temperature prior to casting in a gel tray. DNA samples were mixed with 6x DNA loading buffer containing 15 % ficoll to allow the DNA samples to sink to the bottom of the wells. Orange G was added to check the migration of the samples. The samples were electrophoresed at 100 V for 90 min. Ethidium bromide stained DNA products were visualized under a UV transluminator (Bio-rad).

The DNA bands were compared with the size of DNA markers that were electrophoresed along with the DNA samples. The standard markers used were:

- 100 bp DNA ladder (G2101, Promega, 0.65 μg),
- Φ X174 RF DNA/*Hae* III fragments (15611-015, Invitrogen, 100 ng)

2.6.2 Gel extraction and purification

The DNA was purified from selected bands using the QIAX II for agarose gel extraction kit (Qiagen); following the manufacturer's instructions. All the centrifugation steps were done at 10 000 g for 30 seconds at room temperature, unless stated otherwise. Selected bands were cut with a clean, sharp scalpel blade under a UV translumination (UVP Dual-intensity transluminator).

Each 100 mg of agarose plug was combined with 100 μl of buffer QX1 and 10 μl of re-suspended QIAEX II. The sample was mixed thoroughly and incubated at 50°C for 10 min with vortexing every 2 min for the DNA to bind the resin. The sample was then centrifuged. The resulting pellet was washed twice with 500 μl of buffer PE supplemented with 100 % ethanol and recovered each time by centrifugation. To remove any excess of washing buffer, the pellet was allowed to air-dry for 15 min The DNA was eluted by adding 20 μl of 10 mM Tris-Cl, pH 8.5. The pellet was then dissolved by vortexing and incubating at room temperature for 5 min. The DNA suspension was centrifuged and the supernatant fluid containing the recovered DNA was transferred to a 0.5 ml micro-centrifuge tube and stored at -20°C for subsequent applications.

2.6.3 Quantification

Double stranded DNA concentration was read at 260 nm in a spectrophotometer according to Sambrook *et al* (1989). 50 µg/ml of double stranded DNA (dsDNA) has an absorbance at 260 nm equal to 1 $A_{260} = 1 = 50\mu g/ml$ dsDNA.

2.7 Cloning and recombinant bacterial culturing

2.7.1 Ligation

All ligation reactions were performed using pGEM®-T Easy Vector System (Promega). This kit is based upon the fact that the thermostable polymerase enzymes often add a single deoxyadenosine at the 3'-ends of the amplified products (Clark, 1988). The pGEM®-T easy vector is digested with *Eco*R*V*, which creates overhang 3' ends with a 3' terminal thymidine; this consequently prevents recirculation of the vector and enables insertion of the PCR products (Mezei *et al.*, 1998).

60 ng of insert DNA were added 20 ng pGEM®-T Easy Vector with 5 µl of 2 x ligation buffer (60 mM Tris-HCl, pH 7.8, 20 mM $MgCl_2$, 20 mM DTT, 10 % polyethylene glycol, 2 mM ATP), 1 µl of the bacteriophage T4 DNA Ligase (3 Weiss units/µl) and the reaction volume was made up to 10 µl with deionised water. The ligation reactions were performed at 4°C overnight.

2.7.2 Transformation

High efficiency JM 109 competent cells ($\geq 1 \times 10^8$cfu/µg DNA) were thawed on ice for 5 minutes. Two microlitres of ligation reaction were pipetted and placed in a chilled 1.5 ml microcentrifuge tube. To the reaction, 50 µl of competent cells JM109 were added and gently mixed by flicking the tube. The reaction mixture was then placed on ice for 20 minutes. The cells were heat-shocked for 50 seconds at 42°C in a water bath and replaced immediately on ice for an extra 2 minutes. To the transformation reaction 950µl of SOC medium was added and the mixture was incubated at 37°C for 1h 30 min with continuous shaking (~150 rpm). The interruption of the ß-galactosidase coding sequence in the pGEM®-T Easy vector by the DNA insert allowed colored screening (Blue / White colonies) between recombinant cells. Therefore 100 µl of IPTG (100 mM) and 20

of X-Gal (50 mg/ml) were spread onto LB/ Agar plates supplied with 100 μg/ml ampicilin and the plates were left to dry at 37°C for 30 min. 100 μl of transformation culture was plated onto IPTG, X-Gal LB/ Agar plates and incubated at 37°C overnight. For better screening between white and blue colonies the plates were incubated at 4°C for 30 min to aid the color development prior to colony picking.

2.7.3 Overnight bacterial cultures and plasmid purification

2.7.3.1 Bacterial culture

Using a 100 μl pipette tip, recombinant colonies were picked and transferred to a universal tube containing 2 ml of LB broth supplemented with 100 μg/ml ampicilin. The bacterial cultures were incubated overnight at 37°C with gentle shaking (150 rpm).

2.7.3.2 Glycerol stock

20 μl of autoclaved glycerol was added to 80 μl of the overnight bacterial cultures. The solution was well mixed and stored at -70°C.

2.7.3.3 Plasmid DNA Purification

Plasmid DNA isolation was carried out using QIAprep® Spin Miniprep (Qiagen).

Overnight bacterial cultures were harvested by centrifugation at 10 000 x g for 10 min at room temperature after which the supernatant fluid was discarded. The pelleted bacterial cells were resuspended in 250 μl Buffer P1 supplemented with RN*ase* A (final concentration of 2 ng/ml) and 250 μl of buffer P2 (lysis buffer) before incubation for 5 min at room temperature. To the reaction mixture 350 μl of buffer N 3 were added. To pellet the bacterial cell debris and DNA, the mixture was centrifuged at 10 000 x g for 10 min at room temperature. The supernatant fluid was transferred to a silica-gel membrane column and the plasmid DNA was bound to the membrane by centrifugation at 10 000 x g for 1 min at room temperature. The flow-through was discarded and the column was washed by centrifugation at 10 000 g for 1 min with 750 μl of PE buffer supplemented with 100 % ethanol. An extra centrifugation step was performed to

remove residual wash buffer. The DNA was then eluted by adding 50 μl of buffer EB (10 mM Tris-Cl, pH 8.5) directly to the centre of the column and let standing for 1 min. The column was centrifuged at 10 000 x g for 1 min at room temperature. The resulting plasmid DNA was transferred to a 1.5 ml centrifuge tube and stored at -20°C.

2.8 DNA sequencing and sequence analysis

All samples were sequenced using SP6 RNA polymerase promoter sequence site present on the pGEM®-T Easy Vector. Sequencing was performed at Lark Technologies™ (www.lark.com) from bacterial stock prepared in section 2.7.3.2.

Sequences were analysed using DNASTAR *Software* [LASERGEN, Wisconsin, USA]. The sequence data was first trimmed to remove contaminant vector sequence using the Vector search programme (www.ncbi.nlm.nih.gov/VecScreen/VecScreen.html) which removes any DNA segment within the sequence submitted that matches sequences in a specialized non-redundant vector database

2.8.1 Preparation of DNA sequences

Sequence data was clustered into contiguous consensus sequences ("contigs"). Prior to assembly the sequences were filtered to remove primers, adaptors sequences and low complexity regions since this could lead to over-clustering of the sequence data by providing common but artificial regions between the sequences. A parameter of 95% identity over a region of 40 nucleotides (nts) long was applied for sequence clustering.

2.8.2 Blast Search

BLAST (Basic Local Alignment Search Tool) analysis was carried out on assembeled contigs using the Blast x search program (compares a nucleotide sequence query translated in all six reading frames to proteins sequence on the database) on the *Trypanosoma brucei* genome project (http://www.sanger.ac.uk/Projects/T_brucei/) using the trypansoma Omniblast Server. The trypansoma Omniblast server undertakes a search of all available databases at the Sanger institute (Blast n and tBlast x or tBlast n)

and the Blast output displays the five best hits. Finally blast x search analysis was also conducted on NCBI (http://www.ncbi.nlm.nih.gov/BLAST/)

Blast search results were expressed using e-values. The e-value represents the probability that a sequence alignment with the same or more significant score **s** could exit by chance. The score s is the raw alignment between the query sequence and the subject sequence in the database. So the lower the e-value the more significant is the result therefore Blast x results were deemed significant when a score of 80 or more and an e value of e^{-5} were obtained (Pearson 1997).

2.9 Characterisation of differentially expressed proteins on ProteinChip Arrays

2.9.1 Method

Identification of differentially expressed proteins was carried out using three different ProteinChip® Arrays; strong anion-exchange (SAX 2), weak cation-exchange (WCX 2) and reversed phase (H 50). Normal phase (NP 20) ProteinChip® Arrays were used for external calibration of the PBSII reader.

The ProteinChip® arrays were placed into a deep-well type Bioprocessor assembly (Ciphergen Biosystems). Prior to sample loading both the SAX 2 and WCX 2 arrays were equilibrated with 200 µl binding buffer (100 mM Tric-HCl, 0.1 % Triton x 100, pH 8, for SAX2 and pH 4 and pH 6 for SAX 2), while the H 50 was pre-equilibrated with 200 µl of binding buffer (10% acenonitrile in PBS). All the arrays were incubated for 15 min on a Microplate shaker at room temperature (Heidolph Titramax 100). 10 µl of each protein sample obtained in section 5.2.2 was diluted in 190 µl of the relevant binding buffers and applied to the appropriate ProteinChip® array. Samples were allowed to bind for at least 30 min on a Microplate shaker at room temperature. Excess of sample was removed and the ProteinChips®were washed twice with the appropriate buffer for 10 minutes to remove weakly bound proteins. All the arrays were then rinsed twice in 200 µl dionized water for 10 min and allowed to dry at room temperature. An aliquot of

0.8µl sinapinic acid, which is used as an EAM, was applied twice to the spots of each array.

The arrays were read in the Ciphergen ProteinChip® Reader (Ciphergen ProteinChip PBSII Reader) and the data was analyzed with the ProteinChip® Software version 3.0 (Ciphergen Biosystems).

2.9.2 Peak Detection

The mass spectra of proteins were generated by using an average of 150 shots, at a laser intensity of 260-280 arbitrary units. For data acquisition of low molecular weight proteins, the detection range was between 3-50 kDa with the detector sensitivity set at 8 and the laser intensity set at 260. For data acquisition of high molecular weight proteins, the detection range was between 30-100 kDa with the sensitivity detector set at 8 and the laser intensity set at 280.

Peaks with a signal/noise ratio of at least 3 were concidered meaningful. The region below 3 kDa was not concidered, as this was due to matrix and chemical interference. A biomarker identification Wizard was used to confirm the differences in relative intensities between protein profiles. The biomarker Wizard groups the protein peaks of similar molecular weight (0.3% mass window) across the treated and control preparations into clusters, and according to their relative intensity it shows differences in protein level both statistically and visually.

The programme uses two separate peak detection passes. The first pass recognises potential biomarkers and builds up clusters around them; the second pass includes more peaks that were too small to be identified during the first pass. To view the biomarkers wizard, plots were used with a log normalised intensity option, which plots the log of the peak intensity and normalises the average intensity to 0.

Chapter 3

Introduction

3.1 Introduction

The complex transmission of *Trypanosoma brucei spp* between the mammalian host and the tsetse fly vector (*Glossina spp*) induces the parasite to undergo several complex morphological and biochemical changes in order to survive between the two extremely different host environments (Halbig *et al.*, 2004). When trypanosomes are taken up in a bloodmeal by the fly, they pass through the food canal (FC) and crop (Cr) to reach the midgut. Parasites remain in the gut for up to three days before migrating toward the posterior portion of the fly gut where they take refuge within the space formed between the peritrophic membrane and the gut epithelium. The procyclic (insect form) trypanosomes either survive and establish a constant population within susceptible flies or are cleared by refractory flies. In the refractory flies, dying parasites in the tsetse midgut show characteristic of apoptosis including blebs of the pelicules and fragmentation of DNA (Welburn *et al.*, 1989). Once established, a small fraction of parasites will go on to invade the salivary gland where they transform first into epimastigotes and finally to metacyclics (the mammalian infective form of the insect stage). The tsetse vector stage of *T. brucei* life cycle is therefore divided into two phases. The first phase is the establishment of a constant parasite population (procyclics) and the second phase is the maturation of a salivary gland infection (metacyclics).

3.1.1 Establishment

In the tsetse flies, trypanosomes must switch from the bloodstream form to the procyclic form in order to survive within their new host. This switch is believed to occur rapidly in as little as two h and the transformation lasts for two to three day (Vickerman 1985). This transformation to procyclic form is believed to be triggered by a drop of temperature and can be reproduced *in vitro* by incubation of parasites at 27°C (Cross 1973, Ghiotto 1979). The transformation process can be significantly enhanced by adding citric acid and/or *cis*-aconitate to the culture medium (Brun 1981, Czichos 1986).

The establishment of a permanent procyclic population happens in a small percentage of flies. Most flies are refractory to the establishment of a procyclic population; proportions of trypanosome infections are of the value of 1-20% in laboratory reared flies and are dependant on fly species and parasite strains (Hao *et al.*, 2001). One of the factors believed to influence the establishment of parasites within the flies are sugar binding glycoprotein called lectin, a compound found in the fly gut during the feeding process (Grubhoffer *et al.,* 1997). The refractoriness to trypansomes infection was overcome by the addition to the bloodmeal of the sugar D^+ Glucosamine (Welburn *et al.*, 1994) which has previously shown to inhibit the agglutinating activity of tsetse midgut lectins (Ibrahim *et al.*, 1989). Establishment is also believed to correlate with the presence of *Sodalis glossinidius*, a secondary symbiont bacterium of tsetse. It is believed that these bacteria produce a lectin inhibitory sugar through a chitinolytic activity during pupa development (Welburn *et al.*, 1993), this in turn would promote the establishment of parasites infection by inhibition of gut lectin. Dale and Welburn (2001) obtained flies refractory to trypanosome infection by adding antibiotics to the fly bloodmeal. They suggested that teneral flies are more susceptible to trypanosome infection, due to a production of lectin inhibitors during the pupal stage.

Fly immunity has also been investigated in the establishment of midgut infection. Aksoy *et al.* (2001) noticed that the trypansomes infection rate of flies where the immune system was previously challenged by an injection of *E. coli* significantly reduced trypanosome infections compared to the unchallenged group. Similar results were also recorded with *Brugia malayi* and *Plasmodium* upon the bacterial challenge of their respective host vectors (Lowenberg *et al.*, 1996).

3.1.2 Maturation

Maturation of trypanosomes in the tsetse fly starts by the migration toward the salivary glands. Two routes of migration of procyclics to the salivary glands have been put forward. Firstly it has been suggested that trypanosomes do not cross epithelial barriers and migrate forward through the peritrophic space where they exit at the proventricular

valve, and then migrate through the food canal and pass through the salivary ducts into the salivary glands. This was supported by the observation of high number of parasites in the vicinity of the proventriculus (Van den Abbeele *et al.*, 1999). The second route constitutes a short-cut where parasites traverse the peritrophic membrane, the midgut membrane; pass to the hemocoel from where they find and penetrate the salivary glands, this theory is supported by the observation of trypansomes in the tsetse haemolymph (Mshelbwala, 1972; Otieno, 1973).

Maturation is believed to be initiated by several factors. Dale *et al.*, (1995) suggested that the faster a growing population of procyclics forms within the gut, the more likely it is that they will succeed in becoming mature infections. The maturation process is also believed to depend on yet unidentified serum related factors (Maudlin *et al.*, 1984). It has also been noticed that male flies mature and present more salivary gland infections than female flies (Dale *et al.*, 1995).

Density control of parasites is also believed to trigger the transformation of trypanosomes form one form to the other. For example, Vassella *et al.*, (1997) suggested that the transition from long slender dividing bloodstream forms into non-dividing stumpy forms is controlled by a density sensing mechanism which would stop the dividing parasites from overgrowing in the host and assuring the survival of the host and thus the survival of the parasite population. During the insect stage, procyclic forms are faced with the same dilemma once in the ectoperitrophic space. Welburn and Maudlin (1997) speculated that the established trypanosome population within the ectoperitrophic space would in order not to overload the host; either proceed to maturation and/or undertake suicide in the form of programmed cell death (PCD). Procyclic cultures were shown to display signs of PCD of upon treatment with the lectin concanavalin A (Welburn et al 1999).

Maturation of trypanosomes is also believed to be linked to the presence of lectins. In established *T. congolense* infections Welburn *et al.*, (1989) suggested an increase in the

duration of exposure to midgut lectins corresponded to a higher frequency of maturation. A second lectin, believed to be of haemolymph origin, is also believed to be partly responsible in the maturation of the established midgut parasites (Welburn and Maudlin, 1990). It has been suggested also that the maturation of infections in the midgut is a phenomenon predominantly trypanosome genotype related (Maudlin and Welburn, 1988).

3.1.3 Investigation Process

To investigate the mechanisms triggering and regulating the maturation of established trypanosome infections in tsetse, the gene expression profile of established parasite populations were examined.

Genes differentially expressed between cell populations can give insights into the understanding of the biological process occurring under specific stimuli or during specific stages of cellular development and are mostly regulated by messenger RNA (mRNA) expression. mRNAs can be easily translated into cDNA and their abundance can readily be analysed using a combination of technical methods and bioinformatics tools (Davies et al 1999). Characterisation of differentially expressed genes are mostly investigated at the level of mRNA expression using numerous methods such as Micro-array technology , differential display methods, serial analysis of genes expression (SAGE) and sequencing of cDNA libraries.

Rapid amplification of differentially expressed sequences (RADES-PCR) was first described by Murphy (1994). This method was successfully applied to the identification of differentially expressed transcripts between long slender dividing form and non-dividing stumpy forms of *T. brucei* spp. RADES-PCR differs from the basic differentially displayed PCR (DD-PCR) technique in one important aspect; RADES-PCR includes an enrichment step for double stranded cDNA (dscDNA). This step uses the common splice leader; miniexon found in all *T. brucei* mRNAs (De Lange et al 1984; Parsons et al 1984) to enrich for parasite-specific transcripts and it was speculated

that this step facilitates the enrichment of parasites transcript over that of the host (Murphy *et al.,* 1996). Similar techniques based on the splice leader have been previously described *i.e.* the transcription of parasite mRNAs over host mRNAs. Splice leader gene (SL) specific sequences have been used to isolate *Phytomonas spp.* from plants and phytophagous insects (Serrano *et al.,* 1999). In their attempt to characterise differentially expressed gene in male and female *Brugia malayi*, Michalski (1999) used splice leader differential display PCR (SL DD-PCR). SL DD-PCR has also been successfully used in the analysis of differential gene expression in *Echinococcus multilocularis* (Brehm et al 2003) with the method being of particular importance when there is a limited amount of starting material. A full length stage specific cDNA library was also constructed even in the presence of host material contamination using spliced leader (SL) and oligdT primers to survey gene expression during the third larval stage of the human infective filarial nematode *B. malayi* (Blaxter 1996).

Although the process of establishment of parasites in the midgut has been under great scrutiny, little is known about the mechanisms triggering the transition of established procyclics into mammalian infective metacyclic forms. The RADES-PCR technique was used to identify genes differentially expressed between an established midgut population presenting a salivary gland infection and an established midgut population with no salivary gland infection

3.2 Material and methods

3.2.1 Dissection and collection of infected tsetse fly guts

Glossina morsitans spp were infected with *T. b. brucei* Buteba 135 (But 135) as describe in section 2.4.2. Flies were subsequently dissected after 28 days post infection as described in section 2.4.3. Gut and salivary gland infections were confirmed by microscopy. Groups of 10 infected guts were divided into two groups upon the presence or absence of salivary gland infections as follows:

Group (++) - flies with infected guts and infected salivary glands.

Group (+/-) - flies with infected guts and no apparent salivary gland infections.

A third Group (--) representing non-infected guts dissected from flies fed on non-infected blood meal was considered as a control. Fly guts were collected and a total RNA extraction was carried out on each group as described in section 2.5.1.

3.2.2 Differential Display (DD) Methods

3.2.2.1 cDNA Synthesis

3.2.2.1.1 First strand cDNA synthesis for DD:

First strand cDNA was synthesised using SuperScript™ III Reverse Transcriptase (Invitrogen) In a sterile 1.5 ml Eppendorf tube, 500 ng of total RNA of group of each group (++, +-, --) were mixed with 1 µl of 10 mM dNTP Mix (10 mM each dATP, dGTP, dCTP and dTTP at neutral pH); 500 ng olig (dT) to a final volume of 13 µl, the reaction mixture was heated at 65°C for 5 min. The reaction mixture was then incubated with 4 µl 5x First-Strand buffer [250 mM Tris-HCL (pH 8.3 at room temperature), 375 mM KCL, 15 mM $MgCl_2$] 1 µl 0.1 M DTT ; 1 µl Rnase OUT Recombinant Rnase Inhibitor (40 units/µl) , 1 µl of 200 µg/ml of OligodT (Promega) and 1µl of SuperScript III RT (200 units/µl) . The first strand was synthesised by incubating the mixture in a thermocycler (GRI) at 50°C for 60 min. The synthesis was terminated by heat-deactivation of the reverse transcriptase for 15 min at 70°C

3.2.2.1.2 Double stranded cDNA synthesis

The synthesis of double stranded cDNAs from total RNA of trypanosomes was performed using oligo (dT) and miniexon primer. Miniexon represents a common 5' region shared in all trypanosome mRNAs (De Lange *et al.* 1984; Parsons *et al.* 1984) (Figure 3.1). PCR reactions were carried out in a final volume of 50 µl in 0.2 ml PCR tubes (Alpha Laboratories Ltd), DNA amplification was performed in a Peltier thermal Cycler DNA Engine DYAD™ (GRI). 1 µl of single stranded cDNA was added to 5 µl of 10 X Hotstart PCR buffer (Tris-Cl, KCl, $(NH_4)_2SO_4$, 15 mM $MgCl_2$; pH 8.7), 1 µl of 10 mM dNTPs , 1 µl of 10 µM Oligo(dT),1 µl of 10 µM Miniexon and 0.25 µl of 5 units/µl Hotstart DNA polymerase (Qiagen). DNA amplification followed the program in Table 3.1

Group (+ +) Group (+ -)

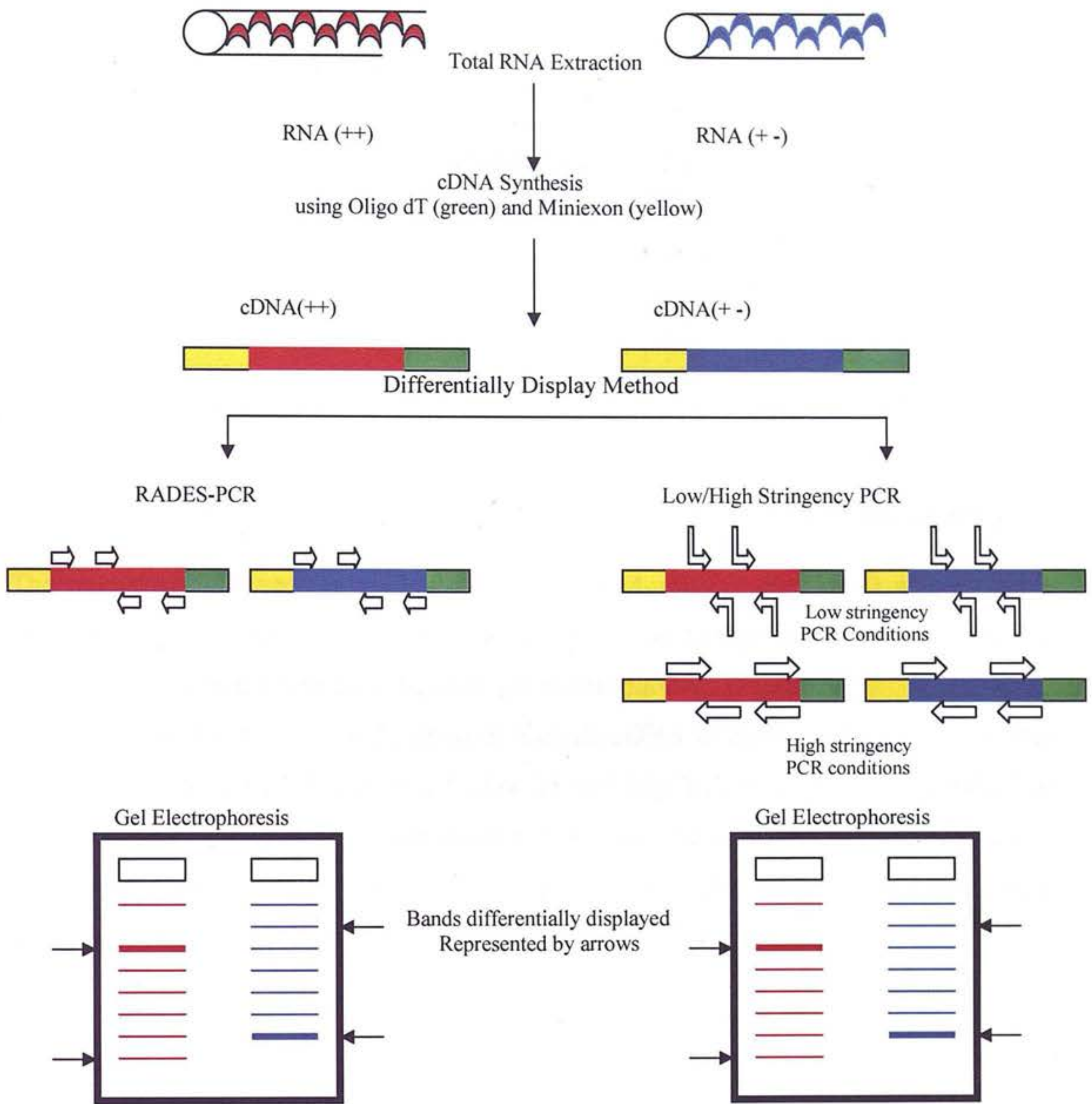


Figure 3.1

Schematic representation of Rapid Amplification of Differentially Expressed Sequence and low/high stringency PCR technique used during the investigation of differentially expressed genes in gut group (++) and (+-).

Double stranded cDNA was synthesised using total RNA extracted from infected fly guts double stranded was amplified using either 10-mer random primer in RADES-PCR and anchored random primer in low/high stringency. Differentially expressed PCR products identification was carried out by electrophoresis on agarose gel.

Table 3.1 1: Amplification steps used for the synthesis dscDNA from group (++, +-, --)

Step	Time (seconds)	Temperature	
Enzyme Activation	900	95°C	
Denaturation	30	94°C	x34
Annealing	30	42°C	
Extension	60	72°C	
Final extension	300	72°C	

3.2.2.2 Rapid Amplification of Differential Expressed Sequences Poly chain reaction (RADES-PCR)-

The procedure followed the method described by Murphy *et al.,* (1995) where double stranded cDNA is amplified by PCR using 10-mer random primers, and the resulting PCR products are resolved using agarose gel electrophoresis (Figure 3.1).

In a 0.2 ml PCR tube, 1 µl of 20 ng/µl of each ds cDNA of each group (++; +-; --) was mixed with 2 µl of 10 X buffer B, 1.6 µl of 25 mM $MgCl_2$, 0.8 µl of 5 mM dNTPs, 1 µl of 20 ng/µl of each random primer (section 2.3.3) and 0.2 µl of 2.5 Units/µl *Taq* DNA Polymerase (Promega), the reaction volume was completed to 20 µl with deionised water. The DNA amplification followed the program in Table 3.2.

Table 3.2 Amplification steps used for RADES

Step	Time(seconds)	Temperature	
Initiation step	300	95°C	
Denaturation	30	94°C	x34
Annealing	30	50°C	
Extension	60	72°C	
Final extension	300	72°C	

3.2.2.3 Low/high Stringincy PCR

The method was based upon the description of Diachenko *et.al* (1996) where an increase in the length of random primer and a combination of low and high stringency PCR parameters were applied to minimise the incidence of artefacts and false positives (Figure 3.1).

In a sterile 0.2 ml PCR tube, 1 µl of the double stranded cDNA from each group was mixed with 0.1 µM of one primer P_n (Section 2.3.2) , 2.5 µl 10 x BD Advantage 2 PCR Buffer (Clontech), 1 µl of 10 mM dNTPs and 0.5 µl of 50 x BD Advantage 2 Polymerase. The reaction mixture was completed to a final volume of 25 µl with deionised water. PCR amplification parameters are as described in Table 3.3.

Table 3.3 Amplification steps used for low/high PCR

<table>
<tr><th>Step</th><th>Time(seconds)</th><th colspan="2">Temperature</th><th></th></tr>
<tr><td>Denaturation</td><td>300</td><td>94°C</td><td rowspan="3">x 1</td><td rowspan="6">Low stringency</td></tr>
<tr><td>Annealing</td><td>300</td><td>40°C</td></tr>
<tr><td>Extension</td><td>300</td><td>68°C</td></tr>
<tr><td>Denaturation</td><td>120</td><td>94°C</td><td rowspan="3">x 2</td></tr>
<tr><td>Annealing</td><td>60</td><td>60°C</td></tr>
<tr><td>Extension</td><td>120</td><td>68°C</td></tr>
<tr><td>Denaturation</td><td>60</td><td>94°C</td><td rowspan="3">x.22</td><td rowspan="3">High stringency</td></tr>
<tr><td>Annealing</td><td>60</td><td>60°C</td></tr>
<tr><td>Extension</td><td>60</td><td>68°C</td></tr>
<tr><td>Final extension</td><td>420</td><td>68°C</td><td></td><td></td></tr>
</table>

3.3 Results

3.3.1 Isolation of Total RNA from *Glossina morsitans morsitans* infected guts

Total RNA extraction was performed on each group of gut (++, +-,--) as described in section 2.5.1. RNA concentrations were calculated as described in section 2.5.4.

The RNA integrity from group (++, +-) was assessed by agarose electrophoresis using 500 ng of total RNA, as describe in 2.5.5. The 28 S and 18 S RNA species were present at 2:1 ratios confirming the integrity of the total RNA preparation; these bands are located between 1.3-1.9 kb and can be seen on Figure 3.2 The presence of genomic DNA contamination was revealed in the RNA preparation as indicated by the high molecular weight bands also visible on Figure 3.2 (Gel A, lanes 1 and 2). A DN*ase* treatment step was undertaken as described in section 2.5.2, which resulted in the elimination of the residual genomic DNA contamination from the RNA preparations (Figure 3.2, Gel B, lanes 1 and 2).

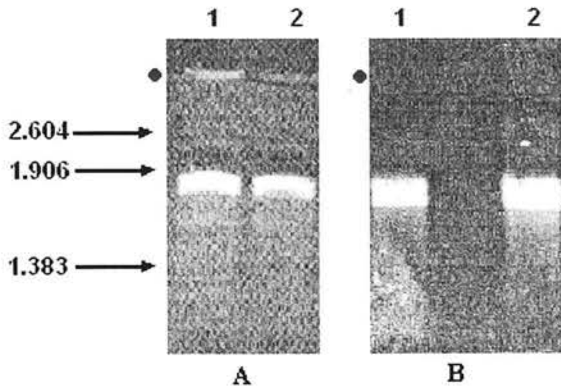


Figure 3.2 *Glossina T.b .brucei* strain BUT 135 infected midgut RNA gel electrophoresis

500 ng of Total RNA was run for 90 minutes at 80 V on a 2 % $NaHPO_4$ agarose gel containing ethidium bromide at a final concentration of 10ng/ml and visualised under a UV transluminator. Lane 1 contains Group (++) and lane 2 Group (+-). The arrows on the side of the gel denote sizes of molecular weight markers in kb.

Gel A revealed a carry over of Genomic DNA in both lane 1 and lane 2 as shown with red dots
Gel B represents the same samples re-run after DNA*ase* treatment which resulted in the digestion of the genomic DNA.

3.3.2 Enrichment for *T.b. brucei* double stranded cDNA (dscDNA) synthesis from Total RNA of infected gut of *G.m. morsitans* using miniexon and oligodT primer sequence

Single stranded cDNA and double stranded cDNA (dscDNA) synthesis was carried out on total RNA from each group (++, +-, --) as described in section 3.2.2.1. 100 ng of dscDNA from each group was run on 2% agarose gel (2.6.1). Figure 3.3 shows the smear of PCR products represents the enriched *T.b. brucei* dscDNA transcripts from group (++) (Figure 3.3 lane 1) and group (+-) (Figure 3.3 lane 2), the products sizes vary from 0.5kb to 5kb. The recommended cleaning step resulted inevitably in slight loss of products (Figure 3.3 Lanes 3 and 4). DscDNA obtained from the control group (--) also showed a smear of enriched PCR products between 0.5 to 5 kb (Figure 3.4), indicating that the host material is also amplified by this method and that in this instance the enrichment step does not specifically recognise parasite material.

A temperature gradient PCR was performed in order to minimise amplification of unspecific transcripts in the control group (--). The increase of temperature seem to eliminate the smear of unspecific products over an annealing temperature of 46°C (Figure 3.4 lane 7) however it was also noticed that such high annealing temperature could not be applied to the current experiments since no transcripts could be obtained using either group (++,+-).

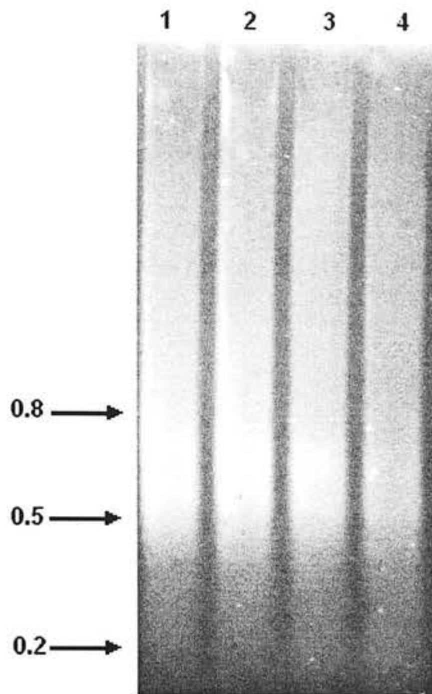


Figure 3.3

Gel electrophoresis of dscDNA obtained with total RNA extracted from guts group (++) and gut group (+-)

Enriched dscDNA gel electrophoresis obtained using 100 ng of total RNA extracted from insect gut. dscDNAs were resolved on 2 % agarose gel and run for 90 minutes in 1x TAE buffer containing ethidium bromide at a finale concentration of 10 ng/ml. Arrows on the side of the gel denote sizes of molecular weight markers in Kb

Lane 1 represents 5 µl of dscDNA obtained from group (++)

Lane 2 represents 5µl dscDNA from group (+-)

Lane 3 represents 5 ul of dscDNA from group (++) after cleaning

Lane 4 represents 5µl of dscDNA from group (+-) after cleaning.

An apparent decrease of product recovery is noticeable in lane 3 & 4 revealing a residual loss in PCR products during the cleaning step

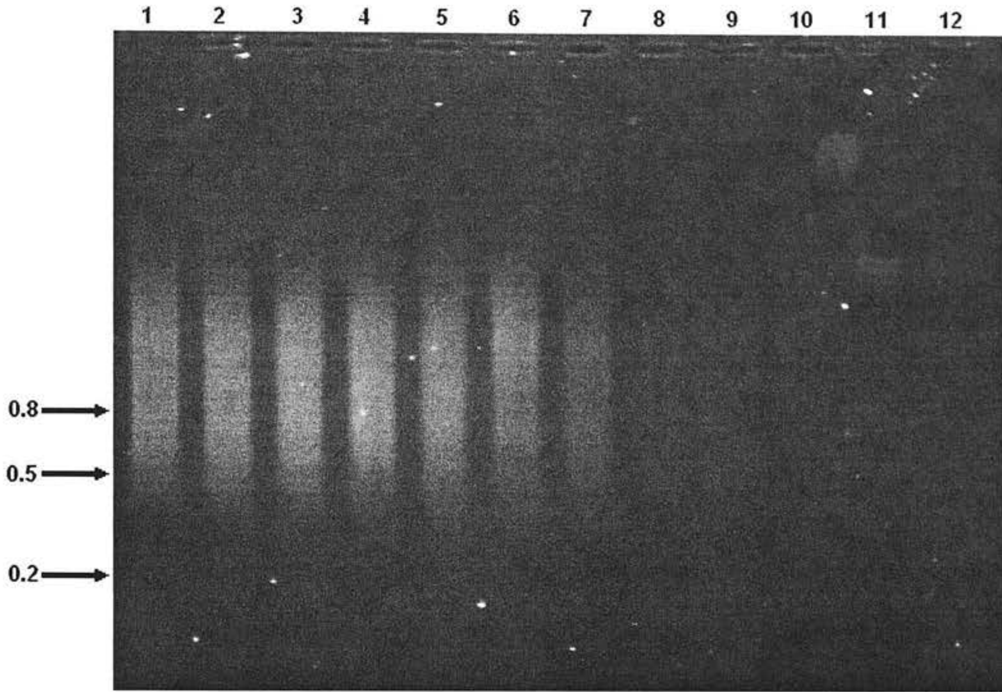


Figure 3.4

Gel electrophoresis of dscDNA obtained with total RNA extracted from control guts (group --)

Gel electrophoresis of enriched dscDNA from total RNA obtained from non-infected fly gut. The PCR products were obtained using a temperature gradient of 40° to 50°C using Peltier thermal cycle (DNA Engine DYAD™).

Lane 1 to 10 represent 100 ng of enriched dscDNA synthesis products run on 2 % Agarose/ 1x TAE . A smear of PCR products representing enriched dscDNA products could be clearly seen between Lanes 1-7 corresponding to an annealing temperature between 40-46°C. An annealing temperature over 46°C resulted either into a poor enrichment (lanes 8 and 9) or no products amplification (lane 10 to 12).

3.3.3

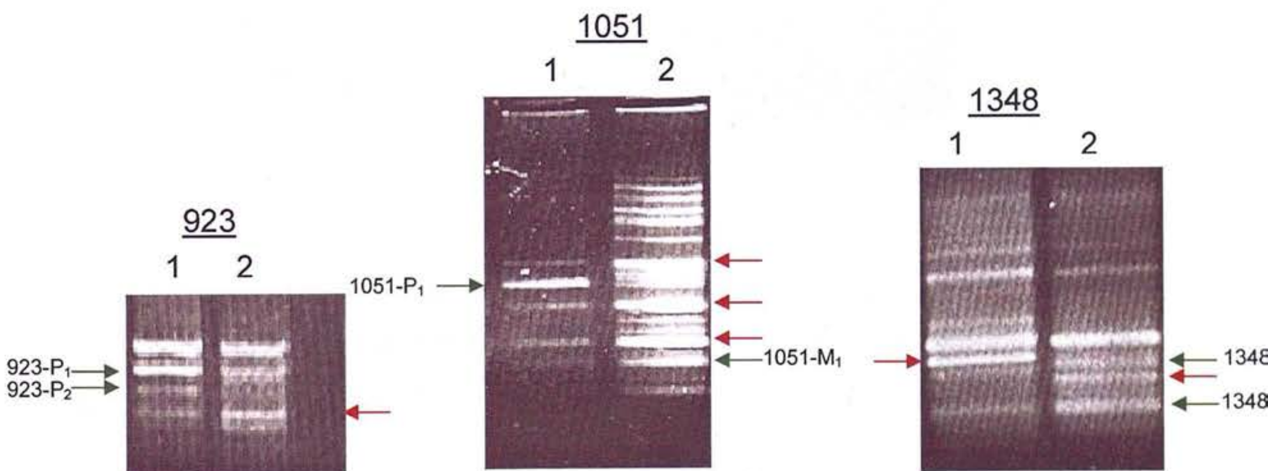
RADES-PCR reactions were performed as described in section 3.2.2.2, the PCR products from each group, obtained with the same sequence primer, were resolved alongside each other on a 2% agarose gel. Differentially expressed transcripts were selected on the basis of appearance, disappearance and/or a difference of two fold or more between the band intensity (Figure 3.1). Generally the size of the RADES-PCR products obtained varied between 0.2 to 0.8 kb and each random primer sequence identified more than one differentially expressed PCR products between group (++, +-). 8 random primers were used and they revealed 28 RADES-PCR products between group (++) and group (+-).

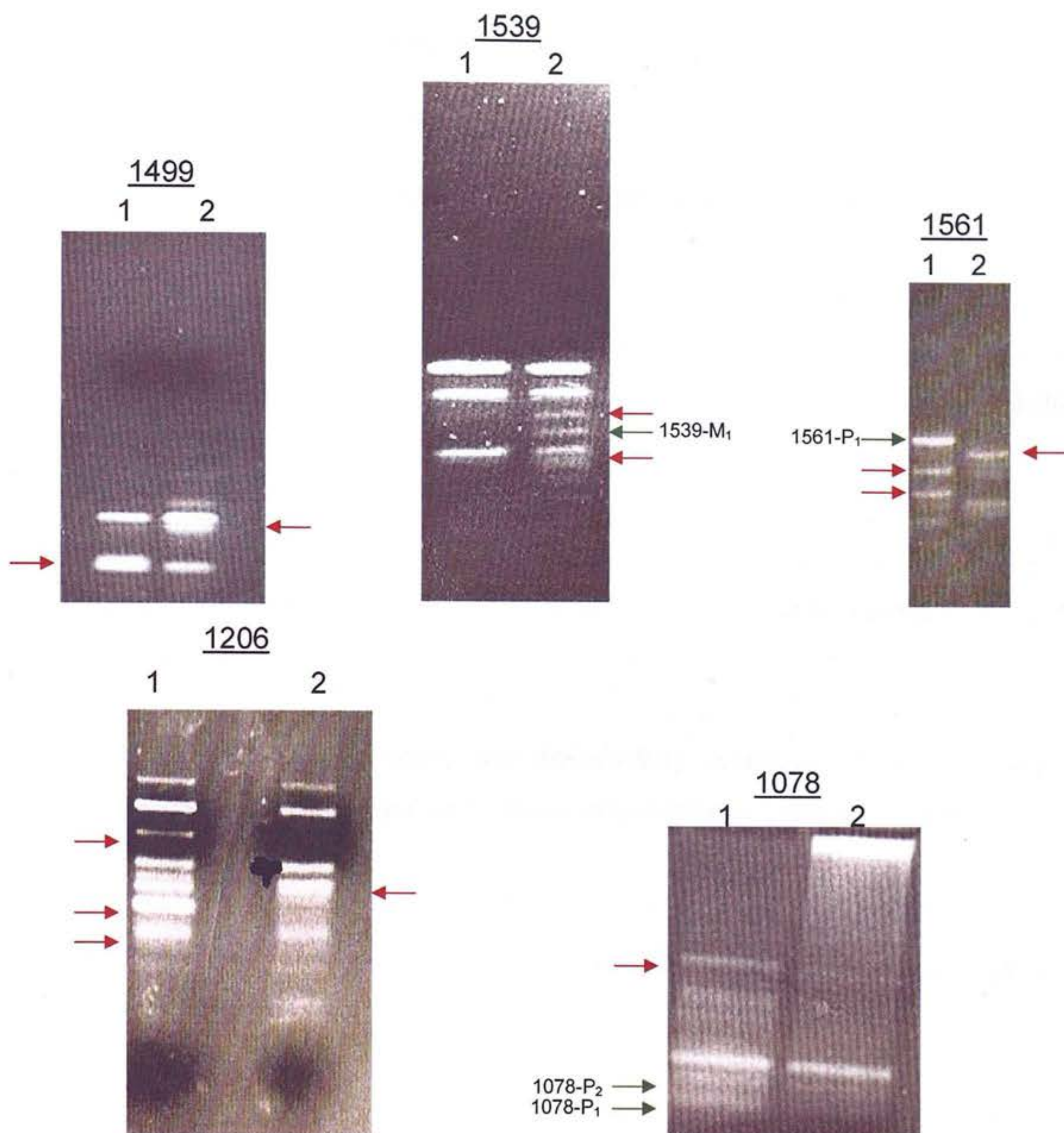
Figure 3.5

Gel electrophoresis of the RADES-PCR products.

The numbers at the top of each gel image denotes the sequence of the random primer (RP) used. Lanes 1 and 2 on each image represents RADES-PCR products obtained with group guts (++) and (+-) respectively. The product of each RP are resolved side-by-side on a 2% (w/v) Agarose, gels were run for 90 minutes at 80 V and visualised under an UV-transiluminator. Values at the side denote the molecular size markers used in kb.

PCR products deemed differentially expressed between group (++) and group (+-) are shown by arrows where green arrow arrows represent RADES-PCR products that were successfully cloned and sequenced and red arrows represent RADES-PCR products that could not be identified.





Primer sequence 923 revealed three differentially expressed PCR products between group (++) and group (+-) (Figure 3.5 - 923).2 RADES-PCR products were isolated in group (++) (Figure 3.5- 923 lane 1) and 1 PCR product was isolated in group (+-) (Figure 3.5 - 923 lane 2).

Primer sequence 1051 revealed one differentially expressed product in group (++) (Figure 3.5 - 1051 lane 1) 4 differentially expressed products in group (+-) (Figure 3.5 - 1051 lane 2).

Primer sequence 1078 revealed three differentially expressed products a all expressed in group (++) (Figure 3.5 - 1078 lane 1).

Primer sequence 1206 revealed three differentially expressed products in group (++) (Figure 3.5 - 1206 lane 1) and one differentially expressed product in group (+-) (Figure 3.5 – 1206 lane 2).

Primer sequence 1348 revealed four differentially expressed products. One PCR product in group (++) (Figure 3.5 - 1348 lane 1) and three PCR products in group (+-)(Figure 3.5 - 1348 lane 2).

Primer sequence 1499 revealed one differentially expressed product in group (++) (Figure 3.5 1499 - lane 1) and one differentially expressed product in group (+-) (Figure 3.5 - 1499 lane 2).

Primer sequence 1539 revealed three differentially expressed products in group (+-) (Figure 3.5 – 1539 lane 2).

Primer sequence 1561 revealed four differentially expressed products three PCR products in group (++) (Figure 3.5 - 1561 lane 1), and one PCR produc in group (+-) (Figure 3.5 - 1561 lane 2).

3.3.3.1 Re-amplification and cloning of identified RADES-PCR products

Re-amplification of the identified PCR products was carried out on gel plugs purified as described in section 2.6.2. Ligation and transformation were carried out as described in section (2.7.1/2). Plasmid purification was carried out as described in section (2.7.3.3). From the 28 RADES-PCR initially identified, 9 products were successful cloned and sequenced. 19 RADES-PCR products failed to re-amplify and were deemed artefacts.

3.3.3.2 Blast Search results of differentially expressed transcripts between group (++) and group (+-).

The characterisation of RNA transcripts identified as differing between the two established gut populations of *T. brucei* trypanosomes, Group (++) with an established mature transmissible infection and Group (+-) which had not established any apparent infection were investigated using the Blast search as described in section 2.8.2 and Table 3.3, sequence DNA was prepared as described in section 2.8.1.

Sequence primer 923 identified 3 transcripts differentially expressed between group (++) and group (+-) (Figure 3.5 – 923). Two of those transcripts referred to as 923- P_1 (537 bp) and 923-P_2 (694 bp) (Figure 3.5 – 923 lane 1) showed strong homology on the *T. brucei* sequence database. The 923-P_1 Blast-x search showed a significant hit on the non-redundant proteins database with hypothetical protein *T. brucei* 927.2.3610 (Score = 268 bits, Expect = $6e^{-71}$). The 923-P_2 Blast-x search revealed a strong hit to Tb07.33N13.310 a conserved hypothetical protein in *T. brucei* (Score = 379.6 bits, Expect = $2.0e^{-108}$). The third transcript, identified in group (+-) (Figure 3.5 - 923 lane 2) failed to re-amplify.

Sequence primer 1051 identified 5 transcripts differentially displayed between the group (++) and the group (+-) (Figure 3.4 – 1051) however, only two products were re-amplified and sequenced. One transcript was identified in the group (++) (Figure 3.3.e. lane 1) will be referred to as clone 1051-P_1 and one transcript was identified in the group

(+-) (Figure 3.3.e.lane 2) and will be referred to as clone 1051-M_1. The sequencing of clone 1051-P_1 revealed a transcript of 563 bp and the sequencing of clone 1051-M_1 produced a transcript of 290 bp. Blast x search with sequence clone 1051-P_1 revealed a significant hit to a multidrug efflux protein [*Yersinia pestis biovar Medievalis* str. 91001] (Score = 83.2 bits, Expect = 1e-^{22}) and blast x search with sequence clone 1051-M_1 revealed a significant hit to Ornithine/acetylornithine aminotransferase [*Cytophaga hutchinsonii*] (Score = 132 bits, Expect = 3e-^{30}).

Sequence primer 1078 identified 3 differentially transcripts group (++) (Figure 3.3.d lane 1) two transcripts were cloned and sequenced. The clones referred to as 1078 P_1 & 1078 P_2 (354 bp and 398 bp respectively).1078 P_1 sequence was of poor quality and consequently did not reveal any significant similarities when subjected to the database search. Blast x search with sequence clone 1078 P_2 revealed a hit with conserved hypothetical YjhS family protein encoded by cryptic prophage CP-933P [*Escherichia coli* O157:H7 EDL933] (Score = 52.0 bits , Expect = 4e-06).

Product identified by sequence primer 1206 did not re-amplify and could not be sequenced and characterised.

Sequence primer 1348 identified 4 transcripts, 3 differentially expressed products in group (+-). Two products were cloned and sequenced referred to as clone1348 M_1 and clone 1348 M_2 (Figure 3.5 – 1348 lane 2) and one differentially expressed product in group (++)(Figure 3.5 – 1348 lane 1) which did not re-amplify. Identified clones 1348 M_1 and 1348 M_2 sequencing revealed transcripts of the size of 307 and 436 bp respectively. A Blast x search with the sequence clone 1348 M_1 found no significant similarity and Blast x search with clone 1348 M_2 revealed a strong hit to Tb04.2L9.920 ,a hypothetical protein [*Trypanosoma brucei*] (Score = 249.4 bits, Expect = 3.2e^{-69}).

Product identified by sequence primer 1499 did not re-amplify and could not sequenced and characterised.

Sequence primer 1539 identified three differentially expressed product in group (+-) (Figure 3.5 - 1539 lane 2).only one transcripts was sequenced and characterised and will be referred to as clone 1539 M_1. The sequencing of 1539-M_1 revealed a clone of 307 bp. A blast x conducted with of 1539-M_1 sequence identified a putative type 4 fimbrial pilin related transmenbrane protein [*Ralstonia solanacearum GMI1000*] (Score = 125 bits, Expect = $4e^{-28}$).

A Blast n on the EST database (Compares the nucleotide query sequence against a nucleotide sequence database) identified a transcript of trypanosome origin (Score = 50.1 bits (25), Expect = 0.003). The similarity between the query transcript and the subject transcript was an overlap of 20 amino acid stretch at the 5' end of the query. The transcribed subject (AA023881) was identified during the study of gene differentially expressed in *T.brucei brucei* using RADES-PCR technique (Osanya *et al.*, 1996). It was deduced that the homology between the query and the subject transcripts could be either due to the miniexon sequence stretch or due to the same sequence primer since the sequences overlap at their extremity. However, miniexon sequence was not identified when examining the transcripts ends neither was the andom primer sequence 1539. Contig was constructed between clone 1539-M_1 and the query sequence and was reconducted. No homology was detected on the extended contig.

Sequence primer 1561 identified three differentially expressed products in the group (++) (Figure 3.4 - 1561 lane 1) and one differentially expressed product in group (+-) (Figure 3.4 – 1561 lane 2). Only 1 product re-amplify and could be sequenced in group (++) referred to as clone 1561-P_1. This clone revealed a size transcript of 554 bp. Blast x search revealed a significant hit (s =180 bits, Expect = $1e^{-44}$) with an outer membrane receptor for monomeric catechols of *Xylella fastidiosa*.

3.3.4 Amplification of differentially expressed sequences between group (++) and group (+-) using low / high stringency -PCR

In an attempt to decrease the incidence of artefacts and increase PCR reaction efficiency we followed a protocol initially described by Diachenko (1996) which uses longer random primer sequences (anchored-random primer 25-29 mers), optimal dNTPs concentration and PCR amplification conditions which are of a higher stringency . A greater number of PCR products were observed under these new conditions with products sizes up to 2 kb (Figure 3.5).

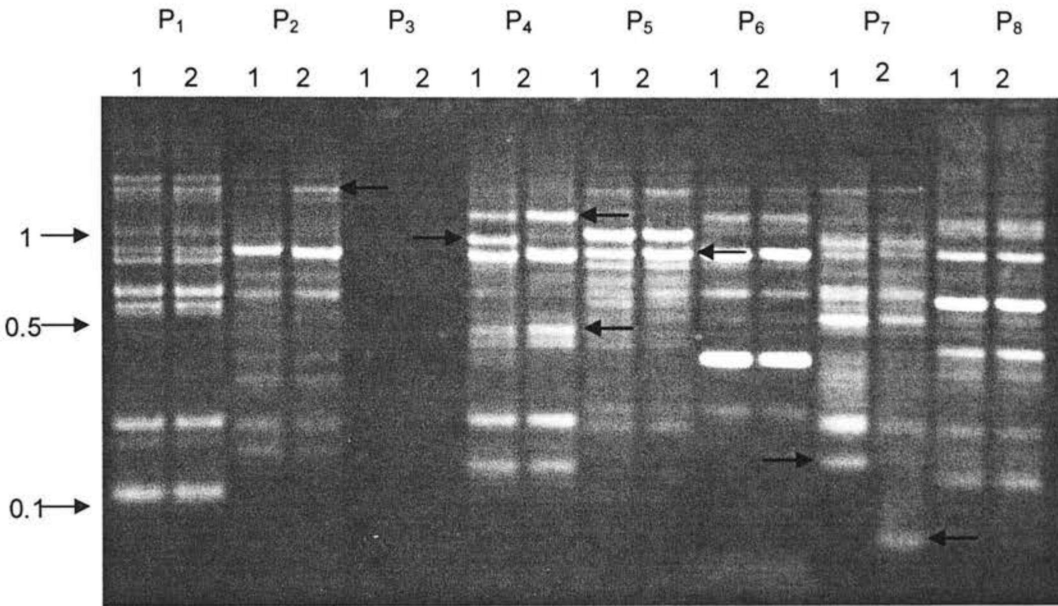


Figure 3.5.1

Gel electrophoresis of low/High stringency DD-PCR products obtained on cDNA prepared on RNA extracted from gut group (++) and gut group (+-)

8 anchored random primer (P_{1-8}) were used on group (++ lane 1) and group (+- lane 2) cDNA . Arrows represents differentially displayed PCR products detected between the two groups of infected guts

P_2 sequence primer revealed one PCR product differentially displayed in the group gut (+-) (Figure 3.6- P_2-lane 2).

P_4 sequence primer revealed one PCR product differentially displayed in group gut (++)(Figure 3.6-P_4- lane 1) and two PCR products in group (+-) (Figure 3.6-P_4- lane 1)

P_5 sequence primer revealed one PCR product differentially displayed in group gut (+-)(Figure 3.6-P_5- lane 2).

P_7 sequences primer identified one PCR product differentially displayed in group (++) (Figure 3.6 - P_7 - lane 1) and PCR orudtc in group (+-) (Figure 3.5 - P_7 - lane 1). None of the PCR products identified could be successfully re-amplified and cloned for sequencing.

Name	Length	Blast x (*T.brucei* Omni BLAST & NCBI)	Organism	Gut group	S	E-value
923-P_1	534 bp	Similar to hypothetical protein Tb 927.2.3	*T brucei*	++	268	6 e^{-71}
923-P_2	694 bp	Conserved hypothetical proteinsTb07.33N13.31	*T brucei*	++	379.6	$2.0e^{-108}$
1561-P_1	554 bp	Outer membrane receptor for monomeric catechols	*Xylella fastidiosa*	++	180	$1e^{-44}$
1348-M_1	307 bp	No significant homology				
1348-M_2	436 bp	conserved hypothetical Tb04.2L9.920	*T brucei*	+-	249.4	$3.2e^{-69}$
1078-P_1	354bp	Poor sequence data				
1078-P_2	398b	conserved hypothetical YjhS family protein encoded by cryptic prophage CP-933P	*E.ColiO157:H7 EDL933*	++	52	$4e^{-06}$
1051-P_1	563bp	Multidrug efflux protein	*Yersinia pestis biovar Medievalis*	++	83.2	$1e^{-22}$
1051-M_1	290b	Ornithine/acetylornithine aminotransferase	*Cytophaga hutchinsonii*	+-	132	$3e^{-30}$

Table 3.3

The RADES clones sequences were compared with database entries using the blast x algorithm against the non-redundant database (http://www.ncbi.nlm.nih.gov/BLAST/) (non-redundant protein translation of GenBank,; Protein Data Bank; SwissProt; Protein Identification Resource Database; Protein Research Foundation Sequence database), and *Trypanosoma* OminBLAST Server(http://www.sanger.ac.uk/cgi-bin/blast/submitblast/t_brucei/omni). The higher score and the lowest Expect values constituted the best match. The length represents the size of the clone transcript after sequencing

3.4 Discussion

The work carried out in this chapter was aimed at investigating the potential differences that might exist between established gut populations of *T.b. brucei* in *Glossina m morsitans* depending on the presence and/or the absence of a mature infection. In order to do this RADES-PCR and low/high stringency PCR were used to investigate differentially displayed genes by looking at the presence of mRNA transcripts. 28 were identified transcripts that differed between these two populations, of which 10 we characterised.

Two transcripts (923-P_1 and 923-P_2) were shown to be present in group (++) *i.e.* the salivary gland infective group; these transcripts were shown to have homology with hypothetical proteins in the *T. brucei* database. On the other hand, transcript 1348-M_2 was shown to be present in group (+-) *i.e.* non-infective salivary gland group; the transcript sequence also overlapped with a hypothetical protein in *T .brucei* database. Gull (2001) characterised these hypothetical genes as "Orphan genes"; he suggested that such genes might bear unique parasitical functions such as parasitism; virulence; survival in environment; host interactions; vector interactions; signaling and communication. However speculation over the expression status of these particular transcripts between the gut groups studied cannot be certain at this stage of the experiment since a large incidence of false positives are usually picked up by differentially display techniques. Therefore, quantification analysis of the identified transcripts between group (++) and group (+-) is imperative.

1348-M_1 clone did not show any significant homology on any of the databases searched. Such transcripts might either represent expressed sequence tag (EST) from the 3'mRNA un-translated region or the clone sequence codes for a product which have not yet been identified

The remaining transcripts showed high homology to prokaryote genes (1561-P_1; 1078-P_2; 1051-P_1; 1051-M_1 and 1539-M_1). Tsetse guts harbor two symbiotic microorganisms

the obligate primary-symbiont *Wigglesworthia glossinidia* and the commensal secondary symbiont *Sodalis glossinidius* (Akman *et al.*, 2002). These microorganisms seem to enable their host to live on restrictive diet, providing compounds related to the B vitamin complexes (Nogge *et al.,* 1981). These transcripts must have been non-specifically amplified during the dscDNA synthesis as they do not originate from *T. brucei* DNA. PCR products were seen to be amplified in the control group (--) and these transcripts may have originated from tsetse gut bacteria that occur normally within the tsetse gut. This could be the result of un-specific binding of splice leader primer sequence on non-trypansomal material and/ or binding of oligo dT on bacterial RNA since polyadenylation of prokaryote RNAs does take place in a variety of bacteria species (Sarkar, 1997)

DD-PCR was originally described by Liang (1992) and the same concept was being copied to develop RADES-PCR technique. Although relatively simplistic and requiring a little amount of starting material, RADES-PCR technique suffered from the same drawbacks already encountered with DD-PCR techniques, a high incidence of artefacts (Covini *et al.,* 1999) and relatively cumbersome methodology although it has been originally described as a simple technique. Low/ high stringency PCR was undertaken to minimise the incidence of artefacts. Although a significant improvement seemed to be achieved in terms of bands number, failure of re-amplification of the right transcripts were still being encountered.

3.5 Future Work

In terms of identification of differentially expressed sequence, a wider range of primer combination should be performed, PCR steps should carried out using hot probe and the PCR products should be resolved on a gel sequencer for better resolution since artefact might be due of co-migration bands. Identified products should imperatively undergo a quantification analysis since DD-PCR has a large incidence of false positive. Since this technique is not exclusive to parasite material as shown on Figure 3.5.

Chapter 4

Identification of genes differentially expressed in response to treatment of *T. brucei* procyclic 8-Br-cGMP

4.1 Introduction

Although cGMP dependant pathway has yet to be characterised in *T. brucei* parasites, Macleod (in press 2005) reported that cGMP could play a role in the survival of *T. brucei* during the establishment in the tsetse fly vector. Therefore the effect of the cGMP analogues 8-Br-cGMP on *T.b. brucei* gene expression was investigated using PCR-Select™ cDNA Subtraction (Clontech). The method combines suppression subtractive hybridisation (SSH) and selective PCR amplification to provide a full set of differentially expressed genes, enriched for low abundant differentially expressed genes. The method has been successfully used in the identification of differentially expressed genes in various biological systems for example; the technique was carried out to identify genes involved in the immune response of adult female *Anopheles gambiae* mosquitoes in response to infection with bacteria and malaria parasites (Oduol *et al.*, 2000). It was also used in the identification of induced genes during oocyst development of *Plasmodium berghei* in *Anopheles stephensi* (Srinivasan *et al.*, 2003) finally the technique has proven to be successful in the identification of genes differentially expressed during the establishment of *T. brucei* in the tsetse fly (Hao *et al.*, 2001).

4.2 Material and Methods

All the primer sequences used in this section can be found in section 2.3.1. Gel electrophoresis conditions were carried out as described in section 2.6.1.

4.2.1 Culture treatment

Parasite cultures of *T.b. brucei,* stock Buteba 135 (But 135), were prepared to a concentration of 10^7 parasites per ml as described in section 2.4.3. These cultures were then halved to provide 2 flasks of identical parasites. One of which was treated with 8-Br-cGMP to a final concentration of 100 μM; these will be referred to as the "Treated procyclic form (TP)" preparation. The second flask or the "Untreated procyclic form (UP)" was treated with the solvent that had been used to prepare the 8-Br-cGMP solution, section 2.4.5.

4.2.2 RNA extraction

4.2.2.1 Total and pol A^+ RNA extraction from TP and UP:

Total RNA extraction from the TP and UP preparations was carried out as described in section 2.5.1. A DN*ase* treatment step was carried out on the total RNA samples prior to poly A^+ isolation as described in section 2.5.2. The poly A^+ RNA isolation from the total RNA preparations as described in section 2.5.3.

4.2.3 Subtracted cDNA Library

The presence of differentially expressed genes resulting from the treatment of procyclic forms with 8-Br-cGMP was investigated using the Clontech PCR-Select™ cDNA Subtraction Kit (Clontech K1804-1)

This method firstly converts the messenger RNA from TP and UP preparations into cDNA. The TP and UP cDNAs are hybridised, resulting in the removal of hybrids sequences. Consequently, the remaining unhybridized cDNAs represent genes that are expressed in the treated procyclic form culture, but are absent from the untreated sample. The genes transcripts are then ligated into a pGEM-T easy vectors® (Promega) which were then used to transform competent cells. The procedure was carried out using the manufacturer instructions.

4.2.3.1 First-Strand cDNA synthesis on poly A^+ mRNA extracted from treated and un-treated procyclic forms.

In a sterile 0.5 ml microcentrifuge tube, 2 μg of poly A^+ mRNA, from either the TP or UP was mixed with 1 μl of "cDNA synthesis primer" (10 μM). The reaction volume was made up to 5 μl with deionised water. The mixture was then spun briefly at room temperature and incubated at 70°C for 2 min in a thermal block. The tubes were chilled on ice for 2 min and again spun briefly. Then, 2 μl of 5 x First-Strand Buffer (250 mM Tris-HCl at pH 8.5, 40 mM $MgCl_2$, 150 mM KCl, 5 mM Dithiothreitol), 1 μl of dNTP Mix (10 mM);1 μl AMV Reverse Transcriptase (20 units/μl) and 1 μl deionised water were added to the mixture and the content of the tube was gently vortexed and briefly

centrifuged. The reaction mixtures were incubated at 42°C for 90 min. The synthesis of the first strand cDNA (sscDNA) was terminated by chilling the reaction on ice.

4.2.3.2 Second-Strand cDNA Synthesis

In two single 0.5 ml sterile eppendorf tubes, double stranded cDNAs (dscDNA) were synthesised by mixing 10 μl of sscDNA of TP or UP with 16 μl of 5x second-strand buffer (500 mM KCl, 50 mM ammonium sulfate, 25 mM $MgCl_2$, 0.75 mM β-NAD, 100 mM Tris-HCl at pH 7.5, 0.25 mg/ml BSA), 1.6 μl of dNTP Mix (10 mM), 4 μl 20 x second-strand enzyme cocktail (DNA polymerase I, 6 units/μl; RNase H, 0.25 units/μl; *E. coli* DNA ligase, 1.2 units/μl) and 48.4 μl of sterile water . The content of the tube was mixed, briefly spun and incubated at 16°C for 2 h. In order to obtain blunt dscDNA ends, a further 30 min incubation was carried out by mixing 2 μl of T4 DNA Polymerase (3 units/μl) with the reaction mixture. The second strand cDNA synthesis was terminated by adding 4 μl of 20 x EDTA/Glycogen Mix (0.2 M EDTA; 1 mg/ml glycogen). The resulting second strand cDNAs were stored at -20°C until further applications.

4.2.3.3 Double strand cDNA purification

To purify dscDNAs, a Phenol: Chlorofom: isoamyl alcohol extraction was performed. One hundred microlitres of phenol: chlorofom: isoamyl alcohol (25:24:1) was added to TP and UP dscDNAs. The reactions were vortexed thoroughly and centrifuged at 14 000 g for 10 min at room temperature. The aqueous phases were transferred to 1.5 ml sterile centrifuge tubes and mixed with 100 μl of chloroform: isoamyl alcohol (24:1), vortexed and then centrifuged at 14 000 g for 10 min at room temperature .The aqueous phases were transferred to sterile 1.5 ml microcentrifuge tubes and mixed with 40 μl of 4 M ammonium acetate and 300 μl of 95 % ethanol. The tubes were centrifuged at 14 000 g for 20 minutes at room temperature. The supernatant fluids were discarded and the resulting pellets were washed with 500 μl of 80 % ethanol. The purified dscDNA pellets were air dried for 10 minutes and re-suspended in 50 μl of deionized water.

4.2.3.4 Rsa I digestion and purification of TP and UP dscDNAs

4.2.3.4.1 Introduction

From the wide repertoire of restriction enzymes, *Rsa* I was chosen because it generated the largest average size of cDNA fragments (600 bp). Firstly a restriction endonuclease digest step was undertaken; this prevents the formation of complex network of long DNA fragments which in turn could prevent hybridisation. Moreover, cutting cDNA in to smaller fragments over-represents differentially expressed genes. It is common that cDNA from the same family of genes have similar coding regions which could form hybrids and be removed during the subtraction process. However, by increasing the number of cDNA fragments from differentially expressed genes, would still allow for non- common coding region to be represented. (Diatchenko *et al.*, 1996)

4.2.3.4.2 Method

In 0.5 ml Eppendorf tube, 43.5 µl of purified TP or UP dscDNA were mixed with 5 µl of 10 x *Rsa* I restriction enzyme buffer (100 mM Bis Tris Propane-HCl at pH 7.0, 10 mM $MgCl_2$, 1 mM DTT) and 1.5 µl of *Rsa* I (10 units/µl). The reactions were vortexed and pulsed centrifuged prior incubation at 37°C for 90 min. The digestion reactions were terminated by mixing 2.5 µl of 20X EDTA/Glycogen Mix (0.2 M EDTA; 1 mg/ml glycogen) to the content of each tube.

The *Rsa* I digestion was followed by phenol: chloroform: isoamyl alcohol purification of the digested dscDNAs as described in section 4.2.3.3. The resulting purified *Rsa* I digested dscDNA pellets were resuspended in 5.5 µl deionized water and stored at – 20°C.

4.2.3.5 Ligation Adpators1 and 2R on dscDNA TP:

The ligation reactions were only carried out on TP dscDNA. This experimental design would only allow the enrichment for the differentially expressed sequences in TP during PCR selective amplification steps (Figure 4.1).

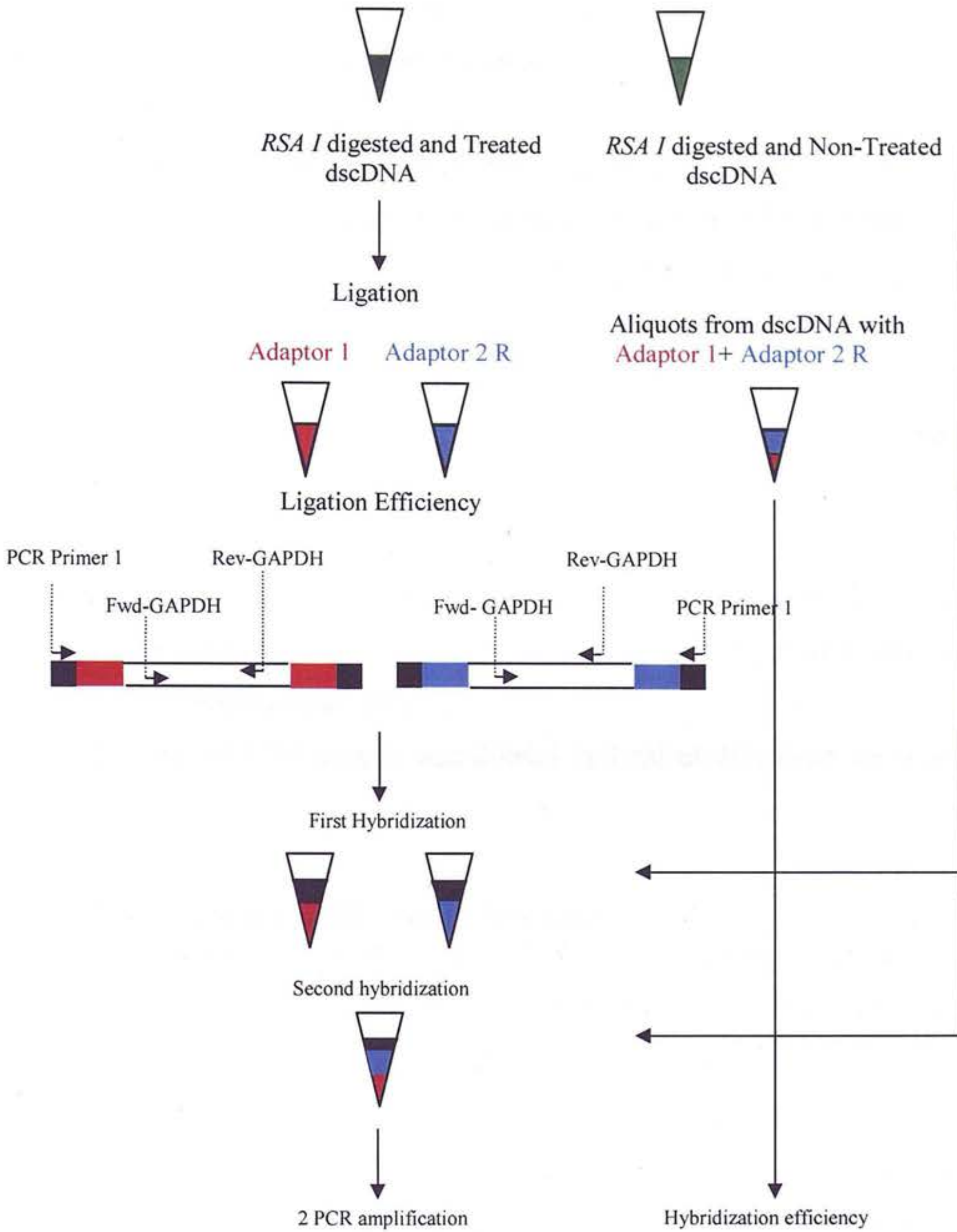


Diagram representing different steps and experimental set-up of PCR select subtracted technique used for the identification of differentially expressed gene in TP sample.
Adaptors 1 and 2 R were only ligated on RSA I digested and treated dscDNA. RSA I digested and non-treated dscDNA sample was used during the hybridisation steps to enrich for rare and differentially expressed transcript specific to dscDNA treated sample.

In a sterile 0.5 ml microcentrifuge tube, 1 μl of the *Rsa* I-digested TP dscDNA was diluted with 5 μl of water. The resulting diluted mixture will now be referred to as dscDNA TP.

In the first 0.5 ml sterile eppendorf tube, 2 μl of dscDNA TP was mixed with 2 μl of adaptor 1 (10 μM). In the second 0.5 ml sterile eppendorf tube, another 2 μl of dscDNA TP were mixed with 2 μl of adaptor 2 R (10 μM). In each tube, 2 μl of 5 x ligation buffer, 1 μl of T4 DNA ligase (400 Units/ μl) and 3 μl of deionized water were also added. In a third 0.5 ml sterile eppendorf tube, 2 μl of the first and second tube reactions were mixed, this resulted in making the unsubtracted treated procyclic form (UTP) sample, which will be used as an internal control for the hybridization.

The reaction mixtures were centrifuged briefly and incubated at 16°C overnight. The ligation reactions were terminated by adding 1 μl of EDTA/Glycogen mix and the ligase enzyme was deactivated by heating the samples at 72°C for 5 min. The adaptor ligated dscDNA TP were stored at -20°C.

One microlitre of UTP sample was diluted in 1 ml of deionized water and stored at -20°C.

4.2.3.6 Ligation Efficiency Analysis

The efficiency of the ligation of the dscDNA TP to adaptors 1 and 2R were tested by DNA amplification of an abundantly expressed transcript *i.e.* GAPDH. This was carried out by combining in two separate PCR reactions GAPDH specific primer (either reverse or forward) with one that spans the adaptor/cDNA junctions (either Adaptor 1 or 2R primers) and secondly GAPDH specific primers as control, this reaction setting would allow, firstly to assess whether the ligation had occurred.and secondly, to assess whether all GAPDH transcripts contained in the sample have both adaptors. The latter was achieved by comparing the intensity of the PCR products obtained from both reactions. If the intensities are equal, this would show that the adaptors ligation had occurred in all the GAPDH transcripts. *T. brucei* GAPDH sequence was obtained by searching the non-redundant nucleotides database on NCBI and specific internal primers were designed using Primer Premier 5 (Biosoft International).

The templates used for DNA amplification were prepared by diluting 1 μl of each adaptor ligated dscDNA TP in 200 μl of sterile water.

To assess whether the adaptor1 was ligated to the dscDNA TP of the first tube, 1 μl of the diluted adaptor1 ligated dscDNA was mixed with 1 μl of PCR Primer 1 (10μM), 1 μl of 3'GAPDH primer (10μM), 2.5 μl of 10X reaction buffer, 0.5 μl dNTPs and 0.5 μl 50x Advantage cDNA Polymerase (Clontech). The reaction was completed to 25 μl with sterile deionised water.

To assess whether adaptor2R was ligated to the dscDNA TP of the second tube, 1 μl of the diluted adaptor 2R ligated dscDNA was mixed with 1 μl of PCR Primer 1 (10 μM), 1 μl of 5'GAPDH primer, 2.5 μl of 10x reaction buffer, 0.5 μl dNTPs , and 0.5 μl 50 x Advantage cDNA Polymerase (Clontech). The reaction was completed to 25 μl with sterile deionised water.

In order to control whether the GAPDH primers amplified the GAPDH transcript contained in each adaptor 1 and adaptor 2R ligated dscDNA TP. An internal control was set as followed: 1 μl of each 5'and 3' GAPDH primers were mixed with 1 μl of the adaptor 1 or adaptor 2R ligated dscDNA TP, 2.5 μl of 10 x reaction buffer, 0.5 μl dNTPs , and 0.5 μl 50 x Advantage cDNA Polymerase (Clontech).

The reaction mixtures were placed in a thermocycler and the DNA amplification procedure was performed as followed (Table 4.1).

Table 4. 1

PCR amplification parameters used to asses the ligation efficiency of Adaptors 1and 2R on TP sample

Step	Time (seconds)	Temperature	
Denaturing	30	95°C	x 25
Annealing	30	65°C	
Extension	150	68°C	

4.2.3.7 Hybridisations

4.2.3.7.1 Introduction:

Two rounds of hybridisation were carried out. The first of which consisted of two separate reactions where an excess of *Rsa I* digested dscDNA UP was added to each of the dscDNA TP ligation preparations and incubated overnight. Four types of hybrid molecules are formed during the first hybridisation (Figure 4.2). An ss cDNA fraction in the TP sample, (Figure 4.2 a) is normalised during the annealing process resulting roughly in an equal concentration of high and low abundant transcripts. This normalisation occurs because these abundant transcripts re-anneal more rapidly due to the second order kinetic of hybridization, the result is the formation of homohybrid mollecules (Figure 4.2 b). Moreover, the type a, molecules are enriched further for differentially expressed transcripts since more common target transcripts are formed with UP sample forming the type c and d molecules (Figure 4.2).

During the second hybridisation, the first hybridisation mixes are combined together with an excess of *Rsa I* digested dscDNA UP. During this stage only the remaining normalised and subtracted ss tester cDNA are able to re-associate (Figure 4.2 b c d and e). Moreover the addition of an excess driver sample enriches more for type e hybrids.

Type e molecules hybrids have exclusive features that differentiate them from the other hybrids formed during the two hybridisations steps (b c d); they bear different sequences at their 5' end, which are represented by sequence of adaptor 1 and 2R. Using complementary primers to these two different sequences in PCR reaction would allow the exponential amplification of over-expressed transcripts in the treated sample. Therefore the subtracted pool is used as template in PCR reaction.

For the first hybridization step, the two components are prepared adding 1.5 µl of UP sample obtained in section 4.2.3.4, to 1.5 µl of each ligated preparation. The mixtures are then added to 1 µl of pre-warmed 4 x hybridization buffer. The reactions were mixed by pipeting up and down and each tube was overlayed with a drop of mineral oil. The

reactions were denatured at 98°C for 1.5 min followed by and an incubation period of 8 h at 68°C in the thermocycler.

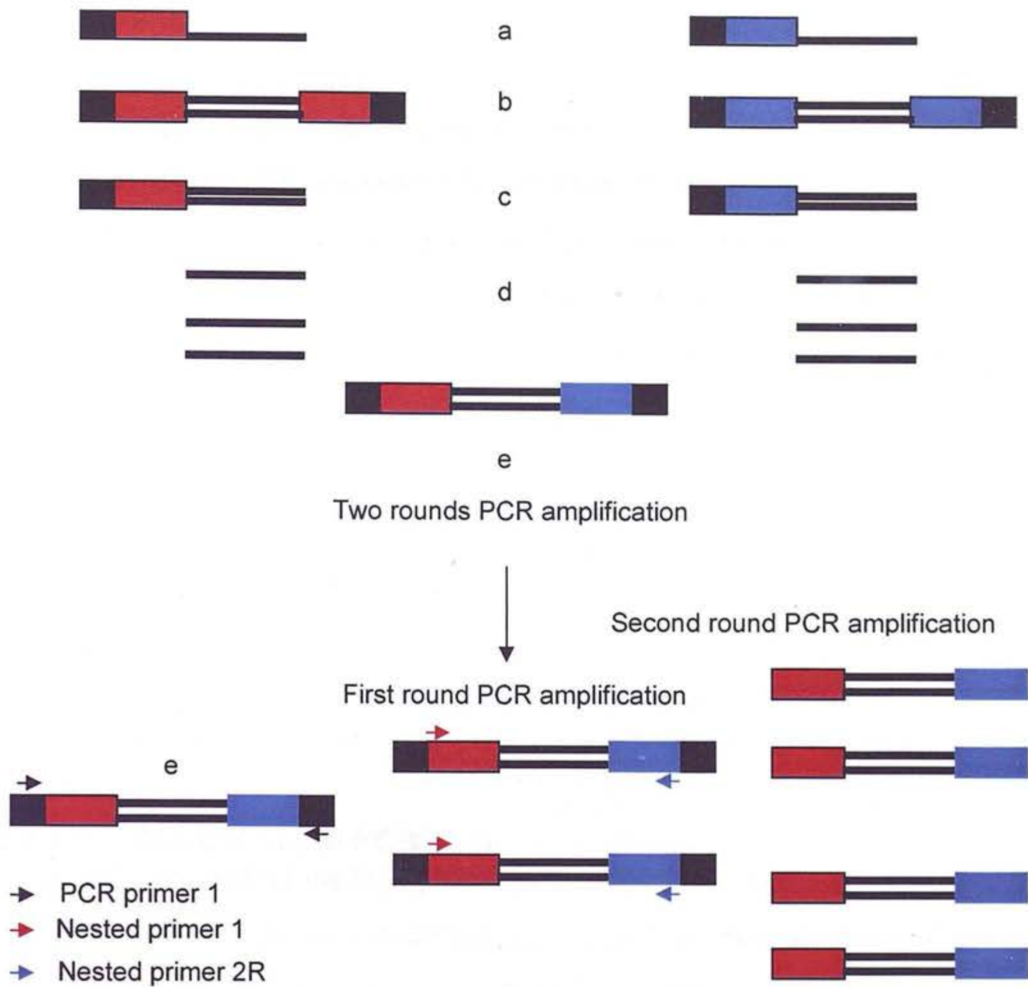
The second hybridisation round was performed immediately after. The UP sample was prepared as follows; 1 μl of 4 x hybridization buffer and 2 μl of dionized water were mixed to 1 μl of the UP dscDNA in a sterile 0.5 ml microcentrifuge tube and overlayed with one drop of mineral oil. This mixture was incubated at 98°C for 1.5 min. The total volume from both reaction obtained in the first hybridization was then added to the freshly denatured UP sample mixture. The samples were mixed together by pipetting up and down and incubated at 68°C overnight in the thermocyler. 200 μl of dilution buffer was added to the mixture, which will be referred to as subtracted treated procyclic form (STP) and heated at 68°C for 7 min and stored at -20°C.

4.2.3.8 PCR Amplification

4.2.3.8.1 Introduction

PCR reactions were carried out, on the one hand to selectively enrich for type e hybrids which represents differentially expressed transcripts in the TP sample, on the other hand to suppress the amplification of non-differentially expressed targets such as type b c and d molecules. PCR suppression of hybrids b happens despite the fact that these hybrids present sequence adaptors at both ends. This suppression of amplification is due to the presence of the same sequence adaptors. Therefore during the PCR annealing step, pan-cake like structures occurs due to the long identical sequence at both ends.
Hybrids c and d would not be amplified because they either lack both adaptors sequences (type d molecule hybrids) or present just one adaptor sequence (type c molecule hybrids) and would amplify linearly. Only Type e molecule hybrids would amplify exponentially because, firstly they bear both adaptors at both ends and secondly these adaptors represent two different stretches of sequences avoiding thus pan-like structure occurring (Figure 4.3). Prior to PCR amplification, the missing strands of the adaptors were filled by a brief incubation, two PCR rounds were then performed firstly

using PCR primer 1 to selectively amplify for type e molecule hybrids and secondly with the nested primer to further enrich for type e hybrids.



a + d: no amplification, missing either one or both adaptors
b: pan-cake like structure, no amplification
c: linear amplification, presence of one adaptor

Figure 4.3

PCR amplification of type e hybrid molecules.

Exponential amplification of type e heterohybrids happens because of the different sequence of adaptors 1 and 2R. Type a and b molecules are missing primer annealing site and are not amplified. Type b molecules and due to suppression PCR effect, form a pan-like structure that suppresses their expenontial amplification. Type c molecules amplify linearly because they bear only one primer site.

4.2.3.8.2 Methods

4.2.3.8.2.1 First round PCR

Two reactions containing a 7 µl of either the UTP or STP sample obtained above, were added to 2.5 µl of 10 x PCR reaction buffer, 0.5µl dNTP Mix (10 mM), 1 µl PCR Primer 1 (10 µM), 0.5 µl of 50 x Advantage cDNA Polymerase mix and 19.5 µl of dionized water. The reactions were mixed well and transferred to the thermocycler. Prior to PCR amplification, the sample were incubated for 5 min at 75°C to extend the adaptors 1 and 2R. PCR parameters were as described in Table 4.2.

Table 4. 2 : PCR parameters used for the selective amplification of hybrids e

Step	Time (seconds)	Temperature	
Denaturing	30	94°C	
Annealing	30	66°C	x 27
Extension	90	72°C	

4.2.3.8.2.2 Second round PCR

A 7 µl aliquot from each of the PCR products obtained above were each diluted in 23 µl of dionized water in 0.5 ml microcentrifuge tube. 7 µl from each dilution was combined to 2.5 µl 10 x PCR reaction buffer, 1 µl Nested PCR primer 1 (10µM), 1 µl of Nested PCR primer 2R (10 µM), 0.5 µl dNTPmix (10µM), 0.5 of 50 x Advantage cDNA Polymerase mix and 12.5µl dionized water. PCR parameters were as described in Table 4.3.

Table 4. 3: PCR amplification parameters used for hybrids e enrichment

Step	Time (seconds)	Temperature	
Denaturing	30	94°C	
Annealing	30	68°C	x 15
Extension	150	72°C	

4.2.3.9 Semi-quantitative PCR analysis of subtraction efficiency

In order to compare the abundance of the known house keeping gene (GAPDH) in the STP and UTP second round PCR products. 1 μl of the STP second round PCR products or 1 μl of UTP second round PCR products were diluted in 10 μl dionized water. In two 0.5 ml micro-centrifuge tubes, 1 μl of these dilutions was mixed with 1.2 μl (10 μM) forward GAPDH primer, 1.2 μl (10 μM) reverse GAPDH primer, 3 μl of 10 X PCR reaction buffer, 0.6 dNTP mix (10 mM), 50 x Advantage cDNA polymerase mix and 22.4 μl of dionized water. PCR parameters are shown in Table 4.4.

Table 4. 4

PCR amplification used during the semi-quantifcation analysis of the substraction efficiency

Step	Time (seconds)	Temperature	
Denaturing	30	94°C	x 33
Annealing	30	55°C	
Extension	120	68°C	

4.2.3.10 Construction of the subtractive library

PCR products obtained in section 4.2.3.8 where used in ligation reaction as described in section 2.7.2. Ligation mixture was used in the transformation reaction of *E.coli* competent cells as described in section 2.7.1.

4.2.3.11 PCR screening of recombinant clones

4.2.3.11.1 Introduction

Recombinants bacterial colonies obtained from the subtraction library where checked for the presence or absence of DNA inserts using nested primers 1 and 2R. Nested PCR primers were used for the identification of subtracted recombinant clones since genuine recombinant clones would have adaptor 1 and 2 R sequence bearing the sequences of the nested primer 1 and 2R respectively on each extremity (type e molecule hybrids).

4.2.3.11.2 Method

A clean toothpick was used to transfer a recombinant white colony to a 0.5 ml microcentrifuge tube containing 2.5 μl 10 x PCR reaction buffer, 1 μl Nested PCR primer 1 (10μM), 1 μl of Nested PCR primer 2R (10μM), 0.5 μl dNTPmix (10μM), 0.5 of 50X Advantage cDNA Polymerase mix and 19.5 μl dionized water. PCR parameters are shown in Table 4.5

Table 4. 5: PCR parameters used for bacterial screening

Step	Time (seconds)	Temperature	
Denaturing	30	94°C	x 25
Annealing	30	55°C	
Extension	120	68°C	

4.2.11.3 Real-time reverse-transcriptase Polymerase Chain Reaction [Real-time RT-PCR]

LightCycler® 1.5 Instrument was used (Roche). Data analysis was carried out using the LightCycler software (Version 3).

Method: For a single reaction, 1 μl of ss cDNA obtained as described in section 4.2.3.1 was mixed with 12.5 μl QuantiTect SYBR Green PCR Master mix, with primers at a concentration of 0.4 μM made up to a final volume of 25 μl with dionized water.

The reaction mixtures were vortexed and transferred to LightCycler® sample capillary tubes (Roche). The tubes were placed in the LightCycler® centrifuge adapters and briefly spun at room temperature to help the reaction mix to settle at the tip of the capillary tube. The tubes were placed in the LightCycler® sample carousel in the LightCycler® System. Fluorescence data acquisition for melting curve and quantification analysis were carried out as follow. Continuous fluorescence data collection at temperature rate transition of 0.1°C/s from 55°C up to 95°C and a single data collection after the extension step. The cycling parameters were as follow.

Table 4. 6

PCR amplification parameters used for Real-time RT-PCR

Step	Time	Temperature	
Initiation	900	95°C	
Denaturing	15	94°C	
Annealing	30	55°C	x 35
Extension	30	72°C	
Cooling	∞	20°C	

4.3 Results

4.3.1 Total and poly A$^+$ RNA extracted *T b brucei* treated procyclic forms with 8-Br-cGMP

Mid-log phase *T. b. brucei* procyclic forms were treated (TP) with 100 μM of 8 Br-cGMP and total and poly A$^+$ RNA were extracted from the cells. Prior to cDNA synthesis for gene expression profiling, checking the quality of the total and poly A$^+$ RNA was a requisite. Two micrograms of TP total and poly A$^+$ RNA were resolved by electrophoresis. Total RNA appeared as three individual ribosomal bands between 1.3 to 2.6 kb (Figure 4.4 lane 1). Poly A$^+$ RNA distribution appeared as a smear running from 0.5 to 6.5 kb with faint bands corresponding to remaining ribosomal RNA products (Figure 4.4 lane 2).

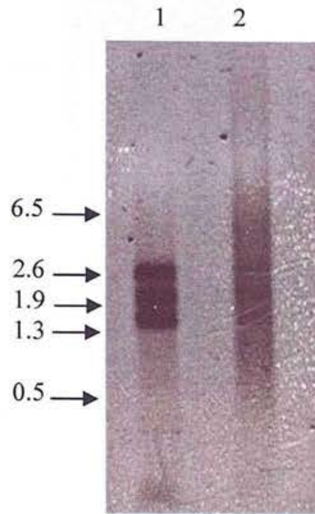


Figure 4.4

Gel electrophoresis of Total and poly A^+ RNA obtained from 8 Br-cGMP treated procyclic form (TP).

0.5 µg of RNA obtained from mid-log phase procyclic cells was electrophorised on a 2 % (v/w) agarose using a 10 mM sodium phosphate running buffer. Lane 1 represents procyclic total RNA and lane 2 represents poly A^+RNA extracted fom total RNA sample. Faint ribosomal bands were still visible in the poly A^+ RNA sample (lane 2). Numbers on the left represents RNA

4.3.2 cDNA synthesis and *RSA* I digestion

The cDNA was synthesised using 2 µg of TP Poly A^+ RNA (Figure 4.5 lane 1). Agarose gel electrophoresis analysis of the cDNA of TP sample showed a smear ranging from 0.5 kb to over 1.5 kb, with singular bands products representing either abundant mRNA products or ribosomal RNA (Figure 4.5 lane 1). Digestion of the TP cDNA resulted in a drop of the average size range, between 0.3 and 1.5 kb (Figure 4.5 lane 2) when compared to the undigested TP cDNA (Figure 4.5 lane 1). This showed that *RSA* I digestion step had occurred.

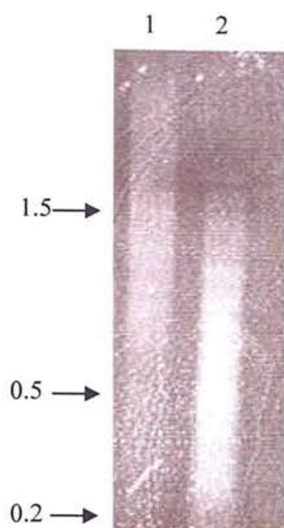


Figure 4.5

Gel electrophoresis of TP cDNA obtained from 2µg of TP poly A^{+}RNA

5 µl of each undigested (lane 1) and *RSA* I digested (lane 2) TP cDNA was electrophoresed on a 2% (w/v) agarose gel. Digested TP cDNA sample showed a drop in product size (lane 2) compared to undigested TP cDNA product sample (lane 1). Singular band product are shown more clearly in lane 2 than lane 1.the drop of size and appearance of band product shows that *Rsa* I digestion had occurred.

4.3.3 Ligation Efficiency

The *Rsa* I digested dscDNA TP was ligated to adaptors 1 and 2R. To assess whether the adaptors 1 and 2R had ligated to the GAPDH transcripts, PCR analyses were carried out as described in section 4.2.3.6 using specific GAPDH and adaptor 1 and 2R primers.
DNA amplification using GPADH specific primers generated a ~ 200bp PCR product that could be seen on Figure 4.6 panel B 1. DNA amplification of the GAPDH transcript using specific reverse primers and Adaptor 1 primer generated a ~300 bp PCR product (Figure 4.6 panel B 2). The ratio of amplification between the two PCR products (Figure 4.6 panel B 1 & 2) was identical which indicated that the ligation of Adaptor 1 was successful on all GAPDH transcripts present in the sample.

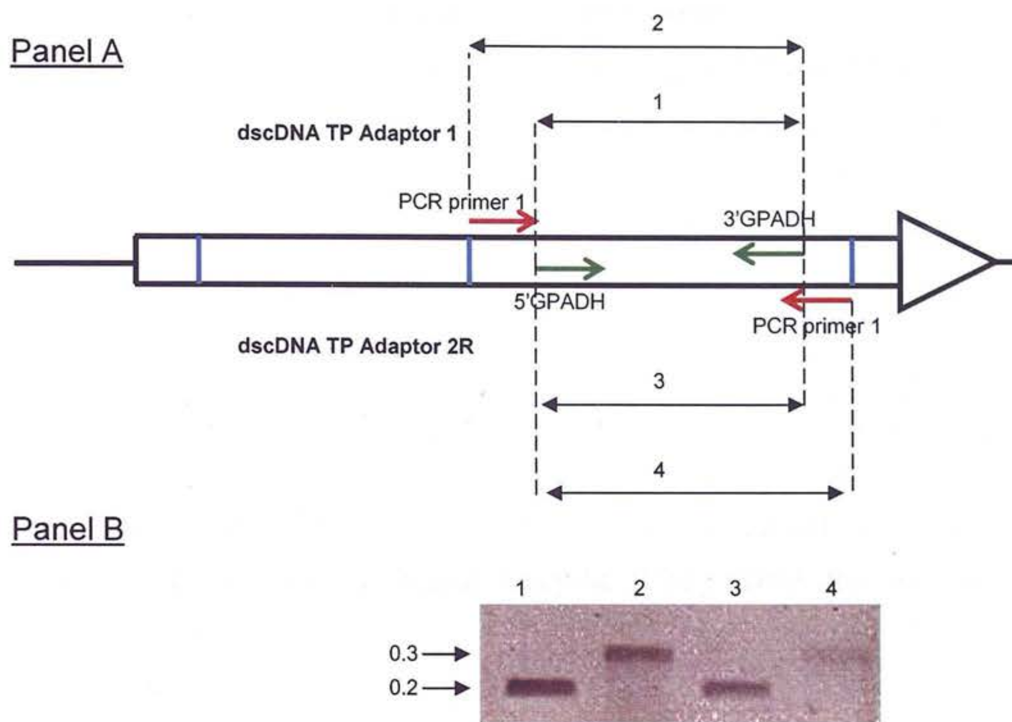


Figure 4.6

ligation efficiency of Adaptor1 and 2R on *RSA* I digested dscDNA TP.

Panel A: graphical describing the designed experiment to asses the ligation of Adaptor 1 and 2R using a combination PCR primer 1 and GAPH specific primers. Vertical blue line represents *RSA I* restriction site on *T.brucei* GAPDH gene sequence. 1 and 2 represent expected PCR products using GAPDH specific primer combination and PCR primer 1- 3'GAPDH combination respectively on *T.brucei* GAPDH. 3 and 4 represent expected PCR products using GAPDH specific primer combination and 5'GPADH - PCR primer 1 combination respectively on T.brucei GAPDH

Panel B: Gel electrophoresis of PCR product 1 2 3 and 4. The ratio of amplification between PCR product 1 and 2 was acceptable however the ligation of adaptor 2 R on dscDNA TP was repeated because adaptor 2 R was not present on all GAPDH transcripts.

Similarly, when DNA amplification was performed using the GAPDH specific primers and the PCR primer 2R with the GAPDH forward primer, products were observed at ~200 bp (Figure 4.6 panel B 3) and at ~300 bp (Figure 4.6 panel B 4) respectively. Although the ligation of adaptor 2R occurred (Figure 4.6 panel B 3), the efficiency has to be proven since the amplification ratio did not meet the requirements (Figure 4.6 panel B 3 & 4). The ligation of adaptor 2R was repeated and after assessment it met the required amplification ratio.

4.3.4 Selective DNA amplification of differentially expressed transcripts in the Subtracted Treated Procyclic (STP) cDNA.

Following the second hybridisation step, a DNA amplification through two-step PCR, of differentially expressed transcripts of the STP cDNA was carried out, as well as PCR amplification of Unsubtracted Treated Procyclic (UTP) cDNA that did not undergo hybridisation rounds.

DNA amplification of STP cDNA appeared as a smear containing distinctive bands running from 200 up to 700 bp (Figure 4.7. lane 1). DNA amplification of the UTP cDNA appeared as a smear running from 200 to above 1500 bp (Figure 4.7 lane 2).
The banding pattern obtained with the STP sample (Figure 4.7 lane 1) was significantly different from the banding pattern obtained with the UTP sample (Figure 4.7 lane 2) which indicated that the hybridisation reaction has taken place in the STP sample.

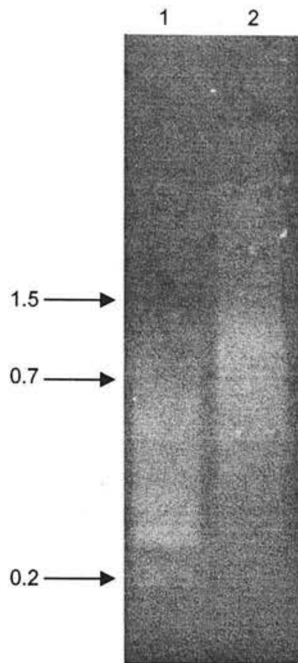


Figure 4.7

Gel electrophoresis of 2 round PCR products obtained from subtracted treated procyclic from (STP) and unsubtracted treated procyclic form samples.

Lane 1 represents amplification of STP cDNA and lane 2 represents amplification of UTP cDNA. Differences in banding patterns displayed between the two samples suggest that hybridization had successfully taken place in the STP sample.

4.3.5 Semi-quantitative PCR analysis of subtraction efficiency

To test the efficiency of the subtraction in STP cDNA, DNA amplification of GAPDH transcript was measured in both subtracted and unsubtracted TP cDNA.

20 ng of each cDNA was amplified using GAPDH specific primers. An aliquot of 5μl of the resulting product was collected after 18, 23, 28 and 33 PCR cycles and analysed through agarose gel electrophoresis.

In the STP, a 200 bp PCR product weakly appeared after 18 cycles (Figure 4.8 Subtracted, cycle18) whereas at the same number of cycles, this product appeared strongly in the UTP sample (Figure 4.8, unsubtracted, cycle 18).

From cycles 23 to 32, the intensity of the amplified GAPDH transcript in the STP cDNA was less (Figure 4.8, subtracted, cycles 23, 28 and 32) than the intensity of amplified GAPDH transcript in the UTP cDNA (Figure 4.8, unsubtracted, cycles 23, 28 and 32). This demonstrated the reduction in the abundant common genes such as GAPDH transcript in the STP cDNA and shows that the subtraction was successful.

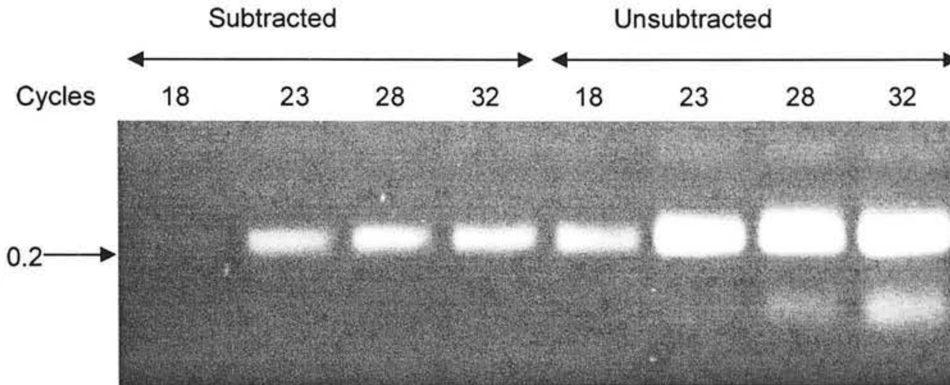


Figure 4.8

Semi-quantitative PCR analysis of hybridization efficiency of subtracted treated procyclic sample.

Abundance of GAPDH transcript in STP and UTP cDNA samples were assessed after 18, 23, 28 and 32 PCR cycles on a 2 % agarose gel electrophoresis. The reduction of GAPDH transcript in STP cDNA sample is evident when compared to the UTP cDNA sample and proves that the subtraction of such abundant transcripts had taken place.

4.3.6 Construction of the STP cDNA library

The PCR amplified STP cDNA obtained in Figure 4.7 lane 1 was ligated to pGMT easy vector. The ligation reaction was then used to transform *E. coli* JM 109 competent cells and blue/white screening was used to identify 1063 recombinant (white) bacterial colonies.

4.3.7 PCR screening of recombinant clones

All of the 1063 white colonies were checked for the presence of type e hybrids using the nested PCR primers. The size of the resulting PCR products, varied between 200 bp and 0.7 kp (Figure 4.9), corresponded to the size range described in the STP cDNA smear (Figure 4.7 lane 1). White colonies bearing double clones were discarded (Figure 4.9 top arrow). Identical insert sizes were encountered (Figure 4.9, down arrows). These identical inserts either corresponded to the same product being present in more than one copy or more commonly to different dscDNAs of the same size, therefore both type were considered for sequencing.

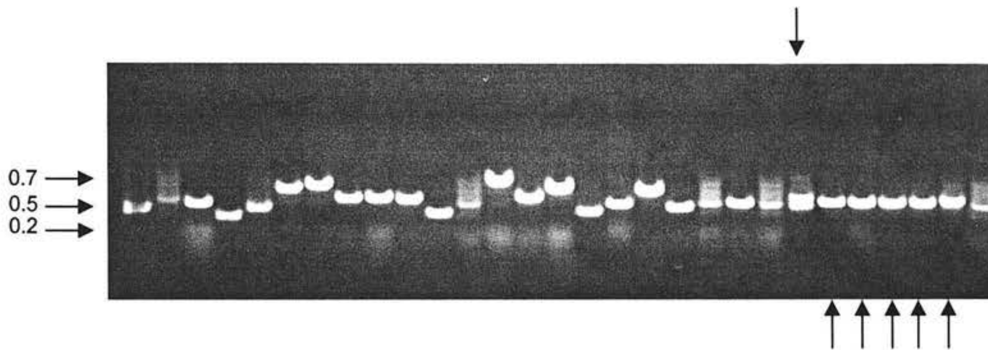


Figure 4.9

PCR screening of 29 recombinant bacterial colonies using nested primer 1 and 2R

PCR products were resolved on a 2 % agarose gel. Top arrow corresponds to double inserts clones and bottom arrows correspond to identical insert representing either the same product insert product or different PCR product of the same size

4.3.8 Sequence analysis of differentially expressed cDNAs from 8 Br-cGMP treated *T. b. brucei*

A total of 1063 recombinant colonies were selected and grown overnight. Bacterial cultures were prepared as glycerol stock and sent to Lark Technology (Essex, UK) for plasmid purification and cDNA sequencing. From 1063 recombinant colonies, 23 failed to grow. A total of 1040 recombinant colonies generated a sequence.

The sequence data analysis and consensus constructs were carried out as described in section 2.8. In total 259 contigs were formed using *seqMan* (DNAstar). The DNA sequences of all constructed contigs can be found in Appendix II.

4.3.8.1 Blast search homology

Blast search homology searchs were carried out as described in section 2.8. As a first step contigs sequences were blasted against the *T. brucei* Ominiblast genome as a stepping stone where more detailed information could be obtained on the query sequence as several algorithm searchs are combined at one glance. Among the 259 contigs constructed [Appendix II], 166 contigs presented significant similarities to annotated genes on the *T. brucei* database and blast search results are resumed in Table 4.7 and Appendix III. Among the 208 constructed contigs sequences 22 did not reveal any significant similarities to annotated genes and where classified as unknown genes. 71 contig sequences were below sequence thresholds and were not considered for blast similarity search.

Contigs shown in Table 4.7 presented strong homology to enzymes involved in polyamines and redox cellular pathway and were selected because they related to Macleod findings where he reported that reactive oxygen species could be responsible for the clearance of parasites in the tsetse midgut and that cGMP could play a role in the survival of the parasites during their establishment in the tsetse midgut.

Name	Blast result	Species	E-value	Size (bp)
Contig 43	Tryparedoxin peroxidase	*T.brucei*	$1.3\ e^{-40}$	278
Contig 86	S-adenosylhomocysteine hydrolase	*T.brucei*	$2.0e^{-28}$	283
Contig 56	Arginine kinase	*T.brucei*	$3.2\ e^{-46}$	307
Contig 57	Selenophosphate synthetase	*T.brucei*	$2.5e^{-85}$	529
Contig 155	Arginase	*T.brucei*	4.9e-73	432
Contig 120	Tryparedoxin	*T.brucei*	$1.2e^{-23}$	247
Contig 188	S-adenosylmethionine synthetase	*T.brucei*	$4.0e^{-69}$	405
Contig 110	unknown Similar to Alpha crystallin B chain	*Homo sapiens* *Pan troglodytes*	$5.e^{-31}$ $5.e^{-21}$	357

Table 4.7

The contig sequences were compared with database entries using the blast x algorithm against the non-redundant database NCBI (http://www.ncbi.nlm.nih.gov/BLAST/) (non-redundant protein translation of GenBank,; Protein Data Bank; SwissProt; Protein Identification Resource Database; Protein Research Foundation Sequence database), and Trypanosoma OmniBLAST Server(http://www.sanger.ac.uk/cgi-bin/blast/submitblast/t_brucei/omni). The higher Score and the lowest Expect values constituted the best match. The length represents the size of the clone transcript after sequencing Contig 110 did not reveal any significant homology on *T.brucei* database however it showed strong homology to *Homosapien* alpha crystalline B chain on NCBI database

Contig 43 was formed from two sequence clones and had a consensus sequence of 278 bp. Blast.x search on omniblast *T.brucei* database revealed high homology with Tryparodoxine peroxidase (TXNPx) protein sequence (Score = 424 , Expect = $1.3e^{-40}$). The protein sequence was classified among AhpC/TSA family (Accession number: PF00578). Peroxiredoxins (Prxs) are regarded as a family of antioxidant enzymes. The regulations of Prxs are believed to be controlled by changes to phosphorylation, redox and possibly oligomerisation states. Prxs are divided into three classes: typical 2-Cys Prxs; atypical 2-Cys Prxs; and 1-Cys Prxs. All Prxs share the same basic catalytic mechanism, in which an active-site cysteine (the peroxidatic cysteine) is oxidized to a sulphenic acid by the peroxide substrate. Tryparodoxine peroxidase plays a role in the defense against oxidant damage in the Trypanosomatids (Ariyanayagam and Fairlamb 2001) by decomposing peroxides using tryparedoxin (TXN) as an electron donor (Trujillo *et al.,* 2004).

Contig 120 sequence was formed from two sequence clones, which consensus revealed a 247 bp sequence size. Homology search identified a strong hit to *T. brucei* tryparedoxin (Score = 264, Expect = $1.2e^{-23}$). Tryparodoxin is exclusively present in protozoan parasites and acts as a thiol-disulfide oxidoreductase in *T. brucei* (Luëdemann *et al.,* 1998).

Contig 57 was formed from two clones and presented a size of 529 bp. Blast x revealed strong homology with a *T.brucei* putative selenophosphate synthetase
(Score = 846, Expect = $2.5e^{-85}$). The predicted protein sequence belonged to AIR synthase related protein, C-terminal domain. it is believed this family of proteins include Hydrogen expression/formation protein HypE HYPE_ECOLI, AIR synthases PUR5_ECOLI EC:6.3.3.1, FGAM synthase PURL_LACCA EC:6.3.5.3 and selenide water dikinase SELD_ECOLI EC:2.7.9.3.It is believed that AIR synthase related proteins contain a sulphate binding site. Selenophosphate synthetase produces selenophosphate from selenide and ATP. Selenophophate is the selenium donor in the synthesis of selenocysteine which are incorporated into selenoproteins. One of the

major roles of selenoproteins is to act as antioxidant or to catalyze oxidation and reduction reactions (Rayman 2000; Stadtman 1996).

Contig 86 was formed from four sequence clones which formed a 283 bp clone. Blast X search revealed a strong homology to a putative S-adenosylhomocysteine hydrolase (SAHH) (Score = 309, Expect = $2.0e^{-28}$). AdoHcyase catalyses the reversible hydration of S-adenosyl-L-homocysteine (SAH) into adenosine and homocysteine in the prescence of NAD^+ as cofactor (Kathleen *et al.,* 1994).

Contig 188 was formed formed of two sequence clones which formed a 405 bp contig. Blast x results revealed a strong homology to S-adenosylmethionine synthetase (Score = 693, Expect = $4.0e^{-69}$). The enzyme catalyses the formation S-adenosylmethionine (AdoMet) from methionine and ATP. AdoMet is a methyl donor for transmethylation and is also the propylamino donor in polyamine biosynthesis (spermide and spermine) in trypanosomatids (Reguera *et al.,* 1999).

Contig 155 was formed of sequence clones which consencus formed a 432 bp contig sequence. Homology search on *T. brucei* database revealed a strong hit with putative arginase enzyme (Score = 730, Expect = $4.9e^{-73}$). Arginase catalyses the conversion of L-arginine into L-ornithine and urea. L-ornithine encourages parasite growth and is a precursor of polyamines synthesis (Vincendeau *et al.,* 2003).

Contig 56 was formed from four sequence clones which formed a 307 bp cDNA clone. Balst x revealed a strong homology to *T.brucei* Arginine kinase (Score = 477, Expect = $3.2e^{-46}$). The corresponding protein sequence belonged to the ATP:guanido phosphotransferase (Accession number: PF02807). These proteins reversibly transfer phosphate from ATP to various phopahagenes such as phosphoarginine and phophocreatine which play an important role as energy reserve (Pereira *et al.,* 2003). In *T. cruzi*, a correlation it been established between arginine kinase activity, arginine uptake and parasite stage and cell growth (Pereira *et al.,* 2002). Pereira (2003), proposed

that arginine kinase is a constituent of an adaptive response to nutritional stress conditions

Contig 110 a 357 bp size transcripts revealed no similarity to any *T. brucei* database using either Blast X and / or Blast N programmes. No homology was detected on kinetoplastids database (website). However, the contig showed firstly a strong homology to an unknown *Homo sapiens* protein (Score = 134, Expect = 5.0 e^{-31}) and secondly a strong hit to Alpha crystalline B chain of *Pan troglodytes* when blasted on NCBI. Crytallin proteins are abundant in eyes lens, alpha and beta crytallin are found in numerous tissues outside the lens (Bhat *et al.,* 1989) with alpha chain crytallin behaving in a lot of aspects like a small heat shock protein (*shsp*) (Wilfied *et al.,* 1993).Shsp are generally more expressed in plants (Veirling *et al.*, 1991) and are among the most diverse structurally and functionally in the super family of stress proteins (Franck *et al.*, 2004). A crystallin alpha domain was described in parasites such as the major egg antigen of the blood parasite *Schistosoma mansoni* shares a common sequence block with alpha-crystallins (Nene *et al.,* 1986). In bacteria an 18 kDa immunodominant antigen of *mycobacterium leprae* belonged to the alpha-crystallin protein family (Nerland *et al.,* 1988). Over-expression of alpha B-crystallin is associated with several neurodegenerative disorders such as Alexander's and Alzheimer's diseases, dementia and scrapie (Bhat *et al.,* 2003, Sax *et al.*, 1994).

4.3.9 Relative Quantification of arginase transcript during treatment of *T.b.brucei* with 8-br-cGMP using real-time reverse transcriptase PCR (real time RT-PCR):

Expression quantification of Arginase gene transcript in the treated procyclic form sample was undertaken using QuantiTect™ SYBR® Green PCR Kit (Qiagen) for real time RT-PCR as described in section 4.2.4.

Real-time reverse transcriptase quantification is a sensitive technique which is preferable to other form of quantitative RT-PCR (Freeman, 1999; Raeymakers, 2000). This

technique relies on the detection of a fluorescence reporter in the PCR mix in real time (Lee *et al.,* 1993; Livak *et al.,* 1995). SYBR Green I fluorescence dye binds to double stranded DNA for the detection of amplified PCR product in real time (Morrison *et al.,* 1998). Fluorescence signal increase proportionally during the exponential amplification of the PCR product and the higher the initial copy number the higher the fluorescence. A fixed fluorescence threshold is set and an increase of fluorescence above the baseline is defined as threshold cycle (C_T) which represents the cycle number at which the fluorescence emission surpasses the fixed threshold (Pfaffl 2001).

The quantification analysis of the expression of the arginase transcripts in the treated and untreated procyclic form sample were performed using the Relative quantification methods as described by Pfaffl (2001) using the C_T of a target gene (Arginase), versus a C_T of the reference gene (GAPDH). The reference gene is a stable and unregulated gene (House keeping gene). House keeping gene expression is considered stable under experimental conditions (Marten *et al.,* 1994; Thellin *et al.*, 1999). Normalisation of C_T of target gene (Ariginase) to C_T of the house keeping gene (GAPDH) is used to correct for the difference in the amount of total RNA added to the reactions and for the differences in RT efficiencies. The normalisation also compensate for PCR inhibitor in the samples (Benoy *et al.*, 2004). Comparison of the C_T between the arginase transcripts and GAPDH transcript can only be undertaken if the efficiency of amplification of Arginase gene transcripts is more or less identical to the PCR efficiency of GAPDH. The efficiency (Eff) of the reactions can be calculated by the arithmetic formula: $E=10^{(-1/slope)}$. So that the slope of the log-linear line represents the efficiency of the PCR amplification (Rasmussen 2001). PCR reaction efficiency should be 90-100 % ($E \geq 1.80$). The concept of relative quantification used a mathematical model which calculates "the ratio (R) of Ariginase gene expression, based upon the efficiency and CT deviation in the treated versus the non-treated sample and expressed in comparison to the expression of GAPDH gene expression in the treated and the control sample:

$$\text{Ratio} = (E_{\text{target}})^{\Lambda CT_{\text{Arginase}}^{(\text{non-TPF} - \text{TPF})}} / (E_{\text{GAPDH}})^{\Lambda CT_{\text{GAPDH}}^{(\text{non-TPF-TPF})}}$$

Where, E_{Arginase}: Efficiency of the PCR amplification of Arginase transcript

E_{GAPDH}: Efficiency of the PCR amplification of GAPDH transcript

Λ CP $_{Arginase}$: Deviation of CT of Arginase between non-TPF and TPF

Λ CP $_{GAPDH}$: Deviation of CT of GAPDH between non-TPF and TPF

4.3.9.1 Confirmation of the purity of Arginase and GAPDH PCR products:

The specificity of the RT-PCR products was confirmed using melting curve analysis. This represents the temperature at which 50 % of the DNA has become single stranded. T_m was determined as described in section 4.2.4.

Arginase and GAPDH amplicons presented a T_m of 78°C and 85°C; respectively. Each amplification was represented by a single product (Figure 4.10). Un-specific primer-dimers were not produced during the amplification steps since only one peak was detected in each reaction (Figure 4.10).

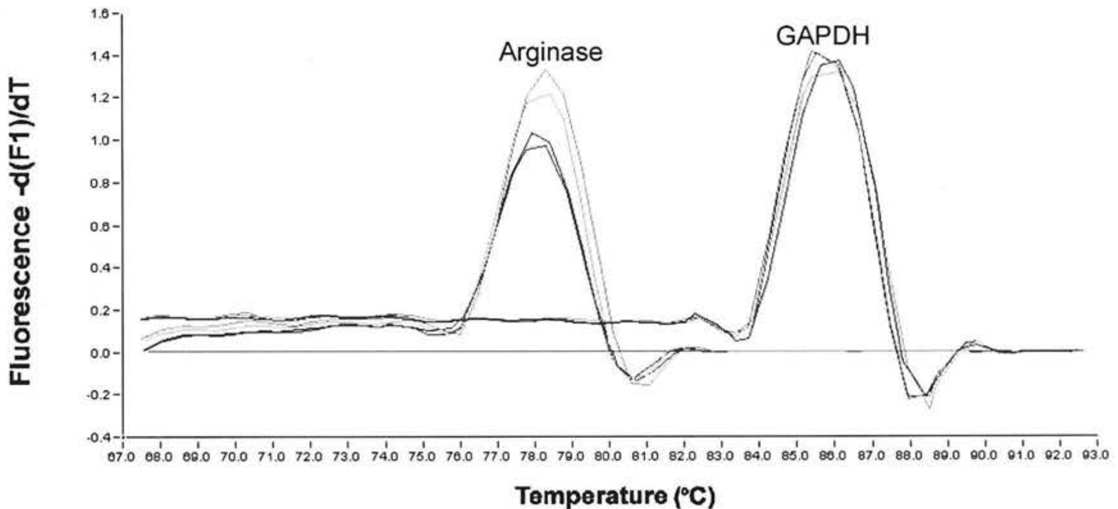


Figure 4.10

Derivative melting curve plots of Arginase and GAPDH amplicons.

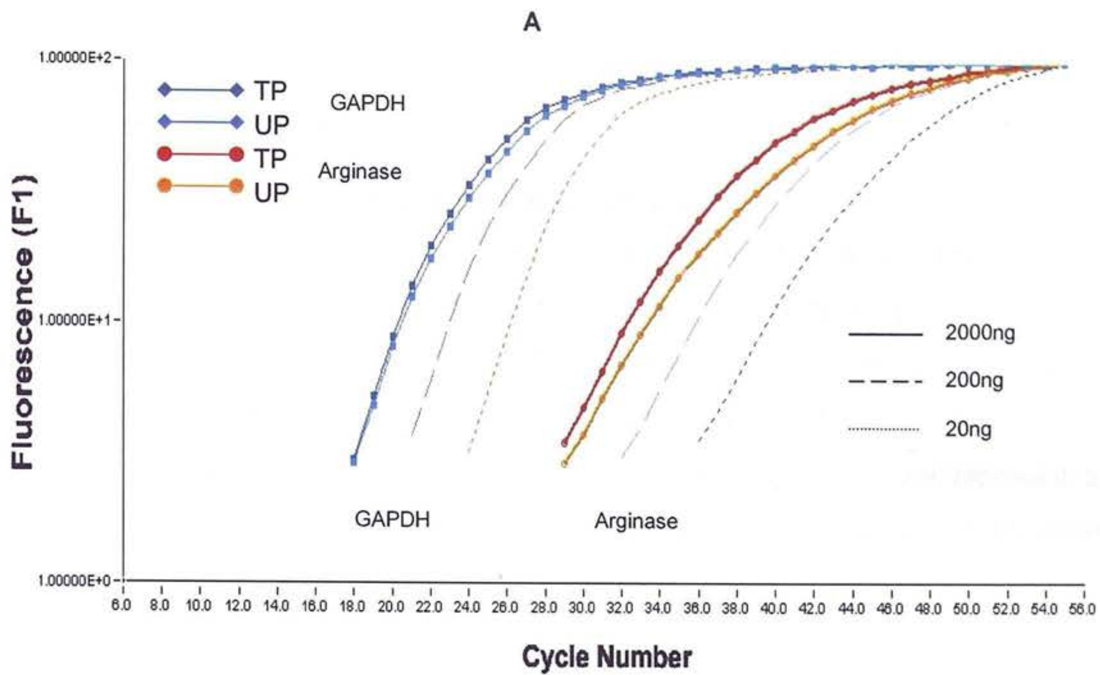
The peaks indicate the melting point of arginase (78°C) and GAPDH (85°C) amplicons. The data were obtained by monitoring the fluorescence of SYBRGREEN during heating from 55°C to 95°C at a temperature transition rate of 0.1°C/s. x axis represents temperature in Celsius degree and y axis represents the derivative of the fluorescence

4.3.9.2 Data analysis

A duplicate of serial dilutions of 2000, 200, 20 ng of *T.b. brucei* ss cDNA obtained from 2 μg of poly A mRNA from TPF and non-TPF sample were used to amplify for arginase and GAPDH gene transcripts as described in section 4.3.9. The PCR reaction efficiency in the arginase reaction was calculated following mathematical formula described in section 4.3.9 and was equal to 2.04. The PCR efficiency obtained with GAPDH PCR reaction was equal to 2.13. The C_T and the slope of each reaction are shown in Table 4.8. The expression of arginase transcripts in the treated vs. the untreated sample was, normalized to the expression of GAPDH transcripts and calculated using the mathematical formula described in section 4.2.4. Arginase was found to be up-regulated by a factor of 3.15 in the treated sample compared to the untreated sample.

Table 4. 7: The C_T obtained from TP and UP of arginase and GAPDH amplicons

	GAPDH		Arginase	
	UP	TP	UP	TP
CT	19.07	18.56	31.9	30.1
	18.73	18.88	32.22	30.42
	$R = 2.04^{(32.06-30.26)} / 2.13^{(18.9-18.72)} = 3.15$			



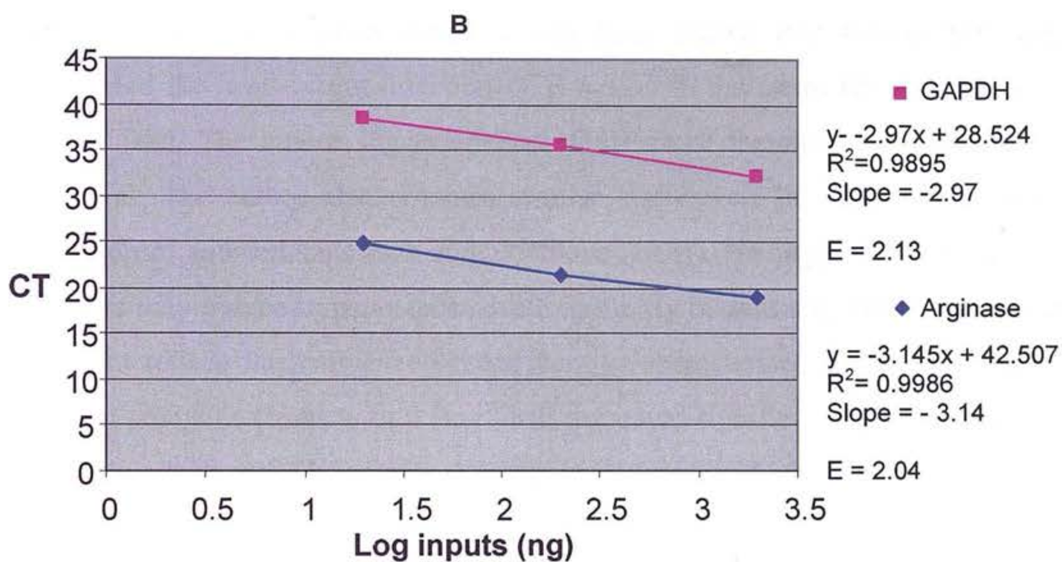


Figure 4.11

Calculation of real-time efficiency rates of arginase and GAPDH amplification.
PCR amplifications were undertaken on serial dilution (2000, 200, 20 ng) of cDNA prepared from poly A^+ obtained from TP and UP samples (A). The efficiency rates of arginase and GAPDH PCR reaction were calculated from the slope of the standards curve where the CT values of each amplicon were obtained on duplicates on each dilution (B). The relative expression of arginase transcript in the TP sample showed an increase when compared in UP sample (A and B). Graph A x axis represents C_T value and y axis represents fluorescence value.

On Graph B the x axix represents logarithm of cDNA inputs and y axis represents C_T corresponding to each dilution.

4.4 Discussion

During the course of this study we isolated and characterised differentially expressed transcripts from the response of *T.b .brucei* But 135 to 100μM of 8-Br-cGMP, an analogue of cGMP. This was achieved using the Clontech PCR-Select™ cDNA Subtraction Kit to create a subtractive cDNA library. 1063 recombinant clones were isolated and sequenced from this library. Of the 256 contigs determined, 166 presented significant homology on *T .brucei* database, while 71 contigs did not represent any significant homology and where classified as unknown. The remaining 19 contigs represented poor sequence data and were discarded.

This study work was developed from previous resultes obtained by the Macleod (Macleod *et al.*, in press 2005). It has been shown that 8-Br-cGMP significantly increased the level of gut infection of *T. brucei* in the tsetse fly vector when added to blood feed. The author suggested that cGMP must therefore be influencing parasites survival. The author also obtained similar high levels of midgut infection upon the addition of anti-oxidants such as glutathione (GSH). He suggested that reactive oxygen species may induce trypanosomes death in the fly midgut and that the prescence of anti oxidant reduce this influence.Several contigs characterised during the treatment of *T. brucei* procyclic forms with 8 Br-cGMP presented significant homology to a number of enzyme which could be playing a role as, or be involved, in the synthesis of anti-oxidant molecules.

Contig 188 shows strong homology to s-adenosylmethionine synthetase. This enzyme catalyses the formation of the s-adenosylmethionine (AdoMet) from methionine and ATP. With AdoMet being the second most biological active compound after ATP (Cantoni *et al.,* 1975 ,Kotb and Geller, 1993; Mato *et al.*, 1997). Among several other cellular implications, AdoMet has been solely described as a methylating agent where is a donor of aminopropyl groups to diamines in the polyamine pathway (Fontecave *et al.,*

2004). *T. brucei* has been reported to rapidly transform exogenous methionine into AdoMet throught the action of a loosely regulated s-adenosylmethionine synthetase (Goldberg et al 1999)(Figure 4.12). Accumulation of AdoMet compounds will induce a state of hypermethylation and inturn the activation of protein methylase II.S-adenosylhomocystein in one of the by-products of methylases but because of the high toxicity of this by-products (Reguera et al 1999), s-adenosylhomocystein is converted into L-homocystein and adenine by s-adenosylhomocystein hydrolase (Figure 4.12) (Palmer et al 1979) we found that contig 86 had a high homology to this gene, which would suggest the involvement of this protein in response to 8-Br-cGMP treatment.. One of the outcomes of homocystein is the conversion into glutathione where in trypanosome; it is conjugated to spermedine for the formation of trypanathione a major cellular anti-oxidant.

One of the first enzymes in polyamines pathway is represented by ornithine decarboxylase, known to be one of the most regulated enzymes in eukaryotes (Satriano 2003). This enzyme metabolizes the decarboxylation of L-ornithine for the formation putrescine and is considered as a regulating enzyme in polyamines synthesis (Jackson *et al.*, 2003). Ornithine plays a major role as a component in the biosynthesis of polyamines, proline, glutamate, and c amino-butyric and is obtained by the action of arginase on arginine in mammalian cells (Cederbaum *et al.,* 2004). Contig 155 showed strong homology to *T. brucei* arginase. Relative quantification analysis of the expression of arginase in the treated sample was analyzed using the real time RT-PCR. The nomalization of arginase transcripts expression was effectuated using GAPDH, a house keeping gene. The arginase target gene was found to be up-regulating by a factor of 3.15. The relative quantification of the arginase gene in the treated simple was normalized to GAPDH gene which was considered as a reference gene to which the expression level of the arginase between the treated and control sample was normalized. We found that the efficiencies of arginase and GAPDH amplification reactions differ which is expected since the primer binding sites, PCR product sequences, and PCR product sizes are different. However the level of expression of GAPDH gene between

the TP and UP samples did not differ (Figure 4.11) which endorses our finding that the increase in level of expression of arginase in the TP sample is likely due to the treatment.

Decarboxylated AdoMet by AdoMet decarboxylase, represents the amino propyl in the synthesis of polyamines such spermedine in trypansosomatids (Reguera *et al.*, 1999). In trypansomes decarboxylated AdoMet is combined to putrescine throught a reaction catalyzed by spermedine synthetase for the formaton of polyamine spermedine where this polyamine is then conjugated to gluthatione by trypanothione synthetase to form a low molecular mass thiol found uniquely in traypansomatids called trypanothione ([*N*1,*N*8-bis(glutathionyl)spermidine] (Fairlamb *et al.*, 1985; Fairlamb *et al.*, 1985).

Trypanathione has several roles, in particular its involvement in the defense against oxidative stress and regulation of the thiol redox system (Fairlamb *et al.*, 1992). Trypansomatids posses a unique thiol-redox system, different from the eukaryotes and prokaryotes systems, as the glutathione/ glutathione reductase system is substituted by trypanothione/ trypanothione reductase (Figure 4.12). Trypansomatids relie on three distincts enzymes to relay the reduction signal from NADPH to alkyl hydroperoxide. Theses enzymes are represented by Trypanothione reductase, tryparedoxine (TXN) and tryparedoxine peroxidase (TXNPx) (Floh'e *et al.*, 1999). Contig 43 and contig 120 showed strong homology to TXN and TXNPx respectively which would suggest the involvement of these proteins in response to the 8-Br-cGMP.

Trypanosomes lack catalase activity (Huynh *et al.*, 2003) and rely instead on tryparedoxin peroxidase to detoxify hydrogen peroxide through a reduction reaction. This depends upon TXN/TR to supply the reduced thiol coupound under the form of reduced trypanathione (Ludemann *et al.*, 1998; Reckenfelderbaumer *et al.*, 2000). Trypansosomes are therfore highly sensitive to thiol depletions (Huynh *et al.*, 2003).

Little is known about the biosynthesis of polyamines precursor such as ornithine in

T. brucei. In the extracellular promastigote form of *Leishmania mexicana*, deletion of arginase genes showed that the enzyme is essential for the survival of the promastigote form and the viability of the parasites could be recovered by the addition of ornithine, spermidine or putrescine, it was concluded that arginase had a singular function as a supplying source in the biosynthesis of polyamines (Roberts *et al.*, 2004). *In vivo* and during *T. brucei* mammalian stage, parasites are highly sensitive to the presence of nitric oxid (NO) produce by the macrophages (MacMicking *et al.*, 1997; Bogdan *et al.*, 2000). Depletion of L-arginine was noticed to lower the rate of parasite clearance in the mammalian host which in turn was linked to the availability of L-arginine for the production of NO by the macrophages (Vincendeau *et al.*, 2002). During mammalian parasite stage, it has been speculated that parasites might be directly responsible for the induction of macrophage arginase in order to reduce NO cytotoxic effect. It was concluded that such ruse could be used by the parasites to evade the innate immunity of the mammalian host in the early stage of trypanosomiasis (Duleu *et al.*, 2004).

Hydrogene peroxide (H_2O_2) and NO may be involved in cells comminucation between various compartment in the tsetse and as such may be encounbtered by *T. brucei* when ingested by the fly (Lehane *et al.*, 2004). During the vector stage the arginase enzyme from *T.brucei* could play a double role in insuring the survival of parasites in the fly midgut. Firstly by depleting the environment of arginine which could be used in the formation of NO by the nitric oxide synthetase of the host fly. Secondly the transformation of arginine into ornithine by *T. brucei* arginase could be then used for polyamines and anti-oxidant synthesis by the parasites.

The arginase transcript of *T. brucei* was shown to be slightly up-regulating during the presence of the c-GMP analogues (Figure 4.11).It has been reported that arginase activity is stimulated by 8-Br cAMP, an analogue of cyclic nucleotide cAMP, in macrophages (Sidney et al 1998). Arginase is involved in regulation of numerous pathways regulation one of which proline synthesis. Proline is considered the main source of energy for these procyclic *T. brucei* forms in the fly midgut (Bursell *et al.*,

1973). In cultured procyclics proline is salvaged from the culture medium (Lamour *et al.*, 2005) and the up-regulation of arginase activity seen during 8-Br-cGMP treatment might not be related to proline consumption since procyclic are auxotroph for proline. This observation may be due to an involvement of this protein in polyamine synthesis than production of proline.

Contig 57 showed a strong homology selenophosphate syntethase. The production of selenoproteins relies upon the synthesis of monoselenophosphate. Selenophosphate synthetase catalyzes the production of the selenium biological donor, selenophosphate from selenium and ATP (Stadtman 1996). In mammals selenoenzymes plays role in the metabolism of reactive oxygen species (Jayakumar, 2004) - and similar role may occure in other organisms such as *T. brucei*

Contig 110 sequence did not reveal any significance homology on any available *T. brucei* gene bank database, however it showed a strong homology to an unknown Homo *sapiens* protein and also to alpha crystalline beta chain (α B) of *Pan Troglodytes*.. Alpha and beta crystalline are the major structural proteins of mammalian lens, they also found in other tissues that the lens where it was found that α B crystalline can function as small heat shock protein and is induced in response to heat and other physiological stress(de Jong *et al.*,1993, Klemenz *et al.,* 1991, Zantema *et al* 1992). Alpha crystalline proteins have also been described as autokinases and it was speculated that theses refractory proteins might also be involved in metabolic pathways (Kantorow *et al.,* 1994). Alpha crystalline plays a role of molecular chaperon where it binds to native protein to prevent protein aggregation during cellular stress (Andley *et al.,* 2000). Over-expression of α A and α B-crystallins protect cells from under going apoptosis (Usha *et al.*, 2002).

Even though cDNa synthesis was undertaken on poly A^+ RNA fraction of *T.brucei*, 38 contigs showed homology to various ribosomal RNA sequences on T.brucei database (Appendix III). Faint ribosomal RNA bands could be still be detectable in poly A^+ RNa sample (Figure 4.4 lane 2)

This work revealed that 8 Br-cGMP, an analogue of c-GMP cyclic nucleotide induces changes in gene expression of *T. brucei* procyclic form. Arginase, an enzyme which catalyzes the formation of ornithine, was shown to be slightly over-expressed under 8-Br-cGMP treatment. Identified contigs such as trypredoxin peroxidase and AdoMet synthetase suggested that polyamines/ thiol redox system could be involved under such treatmenr conditions. However the trypanosomes used in these experiment were grown in culture and are not exposed to an oxidative environment so that the result obtained whilst clearly demonstrating that the influence of 8-Br-cGMP switches on enzymes in the involvement of the synthesis of anti-oxidant, the question arises wether the same pathway would also be involved *in vivo*. Therefore the experiments should now focus on the level of expression of the detected proteins and also proteins such as ornithine decarboxylase and trypanathione synthetase since the latter catalyzes the formation of trypanathione, a unique parasitic anti-oxidant and the former metabolizes the formation of putrescine which is utilize for the formation of the trypanathione.

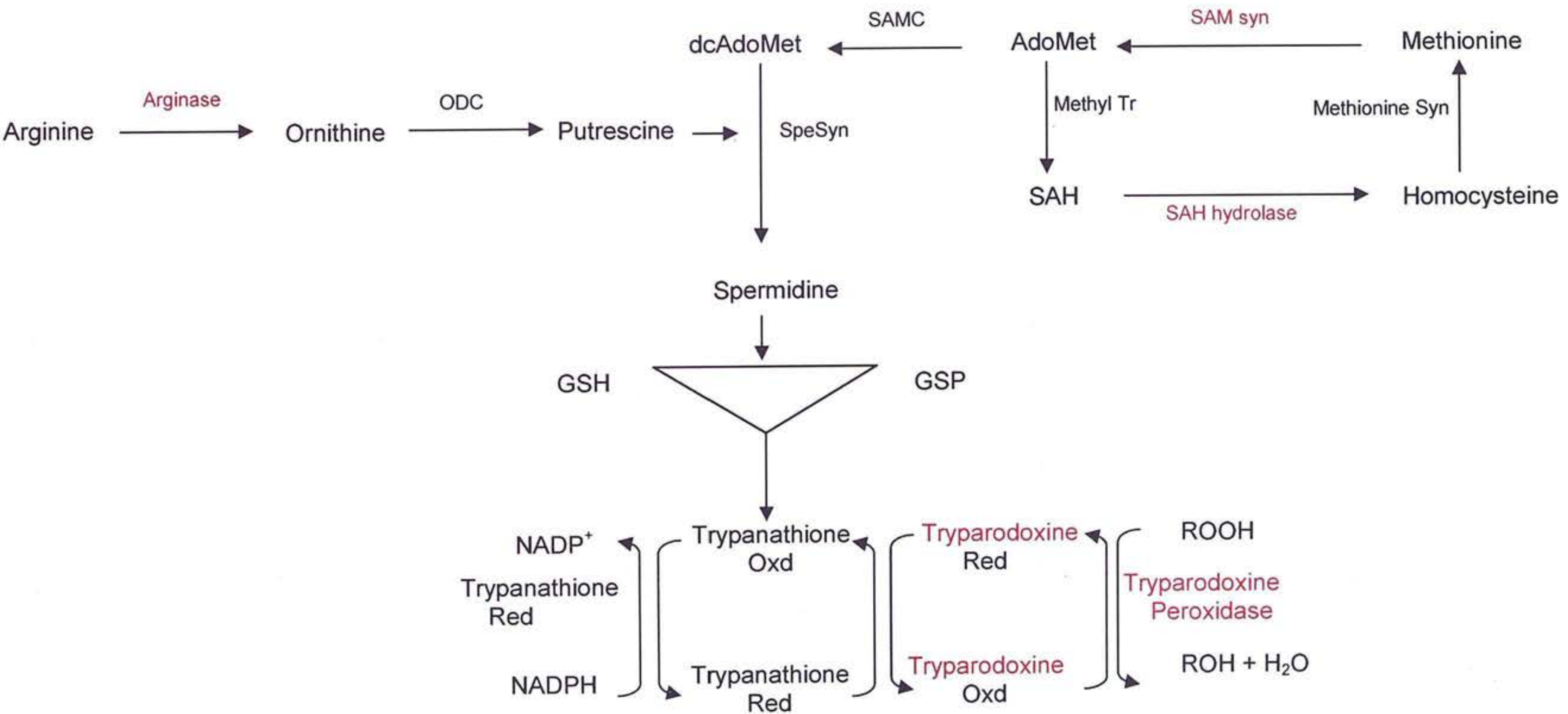


Figure 4.12

Polyamines pathway and peroxide metabolism in *T.brucei*. ODC: ornithine decarboxylase; SAMC: s-adenosylmethionine decarboxylase; Methyltr: methyltransferase; SAM: s-adenosyl methionine synthetase; SAH: s-adenosykmethionine hydrolase;SepeSyn:Spermedine synthetase; AdoMet: s-adenosylmethionine; dcAdoMet:decarboxylateAdoMet

Chapter 5

Proteomic analysis of *T.b. brucei* procyclic forms after 8-Br-cGMP treatment.

5.1 INTRODUCTION

Cyclic nucleotide signalling pathways via cyclic cGMP are present in many different organisms, ranging from protozoans to primates, where they are believed to play a role in many different cellular signalling pathways (Newton *et al* 2004). Cyclic GMP is formed by two distinct guanyl cyclases; (1) a membrane bound receptor molecule called particulate guanylyle cyclase (pGCs) and (2) a heme-containing cytosolic NO receptor enzyme called soluble guanylyl cyclase (Joshua, 2004). The action of cGMP is modulated through cGMP-dependent protein kinases, cGMP-regulated phosphodiesterases and the cGMP-regulation channels (Koesling *et al.*, 2004). In mammalian cells cGMP has a number of diverse roles; for example this molecule is involved in mediating the contraction and relaxation of smooth muscles in response to nitric oxide (Moncada *et al.*, 1992), regulating the movement of sodium ions and water through cellular membranes (Hofmann *et al.*, 1992) and controlling the permeability of the eye rod cell membranes to sodium ions, which are responsible for signal initiation along the optic nerve (Stryer, 1986).

In the *Trypanosoma* genus, cyclic nucleotides are involved in parasite differentiation (Vassella *et al.*, 1997) and evasion of host-immune responses (O'Beirne *et al.*, 1998). The cGMP pathway was first described in the genus *Trypanosoma* by Paveto (1995) who reported high levels of intracellular cGMP in *T. cruzi.* It has also been demonstrated *in vitro* that mobility of epimastigote forms of *T. cruzi* is enhanced upon the addition of c-GMP analogues (Pereira *et al.*, 1997). In *T. brucei*, it has been reported that a regulatory sub-unit of Protein kinase A (TbRSU/PKA) is stimulated by the presence of cGMP rather than cAMP and it was concluded that the TbRSU/PKA might represent the modulator of cGMP effects in *T. brucei* (Shalaby *et al.*, 2001).

Cyclic GMP is degraded by cyclic nucleotide phosphodiesterases (PDEs) **(Rybalkin *et al.*, 2001). In *T. brucei* PDE that have been characterized, namely TbPDE1 and TbPDE2. The TbPDE2** family of proteins have been found to **contain a GAF domain (**Seebeck *et al.*, 2001)**; one of the** functions of these domains, activation of the protein through the binding of cGMP, has been well characterised in proteins from mammalian cells (Charbonnea *et al.*, 1990). However in *T. brucei,* the TbPDE2 activity is not affected by the presence of cGMP and the protein has been shown to favor binding to cAMP (Seeback *et al.*, 2001). In other Trypanosomatidae, the role of cGMP in the parasite life cycle is being investigated. Cyclic GMP-dependent

enzymatic activity has been identified in *Leishmania amazonensis* (Géige *et al.*, 2003). It has also been suggested that the cGMP pathway in *Plasmodium falciparum* plays a role in exflagellation (Kawamoto *et al.*, 1990) a process by which the gametocytes transform into mature gametes within minutes of ingestion by a mosquito (Carter *et al.*, 1977)

During the course of this study we investigated proteins that are differentially expressed during treatment of *T.b. brucei* with a c-GMP analogue. Previous work carried out with c-GMP treated *T. brucei* in which 8 Br-cGMP, non-hydrolysable and cell-permeable analogue of cGMP, was found to significantly increase levels of midgut infection in the vector fly when introduced via an infected blood meal (Macleod *et al.*, in press 2005). Procyclic cultures were therefore treated with 8-Br-cGMP in order to ascertain whether this chemical has an affect on the protein expression profile of these parasites.

Identification of proteins differentially expressed in response to the 8 Br-cGMP in the treated parasites culture was carried out using SELDI-TOF Mass spectrophotometry. This technique uses a series of ProteinChip® arrays with different chemical surface properties, to capture proteins according to their biochemical properties. The presence and absence of proteins are determined through analysis of computer generated spectra which illustrate the protein profile of the sample.

5.2 MATERIAL AND METHODS

5.2.1 Culture treatment

T.b. brucei Buteba 135 stocks were used to infect flies as described in section 2.4.2. Infected fly guts in Cunningham medium were used to initiate procyclic parasite cultures, as described in section 2.4.3. These parasite cultures were prepared to a concentration of 10^7 parasites per ml and split into two stocks, the first of which was treated with 8-Br-cGMP at a final concentration of 100 μM; these will be referred to as the "**treated**" preparation. The "**control**" trypanosome preparation was treated only with the solvent that had been used to prepare the 8-Br-cGMP solution (2.4.5).

5.2.1 Sample preparation

A 1 ml aliquot of cell culture from each sample was washed twice in a buffer containing 10mM Tris–Cl at pH 7 and 25 mM sorbitol. The cells were harvested after each wash by a centrifugation at 4000 g for 5 min at 4° C. Cells were then lysed by the addition of 200 μl of Protein lysis buffer [7 M Urea, 2 M Thiourea, 4% CHAPS, 1 Tablet Complete, EDTA-free Protease inhibitor (Roche)]. After the second wash the pellet was thoroughly mixed by vortexing and the mixture was stored immediately at -70°C.

The predicted protein concentration of each sample was determined following the protocol described by van Deursen (2003), where typically 1 x 10^7 trypanosomes provide 100 μg of total proteins.

5.2.2 Surface-Enhanced Laser Desorption/Ionization Time-of-Flight Mass Spectrophotometer (SELDI-TOF- MS)

SELDI-TOF MS analyses were carried out at the Functional Genomics Units, Moredun Research Institute, Edinburgh. The analyses were performed using the Ciphergen ProteinChip PBS II Reader (Ciphergen, Palo Alto, USA).

The SELDI-TOF MS method consists of four steps:

1. A crude cell lysate is applied directly onto the surface of the ProteinChip®; this is referred to as the "spot".
2. The ProteinChip® surface is then subjected to a washing step where non-specifically bound proteins and buffers are eliminated.
3. An Enhancing Absorbing Molecule (EAM) is added to the samples to facilitate the ionisation and desorption of the proteins.
4. The captured proteins are analyzed in a ProteinChip® reader (Ciphergen ProteinChip PBSII Reader).

Upon the laser firing onto the spot, the lasers energy induces ionisation and a change in the status of captured proteins, from crystallized into gaseous from. The positively charged gaseous particles are repulsed from the surface of the ProteinChip®, which is also charged positively, by a continuous voltage. The continuous voltage applies the same energy to all proteins; this causes protein particles to fly away from the surface of the ProteinChip, and travel according to their mass and not by their charge. The protein Chip reader records the time of flight of the particles and derives a relative mass for each protein, this is illustrated as a mass/charge ratio

5.3 RESULTS

5.3.1 SELDI-TOF-MS Proteins Analysis

5.3.1.1 External calibration

External calibration of the PBS II Reader was performed before each run. Multiple calibration points were determined using 4 proteins standards with masses ranging between 18 -150 kDa (see Table 5.1).

Table 5.1

Comparison of MW of protein standards provided by manufacturer and MW of protein standards obtained with PBS II

Protein standard	Masses provided by the Manufacturer (kDa)	Obtained Masses on PBSII (kDa)
B-Lactoglobulin	18. 363	18. 7014
Horse Radish Peroxidase	43. 240	43. 6888
Bovine Serum Albumin	66. 433	67. 2479
Immunoglobulin G	147.300	148. 7978

Proteins standards were docked directly onto a NP 20 ProteinChip®array and processed as described in section 2.9.1. The obtained molecular weights of each standard are shown on Table 5.1 and Figure 5.1; the calculated molecular weights corresponded to those provided by manufacturer. The IgG and BSA proteins have doubly charged ions which produced an additional peak at half the assigned mass (74 and 34 kDa respectively). Doubly charged ions drift faster than single ions from the same species (Figure 5.1). The protein ion charge state is determined both by the nature of the polypeptide, such as numbers and identities of basic residues, and by the mechanisms for ionisation (Min He et al 2003). Doubly charged proteins ions were also used for the external calibration of PBS II

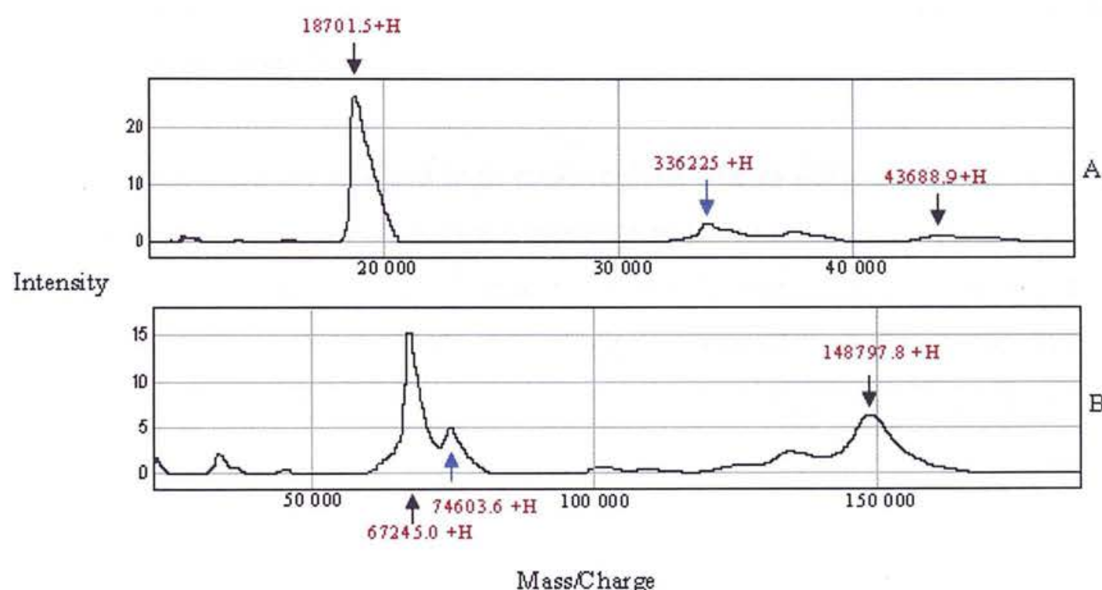


Figure 5.1

External Calibration of PBS II Reader

4 protein standards for external calibration were docked on a normal phase ProteinChip® array (NP 20) B-Lactoglobulin (18 kDa); Horse-radish Peroxidase (43 kDa); Bovine Serum Albumin (67 kDa) and IgG (148 kDa). The Y axis corresponds to the relative mass intensity and the X axis represents the molecular weight calculation in Daltons (Mass/Charge). Panel A represents low molecular weight proteins, which the detection size range was between 3-150 kDa. Panel B represents data acquisition of high molecular weight proteins, the detection size range of which was between 30-100 kDa .The blue arrows represent double charged version of BSA (33 kDa) and IgG (74 kDa) which cause a peak at half the assigned mass (67 kDa) and (148 kDa) respectively.

5.3.1.2 Optimisation and SELDI-TOF MS analysis

The ProteinChip assay selected for this optimisation were; a strong Anion Exchange (SAX 2), a weak anion exchange (WCX 2) and a hydrophobic interaction chromatographic surface (H 50). These were initially chosen to develop the SELDI – TOF MS analysis protocol for the protein samples extracted from the control procyclic forms as they provided a broad range of conditions under which the crude protein extract could be investigated.

A spot of 50 ng of protein was found to provide the optimum conditions for sensitivity and protein peak resolution with all the ProteinChip® surface arrays.

5.3.1.2.1 Optimisation of Protein profiling using WCX2

WCX 2 is a weak cationic exchange surface; optimisation was carried out at pH 4 and 6. The anionic surface of this matrix binds positively charged residues (such as lysine, arginine, and histidine). The WCX 2 ProteinChip®array was prepared for peak detection as described in section 2.9.1.

Peak determination for low and high mass regions was as described in section 2.9.2. Reading of the low mass region of the spectrum revealed a flat spectrum with a single unique peak occuring at 44 kDa (Figure 5.2 A and B), no protein peaks were detected in the high mass regions, (Table 5.2). An extra manual peak inspection failed to recognise any extra protein signatures, as each selected peak was below the signal/ratio threshold and was therefore considered as noise.

Increasing the washing stringency to pH 6 did not reveal additional protein peaks. On the contrary, the peak previously detected at pH 4 did not appear under these conditions (Figure 5.2 B and C). The positive charge on the 44 kDa protein was not sufficient at pH 6 to bind avidly to the WCX2 surface, but became so at pH 4 when more NH_3^+ ions were available.

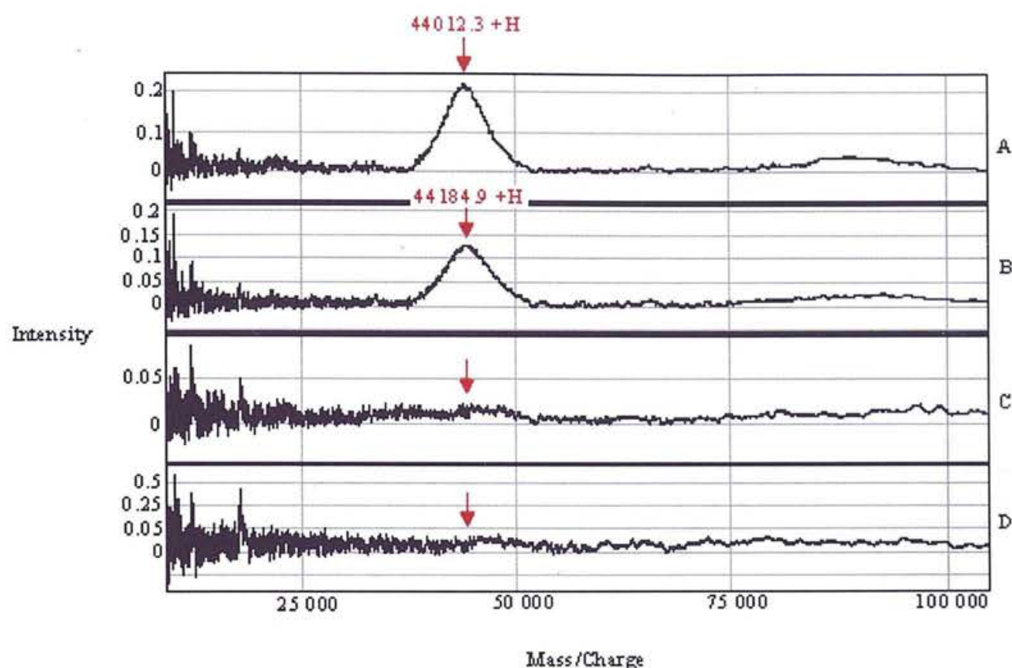


Figure 5.2

SELDI-TOF MS data obtained with WCX 2 ProteinChip array loaded with a *T.b. brucei* protein extract at pH 4.0 and at pH 6.0.

The X axis represents the molecular weight calculation in Daltons (M/Charge).The Y axis corresponds to the relative mass intensity and. Panels A and B present duplicated runs at pH 4.0. A unique protein peak was revealed at 44 kDa. Panel C and D represent a duplicate run of *T.b. brucei* protein extract with a higher stringency wash in which no protein peaks were detected.

5.3.1.2.2 Optimisation of Protein profiling using SAX2.

The cationic surface of the SAX 2 ProteinChip array binds and retains negatively charged residues (for example, aspartic acid, and glutamic acid residues). ProteinChip® preparation and peak detection were carried out as described in section 2.9.1/2. Between ProteinChip® reproducibility was evaluated by visually comparing the five proteins peaks that showed the highest amplitudes (11, 12, 14, 15 and 17 kDa-Figure 5.3). All the selected peaks showed both consistency and reproducibility.

A combination of the automatic Wizard option and manual detection for protein peaks revealed a complex spectrum with 36 protein peaks (Table 2). The majority of the proteins were found in the low mass region (32 peaks), and 4 peaks proteins were detected in the high mass region. The spectrum for the protein region of 3-30 kDa (Figure 5.3) clearly shows the five main peaks, however zooming into the sub region between 20-30 kDa revealed additional peaks; these had been masked due to the very

high abundance of the five high intensity proteins (Figure 5.3). A complete list of proteins identified is provided in Table 5.2.

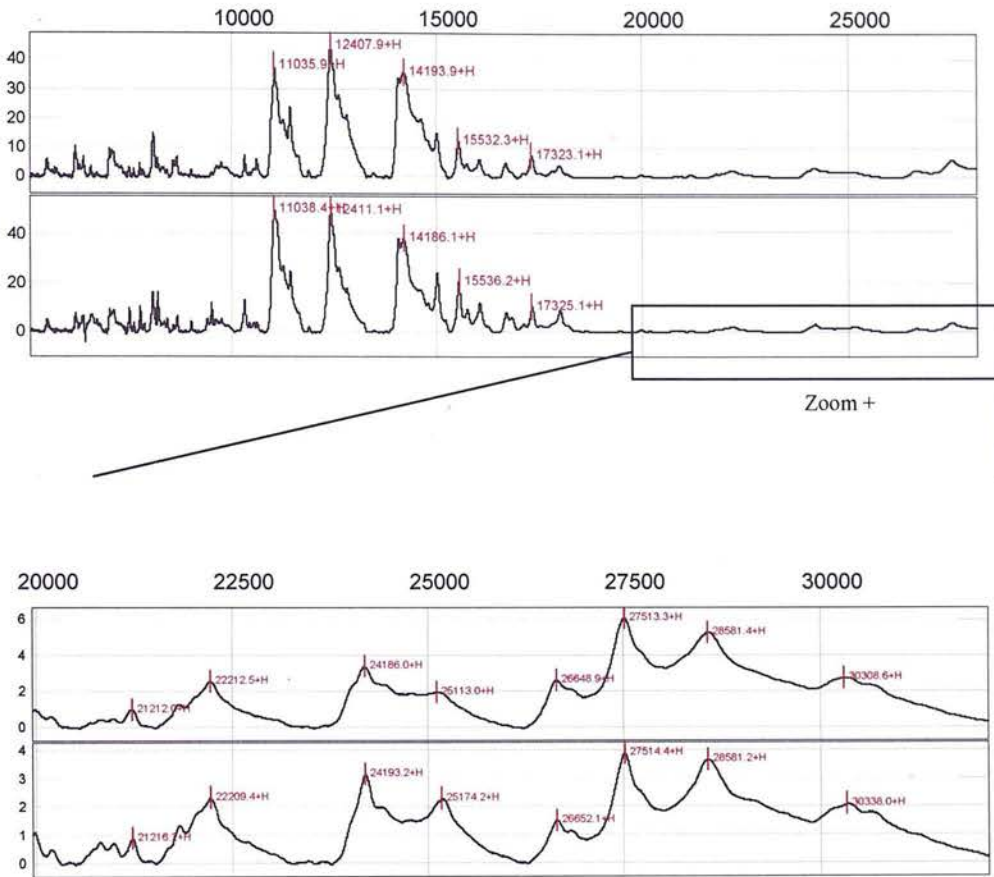


Figure 5.3

SELDI-TOF MS data obtained with a SAX 2 ProteinChip array loaded with *T b brucei* protein extract. Between chips reproducibility of mass spectra is shown by comparing five proteins peak with the highest consistent amplitudes (11, 12, 14, 15 and 17 kDa).Zooming in over the 20-30 kDa revealed additional protein peaks.

5.3.1.2.3 Optimisation of Protein profiling using H 50

H 50 has a hydrophobic surface that contains multiple chains of 16 methylene groups (C 16). The C16 binds to hydrophobic residues on protein surfaces (for example, alanine, valine, leucine, isoleucine, phenylalanine, tryptophane, tyrosine), The ProteinChip® preparation and peak detection were carried out as described in section 2.9.1/2.

38 peaks (Table 5.2) representing protein/ peptides masses were detected, ranging from 3 kDa to 66 kDa (Figure 5.4). Zooming into the sub region between 3 to 5 kDa revealed additional proteins peaks (Figure 5.4 C) as listed on Table 5.2.

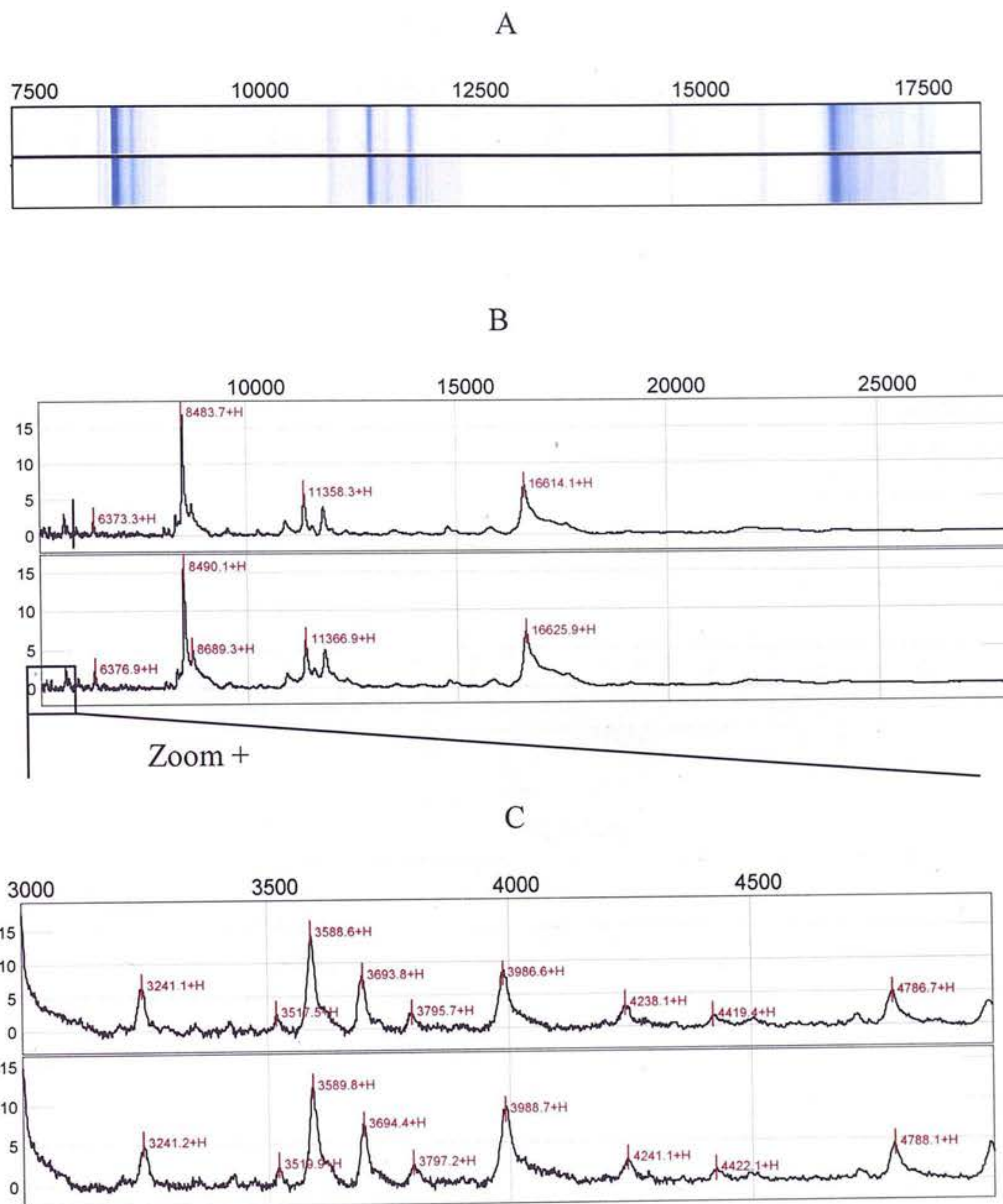


Figure 5.4

SELDI-TOF MS data obtained with H 50 array surface with *T.b. brucei* protein extract.
In panel B Proteins are shown in peak display and in panel A the protein are displayed in virtual gel view". Panel A transforms peak height into band intensity and was used to enhance inter-ProteinChip reproducibility. Panel C represents a zoom in, over the region of 3 to 5 kDa where more proteins could be visualized.

5.3.1.3 Summary of SELTI-TOF MS optimisation.

These results showed clear differences between the protein profiles produced by WCX 2, SAX 2 and H 50 proteinChip surface arrays. The chemical surfaces of each array can be used to capture *T b brucei* proteins selectively, according to their unique biochemical properties (Figure 5.5).

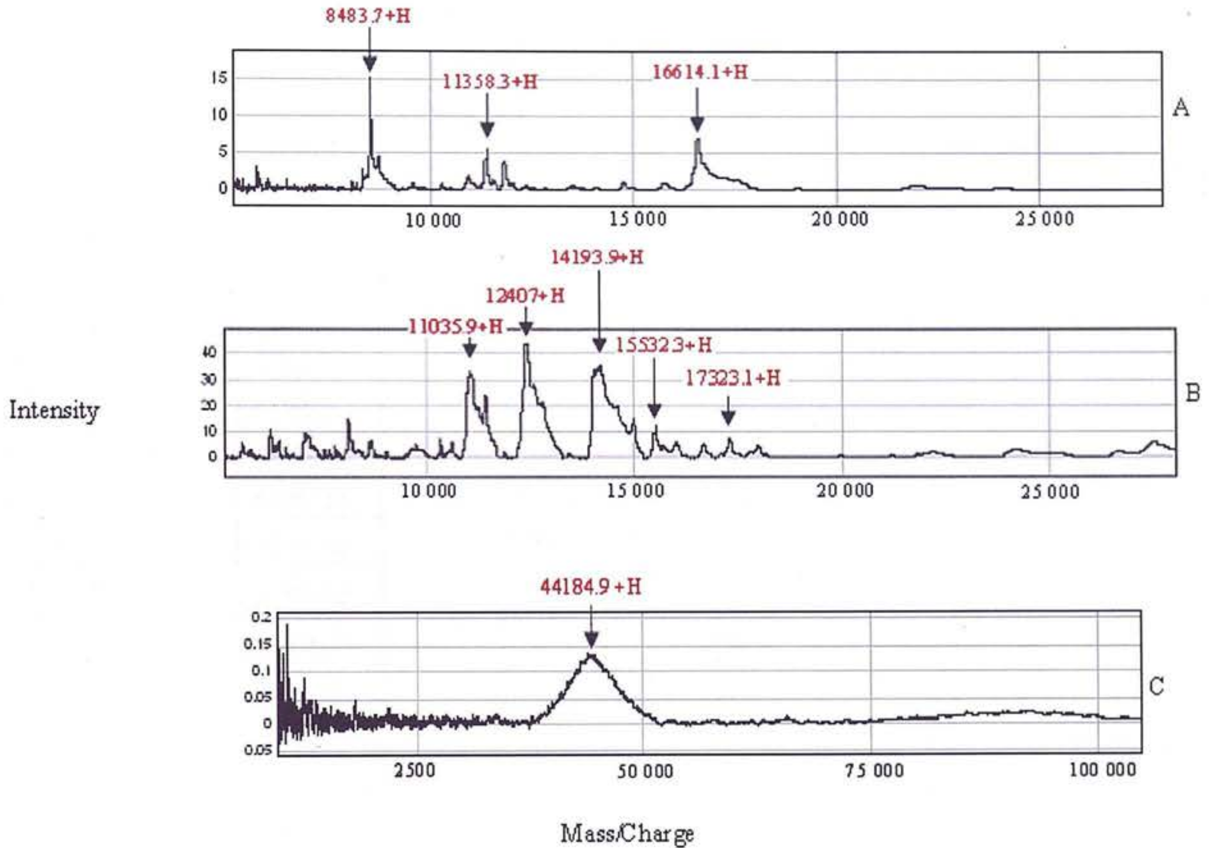


Figure 5.5

SELDI-TOF MS data from different array surfaces loaded with *T b brucei* protein extracts. Panel A represents H 50: hydrophobic interaction surface; Panel B represents SAX 2: strong anionic surface; Panel C represents WCX 2: weak cationic surface.

SAX 2 and H 50 ProteinChip® arrays were selected to investigate proteins differentially expressed in *T.b. brucei* under the influence of 8-Br-cGMP, as they produced busy and reproducible spectra. Because the number of proteins retained by WCX2 was limited, this array was not used in later analysis.

Table 5.2

Peak proteins detected with 3 different proteinChips surface.

SAX Mass (Da)		H 50 Mass (Da)			WCX Mass (Da)	
Low	High	Low		High	Low	High
5524.44	52432.83	3419.06	21918.67	53477.67	44757.3	
6208.18	58263.75	3584.81	24053.77	66437.98		
7033.76	66319.97	3689.15	27541.04			
8089.51	78697.82	3791.72	33057.81			
8667.71	80418.22	3986.7	33352.46			
10314.66	94731.81	4236.77	44185.58			
11035.91		4338.02	48299.44			
11416.57		4387.33	53477.67			
12407.87		4497.58	66437.98			
14193.79		4569.11				
15004.09		4667.07				
15532.3		4695.67				
16047.43		4782.1				
16692		4986.82				
17323.13		5339.34				
18007.06		5674.99				
19459.68		5748.83				
19989.83		6370.03				
21211.97		6543.69				
22212.55		8145.8				
24186.03		8492.01				
25174.19		10905.51				
26648.88		11273.47				
27513.25		11357.12				
28581.41		11803.38				
30308.62		13445.23				
33180.99		14789.65				
34326.92		15756.56				
44359.76		16593.32				

5.3.1 Identification of proteins differentially expressed during treatment of *T.b. brucei* with 8-Br-cGMP

5.3.2.1 Protein design

Three independent procyclic cultures of *T.b. brucei* (A, B and C) were treated with 100μM of 8 br-cGMP (T_A, T_B and T_C) as described in section 5.2.1. Each experimental preparation was partnered by a control (C_A, C_B and C_C) treated with solvent alone.

Treated parasite samples were checked by microscopy 24 hours post treatment for morphological changes; none were observed. Parasites were then harvested by centrifugation and proteins extracted as described in section 5.2.2. Differences between the control and treated protein preparations were investigated using SAX2 and H 50 ProteinChip® assays, following the protocol described in section 2.9.1, with external calibration performed as described in section 5.3.1.4. ProteinChips® were analyzed on PBS II (version 3.0) and each spot was analysed over a range of low and high molecular weights.

5.3.2.2 Identification of differentially expressed proteins using H 50 ProteinChip

The ProteinChips preparation followed the protocol described for the H 50 surface array in section 2.9.1. Data analysis of the protein profiles did not reveal any reproducible differences between the treated and control samples under the experimental conditions used and described in section 5.3.1.2.3.

5.3.2.3 Identification of differentially expressed proteins using SAX 2 ProteinChip array surface

The ProteinChip® preparation followed the protocol described for the SAX 2 surface array in section 5.3.1.2.2.

5.3.2.3.1 Data analysis

Automatic and manual detection of protein peaks from both the treated and control groups were carried out, with peak clustering across the experiments (A, B, C) based on a 0.3% mass window. The resulting protein profiles were compared for differences in the presence and relative levels of polypeptides.

Protein profiles obtained from the treated and control samples showed obvious differences in the low protein mass region between 27-34 kDa. These differences

were detectable in all three sets of experiments (Figure 5.6). Protein peaks at 27, 28 and 30 kDa showed a decrease in relative intensity in treated samples when compared to their control counterpart (Figure 5.6), On the other hand proteins peaks at 33 and 34 kDa showed an relative increases between the treated and control preparations (Figure 5.6).

5.3.2.3.2 Data normalisation

To further investigate differences between treated and control samples at the spectral level, a normalisation step was undertaken on relevant protein peaks (Figure 5.6). This step compensates for any shifts due to sample loading variation and slight shifts in mass due to imperfection on the ProteinChip® array surface. Normalised peak intensities were subject to biomarker identification using the Biomarker Wizard as described in section 2.9.2.

To highlight differences in protein expression between the treated and control samples based on their relative intensity over the 24-35 kDa protein regions (Figure 5.7). Overlapping peak intensities such as 24 kDa and 25 kDa (Figure 5.7) were considered showing no differences between control and treated sample and were used as internal controls. Differences in peak intensity were detected in regions between 27 to 34 kDa (Figure 5.7). The biomarkers Wizard singled out three protein clusters between 27 and 30 kDa as being suppressed in treated samples, with a protein cluster peak at 28.5 kDa being significantly suppressed in the treated samples (Figure 5.7). On the other hand, the biomarker Wizard indicated that 33 and 34 kDa proteins in the treated samples were relatively over-expressed (Figure 5.7).

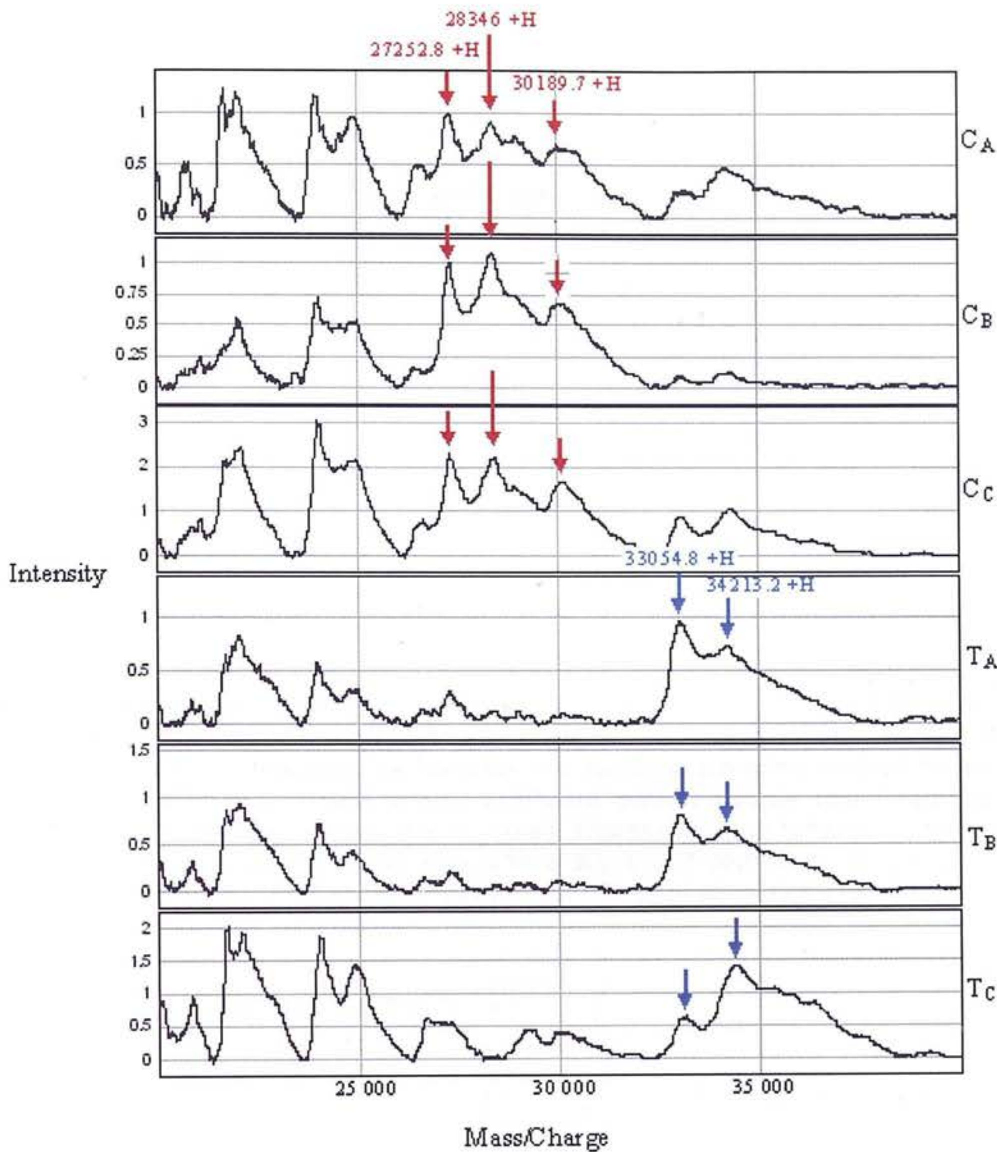


Figure 5.6

SELDI-TOF MS data obtained on SAX with proteins extract from treated and untreated *T.brucei.*

C represents control protein extract and T represents treated protein extract, A, B and C, represent three independent experiments. The spectra show the region between 22-35 kDa in the treated and the control samples. Differential expression of 27; 28.5; 30; 33 and 34 kDa peaks (strong anionic exchange) are clearly shown over the three experiments (A, B, C) between treated and control sample.

Red arrows represent peak proteins suppressed under treatment conditions and blue arrows represent proteins peaks potentially over-expressed during treatment.

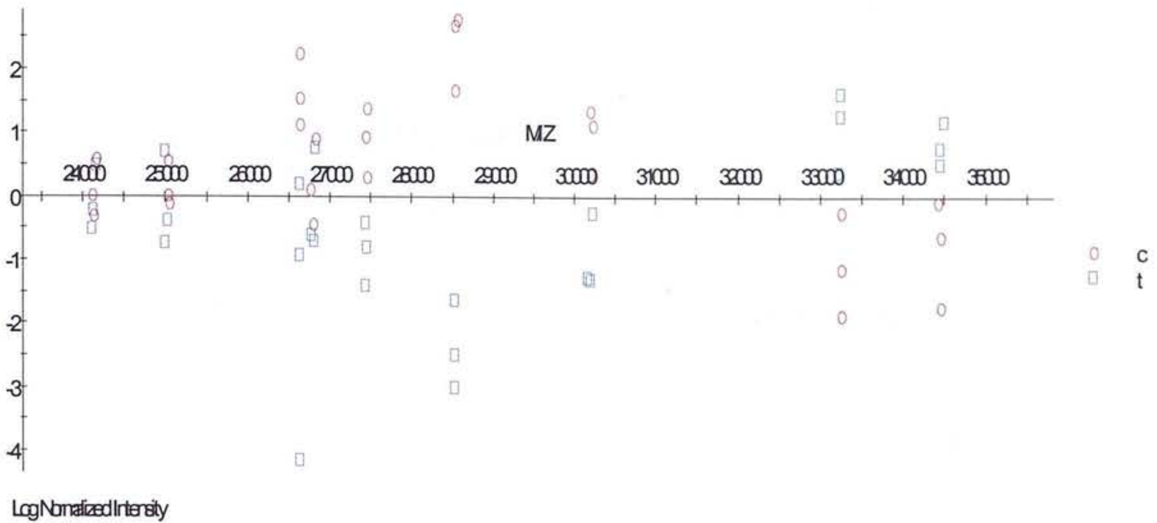


Figure 5.7

Biomarker identification wizards used over the protein region between 24-35 kDa. The wizard uses the log normalized intensity for detecting differences obtained on the SAX ProteinChip; proteins are classified across the samples into clusters according to their molecular weight (0.3% mass window). C (red circles) represent control sample and t represents treated samples (blue squares). Differences in protein expression level between the two groups are apparent over the regions 27-35 KDa with a significant pattern at 28.5 KDa. Overlapping values represent a non-distinctive pattern for differentiation (24 and 25 kDa) and were considered as internal controls.

5.3.2.4 Data base results.

A search in the Entrez Protein database (NCBI):

http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Protein&itool=toolbar) was undertaken as described by Stiles (2004). Briefly, the defined molecular weights from the SELDI-TOF MS analysis and the organism from which the protein originated were entered as a key search word into the PubMed database, to retrieve a list of corresponding proteins.

This approach was adopted for the 5 *T. brucei* proteins highlighted as being affected by the 8-Br-cGMP treatment:

Suppressed:

027000: 030000 [Molecular Weight] AND trypanosoma [Organism]

- Adenylate kinase
- Hypothetical protein Tb927.2.5400
- Hypothetical protein Tb927.2.4980

Over-expressed:

033000: 034000 [Molecular Weight] AND trypanosoma [Organism]

- Hypothetical protein Tb927.2.3970
- Putative mitochondrial carrier protein
- Hypothetical protein Tb927.2.2830

The searches yielded four hypothetical proteins with similar listed molecular weights. These proteins were generated through the *T.brucei* genome project and are based on putative open reading frames. Little additional information can therefore be gleaned from these results. The adenylate kinase (ADK) identified from *T.brucei* has a calculated molecular weight of 29 kDa. ADK are phosphotransferases that catalise the reversible reaction ATP + AMP <==> 2 ADP (Ulschmid *et al.*, 2004). The putative mitochondrial carrier protein had a calculated MW of 33 kDa. Mitochondrial carrier's proteins are found within the inner mitochondrial membrane and include protein such as ADP/ATP translocase (Nelson *et al.,* 1993)

5.4 DISCUSSION

During the course of this study procyclic cultures of *T.b. brucei* BUT 135 were treated with 100 μM 8-Br-cGMP. The treated parasites were compared with untreated samples which served as a control. The samples were compared at the polypeptide level using SELDI-TOF MS in an attempt to identify differentially expressed proteins.

The study initially involved development of a protocol for the generation of suitable spectra for analysis of differentially expressed proteins in *T. b. brucei* using surface enhanced laser desorption/ionisation time of flight mass spectrophotometry (SELDI-TOF MS). Three different ProteinChip® matrices were evaluated during the optimisation process; SAX 2, which binds negatively charged residues, H 50, which binds hydrophobic residues on protein surfaces and WCX 2 which binds positively

charged residues. The SAX 2 and H 50 ProteinChip® array surfaces, which generated complex spectra, with many individual polypeptides, were selected for further analyses. In contrast WCX 2 showed a limited protein/ peptides with a single major peak at 44 kDa (Figure 5.1), and therefore was not used in subsequent experiments.

The 8-Br-cGMP treatment of procyclic forms of *T.b. brucei* did not induce any significant differences from the control parasites on the H 50 hydrophobic ProteinChip®. However, the proteins characterised using SAX 2 ProteinChip® surface array at pH 8, did show significant differences between treated and control cultures. The 8-Br-cGMP treatment caused a down-regulation in polypeptides in the mass region between 27 - 30 kDa, with a particularly strong down-regulation of the 28.5 kDa protein peak. The treatment was also found to induce a slight up-regulation in two protein peaks found between 33- 34 kDa.

In our study, we demonstrated that 8-Br-cGMP, an analogue of cGMP, has a visible effect on the procyclic form of *T. b. brucei*. However, the existence of a cGMP pathway and the downstream effects of this form of signalling in *T.b. brucei* remain to be identified, since guanylyl cyclases have yet to be characterized in these parasites (Naula *et al.,* 2000). However, the effect of cGMP could still be mediated through the cAMP signalling pathway and particularly through a cAMP dependant protein kinase; where it has been found that the TbRSU/PKA sub-units binds selectively to cGMP rather than cAMP (Shalaby *et al.,* 2003).

The *T. brucei* PKA sub-units are reported to be biochemically different from those of mammalian PKA, however relatively little is known on the downstream effectors of cAMP signalling in kinotoplastids (Seeback *et al.,* 2004). Cross activation between cAMP and cGMP signalling pathways have been reported. Schwede (2000) suggested that the presence of high intracellular levels of one second messenger could trigger the cross activation of the other. In *T. cruzi*, PKA activation might induce the phosphorylation of specific protein targets involved in the differentiation of epimastigotes to trypomastigotes (Flavia *et al.,* 1997).

Branerie (1992) isolated a protein in *L. donovani* with similar properties to mammalian PKA. The molecular weight of the proteins was 34 KDa. One of the

proteins differentially expressed characterized during the treatment was also a 34 kDa protein which was over-expressed as a response to cGMP treatment (Figure 5.6).

Down regulation of proteins expression in response to cGMP seen during the course of this study is widely reported in mammalian cells. cGMP was shown to have an inhibitory effect on catecholamine secretion through a cGMP-protein kinase mediated through the regulation of the cytosolic calcium levels (Rodriguez-Pascual *et al.*, 1995). Increased levels of c-GMP can inhibit platelet aggregation through the activation of PKG and the inhibition of PDE (Maurice *et al.*, 1990; Bowen *et al.*, 1991). Braam (2004) studied the effect of Nitric oxide (NO) donors on the gene expression status of endothelial cells. NO increased the level of intracellular cGMP through the activation of soluble guanylyl cyclase, which in turn stimulated the formation of cGMP. They reported that the administration of NO donors decreased expression of genes such as those coding for growth factors, cell structure and adhesion molecules.

Following the approach applied by Stiles (2004) to identify protein peaks within the spectrum generated by the SELDI-TOF MS, the current reports results from the Entrez Protein database (NCBI) for the two main areas of differential expression resulting from the 8-Br-cGMP treatment. The search for proteins listed within the size range of 27-30 kDa, revealed an adenylate kinase (ADK) of 29 kDa. The search for proteins between 33 and 34 kDa listed putative mitochondrial carrier protein of 33 kDa. However, protein identity cannot be based solely on protein molecular weight and further analysis must be undertaken.

In unpublished work of Macleod (personal communication) assessed the effects of 8-Br-cGMP on the survival of *T. b. brucei* in the midgut of tsetse flies, suggesting that cGMP has an anti-apoptotic effect on trypanosomes inside the tsetse midgut. The c-GMP/PKG signalling pathway has also been reported to have anti-apoptotic effects in mammalian cells (Chan *et al.*, 2003). cGMP is believed to suppress caspase activity through activation of cGMP dependent protein kinases (Kim *et al.*, 1997). Expression of *T. b. brucei* metacaspases which are homologous to caspases (Uren *et al.*, 2000) in the yeast *Saccharomyces cerevisiae* lead to growth inhibition, mitochondrial dysfunction and clonal death (Szallies *et al.*, 2002). It is possible that the described

anti-apoptotic effects of cGMP on *T. b. brucei* could involve the proteins identified with SAX 2 ProteinChip® array (Figure 5.6).

During the course of this study we have successfully used SELDI-TOF MS method for the investigation of differentially expressed proteins in *T .b. brucei*. Firstly the method appeared to be ideal for the characterisation of differentially expressed protein in a complex biological system (parasite proteome). In addition we have characterised the biochemical properties of several target proteins which were found to differ in their expressions in the presence of a cGMP analogue. These proteins could constitute one of the downstream effects of a cyclic nucleotide related pathway in *T. b. brucei*. We have shown that cGMP treated parasites show differences in protein expression, and these particular effects could consequently play a role in the anti-apoptotic effect of cGMP and parasite establishment in the insect host observed by Macleod (personal communication).

5.5 FUTURE WORK

The composite data of the differentially expressed proteins obtained by SELDI-TOF MS on the SAX 2 ProteinChip® surface array provides enough information to design a unidimensional liquid chromatography separation system for protein purification. The purification of target proteins identified during the course of this study could be conducted on liquid chromatography columns and electrophoretic techniques. This would then allow further identification of the nature and identity of these proteins through MALDI-TOF analysis for protein/peptide identification and amino acid sequencing.

Chapter 6

General Discussion

6.1 Objectives for this project

Trypanosomes from the Salivarian group are the causative agents of HAT and nagana in African livestock. Although trypanosomes are one of the most studied parasites in the world, there are many questions that still require answers. As the technologies used to study this parasite continue to develop, the means with which to study these organisms are constantly expanding. The aim of this experiment was to apply both established and novel molecular methodologies to the study of establishment of procyclic populations of *T. brucei* within the tsetse vector.

The intention of this project was to define an appropriate experimental system with which to investigate various aspects of establishment of *T. brucei* within the tsetse vector. My approach was to apply a number of established genomic protocols as well as a novel proteomic methodology, with a view to determining two things:
Firstly, what differences exist between procyclic populations capable of maturation compared to those that did not?
Secondly,the treatment of 8 Br cGMP is known to influence the rate of establishment for *T. brucei* within tsetse, but what affect does the compound have on the parasite?

6.2 Experimental approach

In this study, five main experimental approaches were used to characterize differentially expressed molecules in the procyclic form of *T.b. brucei* (Table 6.1).

6.1.1 RADES-PCR

Rapid amplification of differentially expressed sequence is an established method developed specifically for the identification of differentially expressed genes in *T. brucei*. I was able to identify 28 differentially expressed transcripts between group ++ (salivary +/ gut +) and group +- (salivary +/ gut -), nine of which were successfully cloned and sequenced, with 3 clones identifiable on the *T.brucei* gene database. That is not to say however that the other 6 would not be identified in due course inline with the further annotation of the *T. brucei* genome.

The rest of the sequenced transcripts either produced poor quality data or recognized as material of prokaryotic origin. The method was initially described as non-cumbersome, however I found that because of the failure of most of the re-amplification steps I had to optimised the reaction at each re-amplification run. I found that running the RADES-PCR product on a denaturing polyacrylamide gel improved the separation of the products and since artefacts could be the result of the co-migrating bands.

6.1.2 Low/High Stringency PCR.

In attempt to reduce the amount of artefacts encountered during RADES-PCR. I used an anchored random priming approach combined with low/high stringency PCR parameters. We noticed that the efficiency of the amplification improved dramatically since a more specific banding pattern could be seen. The method also identified differentially expressed products between group ++ and group +-. However none of the product initially identified could be reproduced after re-amplification.

6.1.3 Subtractive cDNA library

A PCR-select subtracted cDNA technique was utilised for the identification of differentially expressed genes in *T. brucei*. The method provided a set of genes differentially-expressed during the treatment of *T. brucei* with 8-Br-cGMP (Table 6.1). This allowed me to identify some key enzymes that might be involved in this response to the treatment. A high prevalence of ribosomal transcripts were also identified, possibly the mRNA isolation step might be to blame.

6.1.4 Real Time PCR

Real time PCR is a means for quantifying the amount of starting material within a PCR reaction. The method is based upon the number of cycles at which the PCR reaction reaches the logarithimic stage, unlike other forms of quantitative RT-PCR that detect the amount of final amplified product at the end-point. This is achieved through the detection and quantification of a fluorescent reporter, in this case SYBR green (a non-sequence specific fluorescent intercalating agent). The more starting material present, the more rapidly the logarithimic phase is reached, the crossing threshold (C_T) of the reactions being the point at which the signal surpasses background noise and begins to increase. This approach can also be applied to

analysing the efficiency of a PCR reaction, so that gene expression can be compared relative to a housekeeping gene in order to demonstrate up or down regulation.

Relative quantification however relies on the efficiency of the PCR reaction of the compared transcripts. Different transcripts will automatically be amplified differently since the PCR reaction efficiency relies on several parameters such as the size of the transcript, the G+C content, the priming sequences ect. Therefore the efficiency of the compared transcripts needs to be comparable.

Real time PCR was used to quantify changes in gene expression of an Arginase gene in response to the addition of 8-Br-cGMP to procyclic forms of the *T. brucei* parasite. I was able to show that there was a slight up regulation in the arginase gene in comparison to a housekeeping gene (GAPDH).

6.1.5 SELDI-TOF MS

Surface-enhanced laser desorption ionization - time of flight (SELDI-TOF) is a new proteomic approach. This method utilises the same principles of liquid chromatographic methods for protein purification, however instead of a column system the SELDI-TOF uses chemical and biochemical surfaces. The creation of a protein profile is therefore based on the intrinsic properties of the proteins themselves. This method separates proteins between 3 and 100 kDaltons; the resolution of the protein profile being accurate to less than 1 Dalton (Table 5.2). In addition the application of small amounts of protein will still result in a complete profile. The information gleaned can be applied to large scale protein purifications or further small scale investigations using electrophoresis and MALDI-TOF analysis.

Cost is the main prohibitive to wide scale use of this technique. While the identification of differences in protein expression in response to a treatment provides a good starting point for further investigation, little can be learned directly from simply knowing the biochemical characteristics of the polypeptide (Section 5.3.2.4).

Table 6.1 1

Merits of methods used in this thesis.

Method	Advantages	Disadvantages
7 RADES-PCR	- Rapid	- Cumbersome - High levels of artefacts - Non specific
Low/High Stringency PCR	- Rapid - Improved stringency	- Reamplification and characterisation of identified cDNA problematic.
Subtractive cDNA library	- Provides a full set of differentially expressed genes.	- High incidence of ribosomal genes
Real Time PCR	- Rapid and quantitative means to measure changes in gene expression.	- Optimisation of amplification conditions.
SELDI-TOF MS	- Rapid - Provides sufficient information for further development of separation protocols.	- Limited information for characterisation of targets. - Cost

6.3 Future work

Using the systems developed in this project the scope for further investigating the establishment and maturation of *T. brucei* in the tsetse fly vector is endless.

The large number of genes highlighted by the subtractive cDNA library as activated in response to the 8 Br cGMP treatment could be further analysed by Real Time PCR in order quantify the level to which activation or repression occurs. Once these targets had been confirmed as molecular markers for this system, it would then be possible to analyse the changes that occur over time, in terms of the speed to which particular proteins are required in order to respond to the stimulus of 8-Br-cGMP. In addition it would also be worthwhile investigating other genes involved in the polyamine and trypanothione synthesis pathway.

With the ongoing genome project for *T. brucei*, the continuing analysis and sequencing of the nuclear genome may well result in the further identification of regions with homology to recombinant clones from within our subtractive cDNA library. At present 71 contigs remain classified as unknown; I would expect that over

time this number would reduce in line with completion of the genome sequencing project.

The subtractive cDNA library also highlighted a number of genes that were down regulated in response to the treatment with 8 Br cGMP, similar results observable in the protein profiles generated by the SELDI-TOF MS. Therefore by reversing the subtractive library in order to determine those genes down-regulated in the treated sample by their presence within the untreated sample could be used to isolated and characterize these genes. This library has already been created, so it would simply be a matter of generating and analyzing the recombinant clones.

Given more time it would have been nice to further characterize the proteins highlighted by the SELDI-TOF MS approach. While this technique offers a rapid means to analyze differences in the protein profile of samples based on their biochemical nature, characterization of these differences is hard to achieve without further experimentation. SELDI-TOF MS provides sufficient biochemical information to allow for further separation of the proteins of interest by liquid chromatography, using for example a high throughput system such as the PerSeptive Biosystem. The combination of these two approaches with electrophoretic techniques would allow for identification of these proteins through MALDI-TOF peptide analysis. Once characterized these proteins could then be used as serological targets through the raising of polyclonal antibodies.

Both the proteomic approach of the SELDI-TOF MS and the genomic approach of the cDNA subtractive library could be applied to the process of establishment and maturation of procyclic *T. brucei*. In particular applied to the question of differences between populations of parasites that are able to establish a mature infection compared to those that do not. Given that this would be a direct comparison between two populations it would lend itself well to these approaches. In particular the very nature of the subtractive library method would mean that any contamination arising from the presence of bacterial symbionts within the tsetse fly gut should be removed in the early hybridization steps. With the methodology and expertise already in place this could be a very interesting study. I have already identified 8 transcripts that may distinguish between these parasitic populations, and it could be very interesting to see if the subtractive library similarly identifies these as key elements.

[illegible], E., Ringeli, C., Brun, R., [illegible] (2004). The surface coat of procyclic Trypanosoma brucei [illegible] procyclin [illegible] 101(?), 1513-1518.

[illegible] Cell, 51(?), [illegible]

[illegible] (1996). [illegible] sequence of the [illegible] Nat Genet, 13(?), 403-407.

[illegible] N. L., Smith, D. M., Richardson, J. P., & Prowse, D. W. (1988). [illegible] of different [illegible] by two-dimensional gel electrophoresis. Mol Biochem Parasitol, [illegible]

[illegible] P., [illegible] T. P., & [illegible] (2003). [illegible] in host epithelial [illegible]

[illegible] C. R. (2001). [illegible] evolutionary link between [illegible] proteins. Trends Biochem Sci, 26(12), 458-459.

[illegible] A. R. (2001). [illegible] Mol Biochem Parasitol, 118(2), 199-195.

[illegible] (1992). Isolation and characterization of a cyclic [illegible] protein kinase from [illegible] Mol Biochem [illegible]

[illegible] (2001). The CLIP domain: a large family of [illegible] Trends Biochem Sci, [illegible]

[illegible] Van Laere, S., van Dam, P., Vermeulen, [illegible] & Dirix, L. Y. (2003). Real-time RT-PCR [illegible] for the detection of [illegible] epithelial cells [illegible] of patients with breast cancer. Br J Cancer, 89(10), 1813-[illegible]

[illegible] genes and [illegible] Prog Drug Res, 49, 253-282.

[illegible] C. M. (1998). [illegible] protein kinase [illegible] Biochem Soc Trans, [illegible] 223

References

Acosta-Serrano, A., Vassella, E., Liniger, M., Kunz Renggli, C., Brun, R., Roditi, I., & Englund, P. T. (2001). The surface coat of procyclic Trypanosoma brucei: programmed expression and proteolytic cleavage of procyclin in the tsetse fly. Proc Natl Acad Sci U S A, 98(4), 1513-1518.

Agabian, N. (1990). Trans splicing of nuclear pre-mRNAs. Cell, 61(7), 1157-1160.

Akman, L., Yamashita, A., Watanabe, H., Oshima, K., Shiba, T., Hattori, M., & Aksoy, S. (2002). Genome sequence of the endocellular obligate symbiont of tsetse flies, Wigglesworthia glossinidia. Nat Genet, 32(3), 402-407.

Anderson, N. L., Parish, N. M., Richardson, J. P., & Pearson, T. W. (1985). Comparison of African trypanosomes of different antigenic phenotypes, subspecies and life cycle stages by two-dimensional gel electrophoresis. Mol Biochem Parasitol, 16(3), 299-314.

Andley, U. P., Song, Z., Wawrousek, E. F., Fleming, T. P., & Bassnett, S. (2000). Differential protective activity of alpha A- and alphaB-crystallin in lens epithelial cells. J Biol Chem, 275(47), 36823-36831.

Aravind, L., & Ponting, C. P. (1997). The GAF domain: an evolutionary link between diverse phototransducing proteins. Trends Biochem Sci, 22(12), 458-459.

Ariyanayagam, M. R., & Fairlamb, A. H. (2001). Ovothiol and trypanothione as antioxidants in trypanosomatids. Mol Biochem Parasitol, 115(2), 189-198.

Banerjee, C., & Sarkar, D. (1992). Isolation and characterization of a cyclic nucleotide-independent protein kinase from Leishmania donovani. Mol Biochem Parasitol, 52(2), 195-205.

Bateman, A., & Rawlings, N. D. (2003). The CHAP domain: a large family of amidases including GSP amidase and peptidoglycan hydrolases. Trends Biochem Sci, 28(5), 234-237.

Benoy, I. H., Elst, H., Van der Auwera, I., Van Laere, S., van Dam, P., Van Marck, E., Scharpe, S., Vermeulen, P. B., & Dirix, L. Y. (2004). Real-time RT-PCR correlates with immunocytochemistry for the detection of disseminated epithelial cells in bone marrow aspirates of patients with breast cancer. Br J Cancer, 91(10), 1813-1820.

Bhat, S. P. (2003). Crystallins, genes and cataract. Prog Drug Res, 60, 205-262.

Bhat, S. P., & Nagineni, C. N. (1989). alpha B subunit of lens-specific protein alpha-crystallin is present in other ocular and non-ocular tissues. Biochem Biophys Res Commun, 158(1), 319-325.

Blaxter, M. L., Raghavan, N., Ghosh, I., Guiliano, D., Lu, W., Williams, S. A., Slatko, B., & Scott, A. L. (1996). Genes expressed in Brugia malayi infective third stage larvae. Mol Biochem Parasitol, 77(1), 77-93.

Bogdan, C., Rollinghoff, M., & Diefenbach, A. (2000). Reactive oxygen and reactive nitrogen intermediates in innate and specific immunity. Curr Opin Immunol, 12(1), 64-76.

Bowen, R., & Haslam, R. J. (1991). Effects of nitrovasodilators on platelet cyclic nucleotide levels in rabbit blood; role for cyclic AMP in synergistic inhibition of platelet function by SIN-1 and prostaglandin E1. J Cardiovasc Pharmacol, 17(3), 424-433.

Braam, B., de Roos, R., Dijk, A., Boer, P., Post, J. A., Kemmeren, P. P., Holstege, F. C., Bluysen, H. A., & Koomans, H. A. (2004). Nitric oxide donor induces temporal and dose-dependent reduction of gene expression in human endothelial cells. Am J Physiol Heart Circ Physiol, 287(5), H1977-1986.

Brehm, K., Wolf, M., Beland, H., Kroner, A., & Frosch, M. (2003). Analysis of differential gene expression in Echinococcus multilocularis larval stages by means of spliced leader differential display. Int J Parasitol, 33(11), 1145-1159.

Brun, R., & Schonenberger, M. (1981). Stimulating effect of citrate and cis-Aconitate on the transformation of Trypanosoma brucei bloodstream forms to procyclic forms in vitro. Z Parasitenkd, 66(1), 17-24.

Buechler, W. A., Nakane, M., & Murad, F. (1991). Expression of soluble guanylate cyclase activity requires both enzyme subunits. Biochem Biophys Res Commun, 174(1), 351-357.

Bursell, E., Billing, K. J., Hargrove, J. W., McCabe, C. T., & Slack, E. (1973). The supply of substrates to the flight muscle of tsetse flies. Trans R Soc Trop Med Hyg, 67(2), 296.

Bushman, F. (2002). Constructing the vertebrate genome: evidence from eels that LINEs mobilize SINEs. Mol Cell, 10(5), 961-962.

Butikofer, P., Ruepp, S., Boschung, M., & Roditi, I. (1997). 'GPEET' procyclin is the major surface protein of procyclic culture forms of Trypanosoma brucei brucei strain 427. Biochem J, 326 (Pt 2), 415-423.

Carter, R., & Nijhout, M. M. (1977). Control of gamete formation (exflagellation) in malaria parasites. Science, 195(4276), 407-409.

Cederbaum, S. D., Yu, H., Grody, W. W., Kern, R. M., Yoo, P., & Iyer, R. K. (2004). Arginases I and II: do their functions overlap? Mol Genet Metab, 81 Suppl 1, S38-44.

Chan, S. L., & Fiscus, R. R. (2003). Guanylyl cyclase inhibitors NS2028 and ODQ and protein kinase G (PKG) inhibitor KT5823 trigger apoptotic DNA fragmentation

in immortalized uterine epithelial cells: anti-apoptotic effects of basal cGMP/PKG. Mol Hum Reprod, 9(12), 775-783.

Charbonneau, H., Prusti, R. K., LeTrong, H., Sonnenburg, W. K., Mullaney, P. J., Walsh, K. A., & Beavo, J. A. (1990). Identification of a noncatalytic cGMP-binding domain conserved in both the cGMP-stimulated and photoreceptor cyclic nucleotide phosphodiesterases. Proc Natl Acad Sci U S A, 87(1), 288-292.

Covini, N., Tamburin, M., Consalez, G., Salvati, P., & Benatti, L. (1999). ZFM1/SF1 mRNA in rat and gerbil brain after global ischaemia. Eur J Neurosci, 11(3), 781-787.

Creedon, K. A., Rathod, P. K., & Wellems, T. E. (1994). Plasmodium falciparum S-adenosylhomocysteine hydrolase. cDNA identification, predicted protein sequence, and expression in Escherichia coli. J Biol Chem, 269(23), 16364-16370.

Cross, G. A., & Manning, J. C. (1973). Cultivation of Trypanosoma brucei sspp. in semi-defined and defined media. Parasitology, 67(3), 315-331.

Cunningham, I. (1977). New culture medium for maintenance of tsetse tissues and growth of trypanosomatids. J Protozool, 24(2), 325-329.

Cunningham, I., & Honigberg, B. M. (1977). Infectivity reacquisition by Trypanosoma brucei brucei cultivated with tsetse salivary glands. Science, 197(4310), 1279-1282.

Czichos, J., Nonnengaesser, C., & Overath, P. (1986). Trypanosoma brucei: cis-aconitate and temperature reduction as triggers of synchronous transformation of bloodstream to procyclic trypomastigotes in vitro. Exp Parasitol, 62(2), 283-291.

Dale, C., & Welburn, S. C. (2001). The endosymbionts of tsetse flies: manipulating host-parasite interactions. Int J Parasitol, 31(5-6), 628-631.

Dale, C., Welburn, S. C., Maudlin, I., & Milligan, P. J. (1995). The kinetics of maturation of trypanosome infections in tsetse. Parasitology, 111 (Pt 2), 187-191.

Davies, D. E., Djukanovic, R., & Holgate, S. T. (1999). Application of functional genomics to study of inflammatory airways disease. Thorax, 54(1), 79-81.

de Jong, W. W., Leunissen, J. A., & Voorter, C. E. (1993). Evolution of the alpha-crystallin/small heat-shock protein family. Mol Biol Evol, 10(1), 103-126.

De Lange, T., Berkvens, T. M., Veerman, H. J., Frasch, A. C., Barry, J. D., & Borst, P. (1984). Comparison of the genes coding for the common 5' terminal sequence of messenger RNAs in three trypanosome species. Nucleic Acids Res, 12(11), 4431-4443.

Diachenko, L. B., Ledesma, J., Chenchik, A. A., & Siebert, P. D. (1996). Combining the technique of RNA fingerprinting and differential display to obtain differentially expressed mRNA. Biochem Biophys Res Commun, 219(3), 824-828.

Diatchenko, L., Lau, Y. F., Campbell, A. P., Chenchik, A., Moqadam, F., Huang, B., Lukyanov, S., Lukyanov, K., Gurskaya, N., Sverdlov, E. D., & Siebert, P. D. (1996). Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries. Proc Natl Acad Sci U S A, 93(12), 6025-6030.

Donald, R. G., & Liberator, P. A. (2002). Molecular characterization of a coccidian parasite cGMP dependent protein kinase. Mol Biochem Parasitol, 120(2), 165-175.

Donelson, J. E. (2003). Antigenic variation and the African trypanosome genome. Acta Trop, 85(3), 391-404.

Dragon, F., Gallagher, J. E., Compagnone-Post, P. A., Mitchell, B. M., Porwancher, K. A., Wehner, K. A., Wormsley, S., Settlage, R. E., Shabanowitz, J., Osheim, Y., Beyer, A. L., Hunt, D. F., & Baserga, S. J. (2002). A large nucleolar U3 ribonucleoprotein required for 18S ribosomal RNA biogenesis. Nature, 417(6892), 967-970.

Duleu, S., Vincendeau, P., Courtois, P., Semballa, S., Lagroye, I., Daulouede, S., Boucher, J. L., Wilson, K. T., Veyret, B., & Gobert, A. P. (2004). Mouse strain susceptibility to trypanosome infection: an arginase-dependent effect. J Immunol, 172(10), 6298-6303.

Ellis, D. S., & Evans, D. A. (1977). Passage of Trypanosoma brucei rhodesiense through the peritrophic membrane of Glossina morsitans morsitans. Nature, 267(5614), 834-835.

Esseiva, A. C., Chanez, A. L., Bochud-Allemann, N., Martinou, J. C., Hemphill, A., & Schneider, A. (2004). Temporal dissection of Bax-induced events leading to fission of the single mitochondrion in Trypanosoma brucei. EMBO Rep, 5(3), 268-273.

Evans, D. A., & Ellis, D. S. (1983). Recent observations on the behaviour of certain trypanosomes within their insect hosts. Adv Parasitol, 22, 1-42.

Fairlamb, A. H., Blackburn, P., Ulrich, P., Chait, B. T., & Cerami, A. (1985). Trypanothione: a novel bis(glutathionyl)spermidine cofactor for glutathione reductase in trypanosomatids. Science, 227(4693), 1485-1487.

Fairlamb, A. H., & Cerami, A. (1985). Identification of a novel, thiol-containing co-factor essential for glutathione reductase enzyme activity in trypanosomatids. Mol Biochem Parasitol, 14(2), 187-198.

Fairlamb, A. H., & Cerami, A. (1992). Metabolism and functions of trypanothione in the Kinetoplastida. Annu Rev Microbiol, 46, 695-729.

Ferguson, M. A., Murray, P., Rutherford, H., & McConville, M. J. (1993). A simple purification of procyclic acidic repetitive protein and demonstration of a sialylated glycosyl-phosphatidylinositol membrane anchor. Biochem J, 291 (Pt 1), 51-55.

Ferguson, M. L., Torri, A. F., Perez-Morga, D., Ward, D. C., & Englund, P. T. (1994). Kinetoplast DNA replication: mechanistic differences between Trypanosoma brucei and Crithidia fasciculata. J Cell Biol, 126(3), 631-639.

Fevre, E. M., Coleman, P. G., Odiit, M., Magona, J. W., Welburn, S. C., & Woolhouse, M. E. (2001). The origins of a new Trypanosoma brucei rhodesiense sleeping sickness outbreak in eastern Uganda. Lancet, 358(9282), 625-628.

Flick, K., & Chen, Q. (2004). var genes, PfEMP1 and the human host. Mol Biochem Parasitol, 134(1), 3-9.

Fraidenraich, D., Pena, C., Isola, E. L., Lammel, E. M., Coso, O., Anel, A. D., Pongor, S., Baralle, F., Torres, H. N., & Flawia, M. M. (1993). Stimulation of Trypanosoma cruzi adenylyl cyclase by an alpha D-globin fragment from Triatoma hindgut: effect on differentiation of epimastigote to trypomastigote forms. Proc Natl Acad Sci U S A, 90(21), 10140-10144.

Francis, S. H., & Corbin, J. D. (1994). Structure and function of cyclic nucleotide-dependent protein kinases. Annu Rev Physiol, 56, 237-272.

Franck, E., Madsen, O., van Rheede, T., Ricard, G., Huynen, M. A., & de Jong, W. W. (2004). Evolutionary diversity of vertebrate small heat shock proteins. J Mol Evol, 59(6), 792-805.

Gao, Z. H., Metherall, J., & Virshup, D. M. (2000). Identification of casein kinase I substrates by in vitro expression cloning screening. Biochem Biophys Res Commun, 268(2), 562-566.

Geigel, L. F., & Leon, L. L. (2003). Cyclic 3'-5' guanosine monophosphate-dependent activity in Leishmania amazonensis. Mem Inst Oswaldo Cruz, 98(4), 499-500.

Ghiotto, V., Brun, R., Jenni, L., & Hecker, H. (1979). Trypanosoma brucei: morphometric changes and loss of infectivity during transformation of bloodstream forms to procyclic culture forms in vitro. Exp Parasitol, 48(3), 447-456.

Gibson, W., & Bailey, M. (2003). The development of Trypanosoma brucei within the tsetse fly midgut observed using green fluorescent trypanosomes. Kinetoplastid Biol Dis, 2(1), 1.

Gibson, W. C., de, C. M. T. F., & Godfrey, D. G. (1980). Numerical analysis of enzyme polymorphism: a new approach to the epidemiology and taxonomy of trypanosomes of the subgenus Trypanozoon. Adv Parasitol, 18, 175-246.

Gilman, A. G. (1987). G proteins: transducers of receptor-generated signals. Annu Rev Biochem, 56, 615-649.

Goldberg, B., Rattendi, D., Lloyd, D., Yarlett, N., & Bacchi, C. J. (1999). Kinetics of S-adenosylmethionine cellular transport and protein methylation in Trypanosoma brucei brucei and Trypanosoma brucei rhodesiense. Arch Biochem Biophys, 364(1), 13-18.

Graham, S. V. (1995). Mechanisms of stage-regulated gene expression in kinetoplastida. Parasitol Today, 11(6), 217-223.

Grubhoffer, L., Hypsa, V., & Volf, P. (1997). Lectins (hemagglutinins) in the gut of the important disease vectors. Parasite, 4(3), 203-216.

Gull, K. (2000). Genomics and post-genomics in parasitology: genome babble or a real opportunity? Biochem Soc Trans, 28(5), 541-544.

Gull, K. (2001). The biology of kinetoplastid parasites: insights and challenges from genomics and post-genomics. Int J Parasitol, 31(5-6), 443-452.

Halbig, K., De Nova-Ocampo, M., & Cruz-Reyes, J. (2004). Complete cycles of bloodstream trypanosome RNA editing in vitro. Rna, 10(6), 914-920.

Ham, P. J., Phiri, J. S., & Nolan, G. P. (1991). Effect of N-acetyl-D-glucosamine on the migration of Brugia pahangi microfilariae into the haemocoel of Aedes aegypti. Med Vet Entomol, 5(4), 485-493.

Hao, Z., Kasumba, I., Lehane, M. J., Gibson, W. C., Kwon, J., & Aksoy, S. (2001). Tsetse immune responses and trypanosome transmission: implications for the development of tsetse-based strategies to reduce trypanosomiasis. Proc Natl Acad Sci U S A, 98(22), 12648-12653.

Harteneck, C., Koesling, D., Soling, A., Schultz, G., & Bohme, E. (1990). Expression of soluble guanylyl cyclase. Catalytic activity requires two enzyme subunits. FEBS Lett, 272(1-2), 221-223.

Havercroft, J. C., Huggins, M. C., Nene, V., Dunne, D. W., Richardson, B. A., Taylor, D. W., & Butterworth, A. E. (1988). Cloning of the gene encoding a 50 kilodalton potential surface antigen of Schistosoma mansoni. Mol Biochem Parasitol, 30(1), 83-88.

Hayaishi, O., & Matsumura, H. (1995). Prostaglandins and sleep. Adv Neuroimmunol, 5(2), 211-216.

Hendriks, E., van Deursen, F. J., Wilson, J., Sarkar, M., Timms, M., & Matthews, K. R. (2000). Life-cycle differentiation in Trypanosoma brucei: molecules and mutants. Biochem Soc Trans, 28(5), 531-536.

Hieter, P., & Boguski, M. (1997). Functional genomics: it's all how you read it. Science, 278(5338), 601-602.

Hofmann, F., Dostmann, W., Keilbach, A., Landgraf, W., & Ruth, P. (1992). Structure and physiological role of cGMP-dependent protein kinase. Biochim Biophys Acta, 1135(1), 51-60.

Hunt, M., Brun, R., & Kohler, P. (1994). Studies on compounds promoting the in vitro transformation of Trypanosoma brucei from bloodstream to procyclic forms. Parasitol Res, 80(7), 600-606.

Inge-Vechtomov, S., Zhouravleva, G., & Philippe, M. (2003). Eukaryotic release factors (eRFs) history. Biol Cell, 95(3-4), 195-209.

Iten, M., Matovu, E., Brun, R., & Kaminsky, R. (1995). Innate lack of susceptibility of Ugandan Trypanosoma brucei rhodesiense to DL-alpha-difluoromethylornithine (DFMO). Trop Med Parasitol, 46(3), 190-194.

Jackson, L. K., Goldsmith, E. J., & Phillips, M. A. (2003). X-ray structure determination of Trypanosoma brucei ornithine decarboxylase bound to D-ornithine and to G418: insights into substrate binding and ODC conformational flexibility. J Biol Chem, 278(24), 22037-22043.

Jankiewicz, U., & Bielawski, W. (2003). The properties and functions of bacterial aminopeptidases. Acta Microbiol Pol, 52(3), 217-231.

Jayakumar, P. C., Musande, V. V., Shouche, Y. S., & Patole, M. S. (2004). The Selenophosphate synthetase gene from Leishmania major. DNA Seq, 15(1), 66-70.

Kamakaka, R. T., & Rine, J. (1998). Sir- and silencer-independent disruption of silencing in Saccharomyces by Sas10p. Genetics, 149(2), 903-914.

Kamisaki, Y., Saheki, S., Nakane, M., Palmieri, J. A., Kuno, T., Chang, B. Y., Waldman, S. A., & Murad, F. (1986). Soluble guanylate cyclase from rat lung exists as a heterodimer. J Biol Chem, 261(16), 7236-7241.

Kantorow, M., & Piatigorsky, J. (1994). Alpha-crystallin/small heat shock protein has autokinase activity. Proc Natl Acad Sci U S A, 91(8), 3112-3116.

Kawamoto, F., Alejo-Blanco, R., Fleck, S. L., Kawamoto, Y., & Sinden, R. E. (1990). Possible roles of Ca2+ and cGMP as mediators of the exflagellation of Plasmodium berghei and Plasmodium falciparum. Mol Biochem Parasitol, 42(1), 101-108.

Kim, Y. M., Talanian, R. V., & Billiar, T. R. (1997). Nitric oxide inhibits apoptosis by preventing increases in caspase-3-like activity via two distinct mechanisms. J Biol Chem, 272(49), 31138-31148.

Klemenz, R., Frohli, E., Steiger, R. H., Schafer, R., & Aoyama, A. (1991). Alpha B-crystallin is a small heat shock protein. Proc Natl Acad Sci U S A, 88(9), 3652-3656.

Koesling, D., Russwurm, M., Mergia, E., Mullershausen, F., & Friebe, A. (2004). Nitric oxide-sensitive guanylyl cyclase: structure and regulation. Neurochem Int, 45(6), 813-819.

Krauth-Siegel, R. L., & Ludemann, H. (1996). Reduction of dehydroascorbate by trypanothione. Mol Biochem Parasitol, 80(2), 203-208.

Kubata, B. K., Duszenko, M., Kabututu, Z., Rawer, M., Szallies, A., Fujimori, K., Inui, T., Nozaki, T., Yamashita, K., Horii, T., Urade, Y., & Hayaishi, O. (2000). Identification of a novel prostaglandin f(2alpha) synthase in Trypanosoma brucei. J Exp Med, 192(9), 1327-1338.

Lacourciere, G. M. (1999). Biosynthesis of selenophosphate. Biofactors, 10(2-3), 237-244.

Lamb, J. R., Tugendreich, S., & Hieter, P. (1995). Tetratrico peptide repeat interactions: to TPR or not to TPR? Trends Biochem Sci, 20(7), 257-259.

Lamour, N., Riviere, L., Coustou, V., Coombs, G. H., Barrett, M. P., & Bringaud, F. (2005). Proline metabolism in procyclic Trypanosoma brucei is down-regulated in the presence of glucose. J Biol Chem, 280(12), 11902-11910.

Lazar, T., Gotte, M., & Gallwitz, D. (1997). Vesicular transport: how many Ypt/Rab-GTPases make a eukaryotic cell? Trends Biochem Sci, 22(12), 468-472.

Lee, L. G., Connell, C. R., & Bloch, W. (1993). Allelic discrimination by nick-translation PCR with fluorogenic probes. Nucleic Acids Res, 21(16), 3761-3766.

Lehane, M. J., Aksoy, S., Gibson, W., Kerhornou, A., Berriman, M., Hamilton, J., Soares, M. B., Bonaldo, M. F., Lehane, S., & Hall, N. (2003). Adult midgut expressed sequence tags from the tsetse fly Glossina morsitans morsitans and expression analysis of putative immune response genes. Genome Biol, 4(10), R63.

Lehane, M. J., Aksoy, S., & Levashina, E. (2004). Immune responses and parasite transmission in blood-feeding insects. Trends Parasitol, 20(9), 433-439.

Lenardo, M. J., Rice-Ficht, A. C., Kelly, G., Esser, K. M., & Donelson, J. E. (1984). Characterization of the genes specifying two metacyclic variable antigen types in Trypanosoma brucei rhodesiense. Proc Natl Acad Sci U S A, 81(21), 6642-6646.

Liang, P., & Pardee, A. B. (1992). Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. Science, 257(5072), 967-971.

Lincoln, T. M., & Cornwell, T. L. (1993). Intracellular cyclic GMP receptor proteins. Faseb J, 7(2), 328-338.

Livak, K. J., Flood, S. J., Marmaro, J., Giusti, W., & Deetz, K. (1995). Oligonucleotides with fluorescent dyes at opposite ends provide a quenched probe system useful for detecting PCR product and nucleic acid hybridization. PCR Methods Appl, 4(6), 357-362.

Lowenberger, C. A., Ferdig, M. T., Bulet, P., Khalili, S., Hoffmann, J. A., & Christensen, B. M. (1996). Aedes aegypti: induced antibacterial proteins reduce the establishment and development of Brugia malayi. Exp Parasitol, 83(2), 191-201.

Ludemann, H., Dormeyer, M., Sticherling, C., Stallmann, D., Follmann, H., & Krauth-Siegel, R. L. (1998). Trypanosoma brucei tryparedoxin, a thioredoxin-like protein in African trypanosomes. FEBS Lett, 431(3), 381-385.

Lukes, J., Jirku, M., Dolezel, D., Kral'ova, I., Hollar, L., & Maslov, D. A. (1997). Analysis of ribosomal RNA genes suggests that trypanosomes are monophyletic. J Mol Evol, 44(5), 521-527.

MacMicking, J., Xie, Q. W., & Nathan, C. (1997). Nitric oxide and macrophage function. Annu Rev Immunol, 15, 323-350.

Madison-Antenucci, S., Grams, J., & Hajduk, S. L. (2002). Editing machines: the complexities of trypanosome RNA editing. Cell, 108(4), 435-438.

Marten, N. W., Burke, E. J., Hayden, J. M., & Straus, D. S. (1994). Effect of amino acid limitation on the expression of 19 genes in rat hepatoma cells. Faseb J, 8(8), 538-544.

Martin, T. E., Moller, A. P., Merino, S., & Clobert, J. (2001). Does clutch size evolve in response to parasites and immunocompetence? Proc Natl Acad Sci U S A, 98(4), 2071-2076.

Maslov, D. A., Podlipaev, S. A., & Lukes, J. (2001). Phylogeny of the kinetoplastida: taxonomic problems and insights into the evolution of parasitism. Mem Inst Oswaldo Cruz, 96(3), 397-402.

Matthews, K. R., Tschudi, C., & Ullu, E. (1994). A common pyrimidine-rich motif governs trans-splicing and polyadenylation of tubulin polycistronic pre-mRNA in trypanosomes. Genes Dev, 8(4), 491-501.

Maudlin, I., Dukes, P., Luckins, A. G., & Hudson, K. M. (1986). Extrachromosomal inheritance of susceptibility to trypanosome infection in tsetse flies. II. Susceptibility of selected lines of Glossina morsitans morsitans to different stocks and species of trypanosome. Ann Trop Med Parasitol, 80(1), 97-105.

Maudlin, I., Kabayo, J. P., Flood, M. E., & Evans, D. A. (1984). Serum factors and the maturation of Trypanosoma congolense infections in Glossina morsitans. Z Parasitenkd, 70(1), 11-19.

Maudlin, I., & Welburn, S. C. (1987). Lectin mediated establishment of midgut infections of Trypanosoma congolense and Trypanosoma brucei in Glossina morsitans. Trop Med Parasitol, 38(3), 167-170.

Maudlin, I., & Welburn, S. C. (1994). Maturation of trypanosome infections in tsetse. Exp Parasitol, 79(2), 202-205.

Maurice, D. H., & Haslam, R. J. (1990). Molecular basis of the synergistic inhibition of platelet function by nitrovasodilators and activators of adenylate cyclase: inhibition of cyclic AMP breakdown by cyclic GMP. Mol Pharmacol, 37(5), 671-681.

McAllister-Lucas, L. M., Sonnenburg, W. K., Kadlecek, A., Seger, D., Trong, H. L., Colbran, J. L., Thomas, M. K., Walsh, K. A., Francis, S. H., Corbin, J. D., & et al. (1993). The structure of a bovine lung cGMP-binding, cGMP-specific phosphodiesterase deduced from a cDNA clone. J Biol Chem, 268(30), 22863-22873.

McCulloch, R., & Barry, J. D. (1999). A role for RAD51 and homologous recombination in Trypanosoma brucei antigenic variation. Genes Dev, 13(21), 2875-2888.

McKnight, G. S., Clegg, C. H., Uhler, M. D., Chrivia, J. C., Cadd, G. G., Correll, L. A., & Otten, A. D. (1988). Analysis of the cAMP-dependent protein kinase system using molecular genetic approaches. Recent Prog Horm Res, 44, 307-335.

Mehlert, A., Zitzmann, N., Richardson, J. M., Treumann, A., & Ferguson, M. A. (1998). The glycosylation of the variant surface glycoproteins and procyclic acidic repetitive proteins of Trypanosoma brucei. Mol Biochem Parasitol, 91(1), 145-152.

Mehlitz, D., Brinkmann, U., & Haller, L. (1981). Epidemiological studies on the animal reservoir of gambiense sleeping sickness. Part I. Review of literature and description of the study areas. Tropenmed Parasitol, 32(3), 129-133.

Mehlitz, D., Zillmann, U., Scott, C. M., & Godfrey, D. G. (1982). Epidemiological studies on the animal reservoir of Gambiense sleeping sickness. Part III. Characterization of trypanozoon stocks by isoenzymes and sensitivity to human serum. Tropenmed Parasitol, 33(2), 113-118.

Melville, S. E., Leech, V., Navarro, M., & Cross, G. A. (2000). The molecular karyotype of the megabase chromosomes of Trypanosoma brucei stock 427. Mol Biochem Parasitol, 111(2), 261-273.

Meyer, S., Temme, C., & Wahle, E. (2004). Messenger RNA turnover in eukaryotes: pathways and enzymes. Crit Rev Biochem Mol Biol, 39(4), 197-216.

Michalski, M. L., & Weil, G. J. (1999). Gender-specific gene expression in Brugia malayi. Mol Biochem Parasitol, 104(2), 247-257.

Morris, S. M., Jr., Kepka-Lenhart, D., & Chen, L. C. (1998). Differential regulation of arginases and inducible nitric oxide synthase in murine macrophage cells. Am J Physiol, 275(5 Pt 1), E740-747.

Morrison, T. B., Weis, J. J., & Wittwer, C. T. (1998). Quantification of low-copy transcripts by continuous SYBR Green I monitoring during amplification. Biotechniques, 24(6), 954-958, 960, 962.

Mowatt, M. R., & Clayton, C. E. (1987). Developmental regulation of a novel repetitive protein of Trypanosoma brucei. Mol Cell Biol, 7(8), 2838-2844.

Mshelbwala, A. S. (1972). Trypanosoma brucei infection in the haemocoel of tsetse flies. Trans R Soc Trop Med Hyg, 66(4), 637-643.

Murphy, N. B., & Pelle, R. (1994). The use of arbitrary primers and the RADES method for the rapid identification of developmentally regulated genes in trypanosomes. Gene, 141(1), 53-61.

Mutharia, L. M., & Pearson, T. W. (1987). Surface carbohydrates of procyclic forms of African trypanosomes studied using fluorescence activated cell sorter analysis and agglutination with lectins. Mol Biochem Parasitol, 23(2), 165-172.

Naula, C., & Seebeck, T. (2000). Cyclic AMP signaling in trypanosomatids. Parasitol Today, 16(1), 35-38.

Nelson, D. R., Lawson, J. E., Klingenberg, M., & Douglas, M. G. (1993). Site-directed mutagenesis of the yeast mitochondrial ADP/ATP translocator. Six arginines and one lysine are essential. J Mol Biol, 230(4), 1159-1170.

Nene, V., Dunne, D. W., Johnson, K. S., Taylor, D. W., & Cordingley, J. S. (1986). Sequence and expression of a major egg antigen from Schistosoma mansoni. Homologies to heat shock proteins and alpha-crystallins. Mol Biochem Parasitol, 21(2), 179-188.

Nerland, A. H., Mustafa, A. S., Sweetser, D., Godal, T., & Young, R. A. (1988). A protein antigen of Mycobacterium leprae is related to a family of small heat shock proteins. J Bacteriol, 170(12), 5919-5921.

Newton, R. P., & Smith, C. J. (2004). Cyclic nucleotides. Phytochemistry, 65(17), 2423-2437.

Novick, P., & Zerial, M. (1997). The diversity of Rab proteins in vesicle transport. Curr Opin Cell Biol, 9(4), 496-504.

Nunes, L. R., Carvalho, M. R., Shakarian, A. M., & Buck, G. A. (1997). The transcription promoter of the spliced leader gene from Trypanosoma cruzi. Gene, 188(2), 157-168.

O'Beirne, C., Lowry, C. M., & Voorheis, H. P. (1998). Both IgM and IgG anti-VSG antibodies initiate a cycle of aggregation-disaggregation of bloodstream forms of Trypanosoma brucei without damage to the parasite. Mol Biochem Parasitol, 91(1), 165-193.

Odiit, M., Kansiime, F., & Enyaru, J. C. (1997). Duration of symptoms and case fatality of sleeping sickness caused by Trypanosoma brucei rhodesiense in Tororo, Uganda. East Afr Med J, 74(12), 792-795.

Oduol, F., Xu, J., Niare, O., Natarajan, R., & Vernick, K. D. (2000). Genes identified by an expression screen of the vector mosquito Anopheles gambiae display differential molecular immune response to malaria parasites and bacteria. Proc Natl Acad Sci U S A, 97(21), 11397-11402.

Otieno, L. H. (1973). Letter: Trypanosoma (Trypanozoon) brucei in the haemolymph of experimentally infected young Glossina morsitans. Trans R Soc Trop Med Hyg, 67(6), 886-887.

Paindavoine, P., Pays, E., Laurent, M., Geltmeyer, Y., Le Ray, D., Mehlitz, D., & Steinert, M. (1986). The use of DNA hybridization and numerical taxonomy in determining relationships between Trypanosoma brucei stocks and subspecies. Parasitology, 92 (Pt 1), 31-50.

Paindavoine, P., Rolin, S., Van Assel, S., Geuskens, M., Jauniaux, J. C., Dinsart, C., Huet, G., & Pays, E. (1992). A gene from the variant surface glycoprotein expression site encodes one of several transmembrane adenylate cyclases located on the flagellum of Trypanosoma brucei. Mol Cell Biol, 12(3), 1218-1225.

Palmer, J. L., & Abeles, R. H. (1979). The mechanism of action of S-adenosylhomocysteinase. J Biol Chem, 254(4), 1217-1226.

Pao, S. S., Paulsen, I. T., & Saier, M. H., Jr. (1998). Major facilitator superfamily. Microbiol Mol Biol Rev, 62(1), 1-34.

Parsons, M., Nelson, R. G., Watkins, K. P., & Agabian, N. (1984). Trypanosome mRNAs share a common 5' spliced leader sequence. Cell, 38(1), 309-316.

Paveto, C., Pereira, C., Espinosa, J., Montagna, A. E., Farber, M., Esteva, M., Flawia, M. M., & Torres, H. N. (1995). The nitric oxide transduction pathway in Trypanosoma cruzi. J Biol Chem, 270(28), 16576-16579.

Pays, E., & Nolan, D. P. (1998). Expression and function of surface proteins in Trypanosoma brucei. Mol Biochem Parasitol, 91(1), 3-36.

Pearson, T. W., Beecroft, R. P., Welburn, S. C., Ruepp, S., Roditi, I., Hwa, K. Y., Englund, P. T., Wells, C. W., & Murphy, N. B. (2000). The major cell surface glycoprotein procyclin is a receptor for induction of a novel form of cell death in African trypanosomes in vitro. Mol Biochem Parasitol, 111(2), 333-349.

Pearson, W. R. (1997). Identifying distantly related protein sequences. Comput Appl Biosci, 13(4), 325-332.

Pedram, M., & Donelson, J. E. (1999). The anatomy and transcription of a monocistronic expression site for a metacyclic variant surface glycoprotein gene in Trypanosoma brucei. J Biol Chem, 274(24), 16876-16883.

Pereira, C., Paveto, C., Espinosa, J., Alonso, G., Flawia, M. M., & Torres, H. N. (1997). Control of Trypanosoma cruzi epimastigote motility through the nitric oxide pathway. J Eukaryot Microbiol, 44(2), 155-156.

Pereira, C. A., Alonso, G. D., Ivaldi, S., Silber, A., Alves, M. J., Bouvier, L. A., Flawia, M. M., & Torres, H. N. (2002). Arginine metabolism in Trypanosoma cruzi is coupled to parasite stage and replication. FEBS Lett, 526(1-3), 111-114.

Pfaffl, M. W. (2001). A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res, 29(9), e45.

Pfeffer, S. R. (1999). Motivating endosome motility. Nat Cell Biol, 1(6), E145-147.
Raeymaekers, L. (2000). Basic principles of quantitative PCR. Mol Biotechnol, 15(2), 115-122.

Rascon, A., Soderling, S. H., Schaefer, J. B., & Beavo, J. A. (2002). Cloning and characterization of a cAMP-specific phosphodiesterase (TbPDE2B) from Trypanosoma brucei. Proc Natl Acad Sci U S A, 99(7), 4714-4719.

Rasmussen, E. B., & Lis, J. T. (1993). In vivo transcriptional pausing and cap formation on three Drosophila heat shock genes. Proc Natl Acad Sci U S A, 90(17), 7923-7927.

Ravel, S., Grebaut, P., Mariani, C., Jamonneau, V., Cuisance, D., Gooding, R. H., & Cuny, G. (2004). Monitoring the susceptibility of Glossina palpalis gambiensis and G. morsitans morsitans to experimental infection with savannah-type Trypanosoma congolense, using the polymerase chain reaction. Ann Trop Med Parasitol, 98(1), 29-36.

Reed, S. L., Fierer, A. S., Goddard, D. R., Colmerauer, M. E., & Davis, C. E. (1985). Effect of theophylline on differentiation of Trypanosoma brucei. Infect Immun, 49(3), 844-847.

Reguera, R. M., Perez-Pertejo, Y., Ordonez, C., Cubria, J. C., Tekwani, B. L., Balana-Fouce, R., & Ordonez, D. (1999). S-adenosylmethionine synthesis in Leishmania infantum promastigotes. Cell Biol Int, 23(8), 579-583.

Reuner, B., Vassella, E., Yutzy, B., & Boshart, M. (1997). Cell density triggers slender to stumpy differentiation of Trypanosoma brucei bloodstream forms in culture. Mol Biochem Parasitol, 90(1), 269-280.

Richardson, J. P., Beecroft, R. P., Tolson, D. L., Liu, M. K., & Pearson, T. W. (1988). Procyclin: an unusual immunodominant glycoprotein surface antigen from the procyclic stage of African trypanosomes. Mol Biochem Parasitol, 31(3), 203-216.

Roberts, S. C., Tancer, M. J., Polinsky, M. R., Gibson, K. M., Heby, O., & Ullman, B. (2004). Arginase plays a pivotal role in polyamine precursor metabolism in Leishmania. Characterization of gene deletion mutants. J Biol Chem, 279(22), 23668-23678.

Robinson, N. P., Burman, N., Melville, S. E., & Barry, J. D. (1999). Predominance of duplicative VSG gene conversion in antigenic variation in African trypanosomes. Mol Cell Biol, 19(9), 5839-5846.

Roditi, I., Carrington, M., & Turner, M. (1987). Expression of a polypeptide containing a dipeptide repeat is confined to the insect stage of Trypanosoma brucei. Nature, 325(6101), 272-274.

Roditi, I., Dobbelaere, D., Williams, R. O., Masterson, W., Beecroft, R. P., Richardson, J. P., & Pearson, T. W. (1989). Expression of Trypanosoma brucei procyclin as a fusion protein in Escherichia coli. Mol Biochem Parasitol, 34(1), 35-43.

Roditi, I., & Pearson, T. W. (1990). The procyclin coat of African trypanosomes (or the not-so-naked trypanosome). Parasitol Today, 6(3), 79-82.

Rolin, S., Hanocq-Quertier, J., Paturiaux-Hanocq, F., Nolan, D., Salmon, D., Webb, H., Carrington, M., Voorheis, P., & Pays, E. (1996). Simultaneous but independent activation of adenylate cyclase and glycosylphosphatidylinositol-phospholipase C under stress conditions in Trypanosoma brucei. J Biol Chem, 271(18), 10844-10852.

Ruepp, S., Furger, A., Kurath, U., Renggli, C. K., Hemphill, A., Brun, R., & Roditi, I. (1997). Survival of Trypanosoma brucei in the tsetse fly is enhanced by the expression of specific forms of procyclin. J Cell Biol, 137(6), 1369-1379.

Rybalkin, S. D., Rybalkina, I., Beavo, J. A., & Bornfeldt, K. E. (2002). Cyclic nucleotide phosphodiesterase 1C promotes human arterial smooth muscle cell proliferation. Circ Res, 90(2), 151-157.

Saito, R. M., Elgort, M. G., & Campbell, D. A. (1994). A conserved upstream element is essential for transcription of the Leishmania tarentolae mini-exon gene. Embo J, 13(22), 5460-5469.

Salditt-Georgieff, M., Harpold, M., Chen-Kiang, S., & Darnell, J. E., Jr. (1980). The addition of 5' cap structures occurs early in hnRNA synthesis and prematurely terminated molecules are capped. Cell, 19(1), 69-78.

Sarkar, N. (1997). Polyadenylation of mRNA in prokaryotes. Annu Rev Biochem, 66, 173-197.

Satriano, J. (2003). Agmatine: at the crossroads of the arginine pathways. Ann N Y Acad Sci, 1009, 34-43.

Sax, C. M., & Piatigorsky, J. (1994). Expression of the alpha-crystallin/small heat-shock protein/molecular chaperone genes in the lens and other tissues. Adv Enzymol Relat Areas Mol Biol, 69, 155-201.

Seebeck, T., Gong, K., Kunz, S., Schaub, R., Shalaby, T., & Zoraghi, R. (2001). cAMP signalling in Trypanosoma brucei. Int J Parasitol, 31(5-6), 491-498.
Serrano, M. G., Nunes, L. R., Campaner, M., Buck, G. A., Camargo, E. P., & Teixeira, M. M. (1999). Trypanosomatidae: Phytomonas detection in plants and phytophagous insects by PCR amplification of a genus-specific sequence of the spliced leader gene. Exp Parasitol, 91(3), 268-279.

Shalaby, T., Liniger, M., & Seebeck, T. (2001). The regulatory subunit of a cGMP-regulated protein kinase A of Trypanosoma brucei. Eur J Biochem, 268(23), 6197-6206.

Simpson, L., & Maslov, D. A. (1994). Ancient origin of RNA editing in kinetoplastid protozoa. Curr Opin Genet Dev, 4(6), 887-894.

Simpson, L., Thiemann, O. H., Savill, N. J., Alfonzo, J. D., & Maslov, D. A. (2000). Evolution of RNA editing in trypanosome mitochondria. Proc Natl Acad Sci U S A, 97(13), 6986-6993.

Spadafora, C., Repetto, Y., Torres, C., Pino, L., Robello, C., Morello, A., Gamarro, F., & Castanys, S. (2002). Two casein kinase 1 isoforms are differentially expressed in Trypanosoma cruzi. Mol Biochem Parasitol, 124(1-2), 23-36.

Spielmann, T., & Beck, H. P. (2000). Analysis of stage-specific transcription in plasmodium falciparum reveals a set of genes exclusively transcribed in ring stage parasites. Mol Biochem Parasitol, 111(2), 453-458.

Srinivasan, P., Abraham, E. G., Ghosh, A. K., Valenzuela, J., Ribeiro, J. M., Dimopoulos, G., Kafatos, F. C., Adams, J. H., Fujioka, H., & Jacobs-Lorena, M. (2004). Analysis of the Plasmodium and Anopheles transcriptomes during oocyst differentiation. J Biol Chem, 279(7), 5581-5587.

Stevens, J. R., Noyes, H. A., Dover, G. A., & Gibson, W. C. (1999). The ancient and divergent origins of the human pathogenic trypanosomes, Trypanosoma brucei and T. cruzi. Parasitology, 118 (Pt 1), 107-116.

Stiles, J. K., Whittaker, J., Sarfo, B. Y., Thompson, W. E., Powell, M. D., & Bond, V. C. (2004). Trypanosome apoptotic factor mediates apoptosis in human brain vascular endothelial cells. Mol Biochem Parasitol, 133(2), 229-240.

Strader, C. D., Fong, T. M., Tota, M. R., Underwood, D., & Dixon, R. A. (1994). Structure and function of G protein-coupled receptors. Annu Rev Biochem, 63, 101-132.

Stryer, L. (1986). Cyclic GMP cascade of vision. Annu Rev Neurosci, 9, 87-119.
Sturm, N. R., & Simpson, L. (1990). Kinetoplast DNA minicircles encode guide RNAs for editing of cytochrome oxidase subunit III mRNA. Cell, 61(5), 879-884.

Sultan, A. A., Richardson, W. A., Alano, P., Arnot, D. E., & Doerig, C. (1994). Cloning and characterisation of a Plasmodium falciparum homologue of the Ran/TC4 signal transducing GTPase involved in cell cycle control. Mol Biochem Parasitol, 67(2), 339.
Tanaka, N., Nakanishi, M., Kusakabe, Y., Shiraiwa, K., Yabe, S., Ito, Y., Kitade, Y., & Nakamura, K. T. (2004). Crystal structure of S-adenosyl-L-homocysteine hydrolase from the human malaria parasite Plasmodium falciparum. J Mol Biol, 343(4), 1007-1017.

Tetley, L., & Vickerman, K. (1985). Differentiation in Trypanosoma brucei: host-parasite cell junctions and their persistence during acquisition of the variable antigen coat. J Cell Sci, 74, 1-19.

Treumann, A., Zitzmann, N., Hulsmeier, A., Prescott, A. R., Almond, A., Sheehan, J., & Ferguson, M. A. (1997). Structural characterisation of two forms of procyclic acidic repetitive protein expressed by procyclic forms of Trypanosoma brucei. J Mol Biol, 269(4), 529-547.

Turner, C. M., Barry, J. D., Maudlin, I., & Vickerman, K. (1988). An estimate of the size of the metacyclic variable antigen repertoire of Trypanosoma brucei rhodesiense. Parasitology, 97 (Pt 2), 269-276.

Ullu, E., Matthews, K. R., & Tschudi, C. (1993). Temporal order of RNA-processing reactions in trypanosomes: rapid trans splicing precedes polyadenylation of newly synthesized tubulin transcripts. Mol Cell Biol, 13(1), 720-725.

Ulschmid, J. K., Rahlfs, S., Schirmer, R. H., & Becker, K. (2004). Adenylate kinase and GTP:AMP phosphotransferase of the malarial parasite Plasmodium falciparum. Central players in cellular energy metabolism. Mol Biochem Parasitol, 136(2), 211-220.

Urakawa, T., Eshita, Y., & Majiwa, P. A. (1997). The primary structure of Trypanosoma (Nannomonas) congolese variant surface glycoproteins. Exp Parasitol, 85(3), 215-224.

Uren, A. G., O'Rourke, K., Aravind, L. A., Pisabarro, M. T., Seshagiri, S., Koonin, E. V., & Dixit, V. M. (2000). Identification of paracaspases and metacaspases: two ancient families of caspase-like proteins, one of which plays a key role in MALT lymphoma. Mol Cell, 6(4), 961-967.

Van Den Abbeele, J., Claes, Y., van Bockstaele, D., Le Ray, D., & Coosemans, M. (1999). Trypanosoma brucei spp. development in the tsetse fly: characterization of the post-mesocyclic stages in the foregut and proboscis. Parasitology, 118 (Pt 5), 469-478.

Van den Abbeele, J., Rolin, S., Claes, Y., Le Ray, D., Pays, E., & Coosemans, M. (1995). Trypanosoma brucei: stimulation of adenylate cyclase by proventriculus and esophagus tissue of the tsetse fly, Glossina morsitans morsitans. Exp Parasitol, 81(4), 618-620.

Van der Ploeg, L. H., Valerio, D., De Lange, T., Bernards, A., Borst, P., & Grosveld, F. G. (1982). An analysis of cosmid clones of nuclear DNA from Trypanosoma brucei shows that the genes for variant surface glycoproteins are clustered in the genome. Nucleic Acids Res, 10(19), 5905-5923.

Van Deursen, F. J., Thornton, D. J., & Matthews, K. R. (2003). A reproducible protocol for analysis of the proteome of Trypanosoma brucei by 2-dimensional gel electrophoresis. Mol Biochem Parasitol, 128(1), 107-110.

Vanhamme, L., & Pays, E. (1995). Control of gene expression in trypanosomes. Microbiol Rev, 59(2), 223-240.

Vassella, E., Den Abbeele, J. V., Butikofer, P., Renggli, C. K., Furger, A., Brun, R., & Roditi, I. (2000). A major surface glycoprotein of trypanosoma brucei is expressed transiently during development and can be regulated post-transcriptionally by glycerol or hypoxia. Genes Dev, 14(5), 615-626.

Vassella, E., Reuner, B., Yutzy, B., & Boshart, M. (1997). Differentiation of African trypanosomes is controlled by a density sensing mechanism which signals cell cycle arrest via the cAMP pathway. J Cell Sci, 110 (Pt 21), 2661-2671.

Vickerman, K. (1969). On the surface coat and flagellar adhesion in trypanosomes. J Cell Sci, 5(1), 163-193.

Vickerman, K. (1994). The evolutionary expansion of the trypanosomatid flagellates. Int J Parasitol, 24(8), 1317-1331.

Vincendeau, P., Gobert, A. P., Daulouede, S., Moynet, D., & Mossalayi, M. D. (2003). Arginases in parasitic diseases. Trends Parasitol, 19(1), 9-12.

Wahle, E. (1995). 3'-end cleavage and polyadenylation of mRNA precursors. Biochim Biophys Acta, 1261(2), 183-194.

Walker, J. E. (1992). The NADH:ubiquinone oxidoreductase (complex I) of respiratory chains. Q Rev Biophys, 25(3), 253-324.

Weiden, M., Osheim, Y. N., Beyer, A. L., & Van der Ploeg, L. H. (1991). Chromosome structure: DNA nucleotide sequence elements of a subset of the minichromosomes of the protozoan Trypanosoma brucei. Mol Cell Biol, 11(8), 3823-3834.

Welburn, S. C., Arnold, K., Maudlin, I., & Gooday, G. W. (1993). Rickettsia-like organisms and chitinase production in relation to transmission of trypanosomes by tsetse flies. Parasitology, 107 (Pt 2), 141-145.

Welburn, S. C., Lillico, S., & Murphy, N. B. (1999). Programmed cell death in procyclic form Trypanosoma brucei rhodesiense --identification of differentially expressed genes during con A induced death. Mem Inst Oswaldo Cruz, 94(2), 229-234.

Welburn, S. C., & Maudlin, I. (1987). A simple in vitro method for infecting tsetse with trypanosomes. Ann Trop Med Parasitol, 81(4), 453-455.

Welburn, S. C., & Maudlin, I. (1989). Lectin signalling of maturation of T. congolense infections in tsetse. Med Vet Entomol, 3(2), 141-145.

Welburn, S. C., & Maudlin, I. (1990). Haemolymph lectin and the maturation of trypanosome infections in tsetse. Med Vet Entomol, 4(1), 43-48.

Welburn, S. C., & Maudlin, I. (1997). Control of Trypanosoma brucei brucei infections in tsetse, Glossina morsitans. Med Vet Entomol, 11(3), 286-289.

Welburn, S. C., Maudlin, I., & Ellis, D. S. (1989). Rate of trypanosome killing by lectins in midguts of different species and strains of Glossina. Med Vet Entomol, 3(1), 77-82.

Welburn, S. C., Maudlin, I., & Molyneux, D. H. (1994). Midgut lectin activity and sugar specificity in teneral and fed tsetse. Med Vet Entomol, 8(1), 81-87.

Yang, H., Zhou, J., Ochs, R. L., Henning, D., Jin, R., & Valdez, B. C. (2003). Down-regulation of RNA helicase II/Gu results in the depletion of 18 and 28 S rRNAs in Xenopus oocyte. J Biol Chem, 278(40), 38847-38859.

Zantema, A., Verlaan-De Vries, M., Maasdam, D., Bol, S., & van der Eb, A. (1992). Heat shock protein 27 and alpha B-crystallin can form a complex, which dissociates by heat shock. J Biol Chem, 267(18), 12936-12941.

Appendix I

RADES contig information.

923-P_1

NTGTTGCGGTGCTCSYnCATTACAGATGTTTCTGCTGCACGTCAAGGTCACTTCACC
AACAAAACGTCWACAGCAACATAACCAAATGCATGATCGAGTGTCAATCAATCAATC
AATCGCTAGTAGTTGGGATTTACCTTCGTTACCAGTGCTTCCACTAGTGGAGAATCC
GGGTCGATTGGTGGAAAAATACTTTGCCTTCCCAACACTGTTCCAATGAGGAAAAGT
ACCACCAACTTCGTGAGGAAGCGAAGGTCCCAAATTAAATTCACTTTTGTGTACTGC
GCGAAGTACTCGCCGGAGATAATGTAGTGGCATTCGTCAAAGAAACGCACCACACGG
TAGTGGAAAAGTTGTCTCTGTTCAGCCAAGTACCGTTTCGTACTGCTCACCGCTGCG
TCAGTTCCGTTGGCGTTGTGTTCGCTGATGGTCTTTTGAATTCCCNCGTCTGNTAGN
GCGGCAAGTTTCTTTGCAGAACGCGCTTGAACCTTACTGGCCGCATTGCATANCGGC
GTCATGATCTTACAGGCAACGGNC

923-P_2

TNGTTGCGAACAGGSTAGACGCTGTTGCACGCCGTGGTTGAAACTCGCTACCGTGGA
AGTCATCTGCTCAGTGGAGGAATAAGAATGAGGTACCTCATTTTGAACTTCATTCCC
CAGCCCATCATCTATTCGCTGCACGCTAGCAGTTTCGCCCCGGCTGAGCGCGCTGTT
TGAACCATAACTCCGGTCAGGATAGACTATGCTCTCACCAATTAGTTCCTGAATGCG
TCTGACGCGTTCCCTCATCAAGTTGAGTGAATTCTGTCCCTCAGAAGCGCTGAACAG
GTTTTGAGGGGTACTTTCCACACTTTCTATGGCACTCGACAAATGTAAAGGCTGCGA
TACCTTTGTGGCTTTCAACCCGCCTGTCTGCACTTGGACTCTCCTCTCCAACGGGAG
AGACTGAGAGCCTACCACCAGCCCCTGCTCTGCGACAACGTGCACGTTCCCCCCGGC
TCGCGGTACGGAGGAAAAGTTAGAATGTGCAAGCTGCTCTCGGGTCCGCCGATCCCC
ATCAAAATTAACTTCCGCGAAGGGAAAGTGGCGAGGATCATCCGCATNTGTCGATGG
NCCTCCCCCACTAAAACGGNGACCATCCACACNAAGGGACGTGCCAGGGNGTCGCTG
CTTGGCAACTCTTGGTCCAGAGTCTTACGCACAAAGAACGGCCAAAGGNGATGAGGN
TTGAGCTACA

1051-P_1

TCKKCCGGTCATGTTCTGTTCGATAATCTGGGCGACGGTATTTTCCAGCGTTTGCGC
AGAGGCGCCGGGATAATTTAGAGGAAATACGGATCGTCGGCGGCGCCCGCCGGGGAT
ATTGCTCTATCGGCAGCGATATGATGGCCATCACGCCAGTCAGGCAAACAATAATTG
CCAGCACCCAGGAAAAAATAGGGCGATCAACAAAAAAATTAGCCATCCARTCATAAC

CTCAAGGCAATAGCTTATTGGGTCGGCGCTACGCGAACCTTCGCACCGGTGCCGGCT
TTGTTGAATAAATCAGAACTGTGACTACGCCCCCCTAGCAGATCGATTGTTATGCGA
TCCGGATGCTGTTCGGCATTCCTAACAGCGTCATTTTGTTAAGCGCTTTGACCATCG
CCATTGCCTCACCTACCTGCGCATCATAGTCACGCAGACTTAGATGATCGCCCACTT
GCTTTTTCCATACATCGTTACTGCCGCTTAGACGCTGATTCACAACGGCGTAGTTAC
GCTCAGACTTYGCCCGGTAAGTCATCAATGGATCGCGCTTGACCGGCCcGA

$1051\text{-}M_1$

TCGGTTCGGTCACATACAATACAACCATGTNCACTACTTACTTTATCCGGATCAGCC
AGCGCATATTTACAGCTGGAAGAAAAATACGGCGCACATAATTACCATCCCCTCCCC
GTTGTGCTCGAAAAAGGAGAAGGTGTTTTTCTCTGGGATGTAGACGGTAAACGTTAT
TACGATTTCCTGAGCGGTTATTCGGCCGTTAACCAGGGTCACTGTCATCCCCGAATC
GTTCGCGCCCTCATTGAACAGGCGCAGAAGCTTACGCTCACATCCCGTGCCTGACCG
GCCGA

$1078\text{-}P_1$

ANAGGAGAGGGCCGGCCCAGAGGGAGGGAACCCNAGGGCGGCCCACGGCAGGGNGGC
CACGNCGGCGGACGAAGGCNANAGAGACANAGCGACGGCGGNGGNGNNGAGNGGNNG
NGAGNGNGGGANNACGGCNGAGCGGGGGCGGGCGCGGNGANCNGGGCGCNANANGGN
CGAAGGGGGGGGAGACCGNGNCNCNCCNNCCNNNNAAACGCCGCNNGAANGAGGGCC
CGCGGCCGGGGCANNAGCCCGNCGCGGGGCGAGNGANACGCGGCGCAGCACAGNGNG
CGGGGGGGGAGNNGGCGNAGGCGGCAACCGCGGCCNCAGACGGAGGAGCGCGNGCAA
CAGGGCGCGGAG

$1078\text{-}P_2$

TTYYYCTTCGCAGCAAAACAATAAAAGTATCGAGTGCGCTTTGGTATTGTTTGAAGT
TTACATCTTTTTCTCCTTGCGCCCAAAGCACCGCAACAAATTTACTTTTCGGAAAAT
GCTTCATAACAAAATTCACACGGTTAAGCGTAGTGTTGTATAAGTAATTTCCTTTGT
TCCATCCGTCACCTTCAAAACCCGAGTTATCCATGGCACACGGTACAATTAAAATAT
CACGCTTATTGTTAATGTATTTCTGATATTCTTTTGCAAAGGTTAACGAAAACCCGA
TTCGTTCCGGTTTAGGGTTAAAATTATGAAGCGGCTCTTTTGCAGGTATGATTTTCA
CATCGCTTTCATTGCGTCCCAATTGGTAAATACCGGGCGAAGGGGAAATCGAANTC

$1348\text{-}M_1$

TGAWTGCTGCGTATTTGATGTTTAAGAGGAGTGTTACCACAAAGCTACTCTCATCAA
TGCTGCCACACTCCCTGCCCACCTAGTCAAAAAGACAAAGTGAATGTAGGATGAATG
AACAAATGAGTGAATGAGCACATTCTGCCCTGTCTCTTCTCTTTCCCTCACGCCTCA
CATCCGGCCTGTGAGGAGACCCTGCCACCTCCACCTCCAAAATACACCCAGAATCTG

ACCTCCCGCTCCGGCTCTGTTCCCACCGTGGTGTGGGCCACCCGCCCCACCACTCCT
GAGCTGTTGCCGCAGCCTTCA

1348-M$_2$

TCACAGNNTCCAAATCCTTGGTTAAGAGGAACCAAAGCAGAATGGCAACCATGATCA
CGAACGAAACAACCGTGGTAGTGGTATCCCTTTGCTCGCCATCCATGGGATTCTCAA
AGTAAAACGTGAGGACGAATATTTCGTCCTCGGGTTTAAGCACTATCTTGGTATCCT
CAGGGCTGAGGTGCACCGTGCAGTACTTACCGCGCAACCGCCAGTCCAACGCCTGCG
TACCGTCACGACTGTTTGCACTGCTGACCAACTTTACCACTTTCGAAAATACGGTTC
GATATACAATTTCCTCTGGAAGACAGGGNGCACACGCAAAGGGTTAAGCGATGACGA
TGGCCGGCAAATTCGTCAGCTGTCGAACTGACTTTGNTGNCCAACGAATTTCAAGCT
GGATCTCATTCCCAACACCCCTCTCCGTTTNGNCCAG

1539-M$_1$

CTCTCCCATATGTCATCTAACTGCAGGNCSGCCGCACTAGTGATTGGGTGGGCGTTG
TATATGATGCCCTCGCCGCAAACCCAGGCGACGCCGCTTGAACCCGGCACCGAAGAT
ATCTTGATGCGCGTGTCGCCGCCTGTGGTGCCGCCCGTCGAATTTGCGTTCACGGCG
TCACAGCGGGATGTCGACACCCTCTCTTTTGCACCTGTGGGATACTCGCGACAGCGT
AACGGCAGCCAGCTCTACGGCACGCTGACAATCCGATCTGGAGCACATGAGCGGCGG
GTCCGCATCAGCTCCGCAGGCCGCGCCCGgCATCTGCGATCCGGCCAAGGACACCAA
AACGTGCGGCGCCAGCGCCTCGGCTAACGACGATTAACCCAAACTCCCGCATTTTTT
GCTCATGTTCTCCGACTTCGACCACGCCATGATGCAGCGCGCCCTGGCGCTGGCCGA
GAAGGGCTTGTTCACCACCACGAA

1561-P$_1$

TGNTTGGGCGTGGTCGGTGACATCACCGCAGCATGGGCAGTGAGCGCAGGCTATACC
GTGATGAACACGAAGGTCACCAGTGGCGCAGTAGTTGCTGCAGACGGTACCCGCGGC
TTGAGCTACACACCAAAACAGGCGTTTACAGCATGGACTACCTACAAGCTGCCAGGC
GGCTTCACGATAGGTGGCGGTGCACGCTTTGTTGATAGTTTGCGTAAGCCTTCGGAT
GGTGCAGTTGGTACCCCGACACACACCGAGTCCTACTGGGTCTACGATGCAATGGCG
ACATACGCAGTTACGAAAAACTTCGACCTGCAGCTTAACCTCTACAACCTGACCGAT
AAAGAATACGTAGCGTCGATCAACAAGAGTGGTTATCGCTACACGCCAGGTATTGAG
CGCTCGGCACGTCTGACCGCAAATATCCGCTTCTAAGTTTATAAAAGTAGCCCGGAG
CTGAAGTTATACTGATAAGCCTCTCCTTAGGGAGAGGCTTTTTTCATTGAGGTAACC
ACCATGATGCTGCATATACCCGAAGTACTGACGCCCACCCA

Appendix II

Subtractive cDNA library contig information

Contig 1

TAGCGTGGTCGCGGCCGAGGTACAAACGGCCGCCGACGACACCGTGGCGAGCGTCAG
GGCGAAGGCAGGTCTAGGTGAGAATGATCTCCTCTTCTACGGCGGTGAGTGCCTGTG
CGACGGTGCATTGCTGTGCGACTACGCGCTGAAGCGGGAGAGCACCATCCAGGCGAT
GCTCCCGGTGGAGGGTGGTAAGGGCAAGAAAAAGAAGAAGAAGGTTTTCACGAAGCC
GAAGAAACCCATTCACCGTCACAAACTCGAGAAGATGCGTGCGTTGAAGTACCTGCT
AGCGTGGTCGCGGCCGAGGTACGATATTTGCTTCCGCACCCTGAAACTGACAACACC
AACGTTCGGTGACCTGAACCACTTGGTGTCTGCTGTTGTGTCCGGCGTCACCTGCTG
CCTGCGCTTCCCTGGTCAGTTGAACTCTGACCTCCGTAAGTTGGCTGTGAACCTTGT
CCCATTCCCGCGTCTGCACTTCTTCATGATGGGCTTCGCCCCGCTGACCAGCCGCGG
CTCGCAGCAGTACCTGCCCGGGCGGCCGCTCGAAcGCCCGGGCAGGTACTTGCCGTG
GCGGGGGTCGCACTTTGTCATCATGGAGGCGGGCTCAAACACAGCGTTCGAGATCTC
AGAGACAGAGAGTTGCTCGTGGTAGGCCTTCTCTGCGGAGATGACTGGTGCATAGCT
TGTCAGCACGAAGTGGATACGTGGGTACCTCGGCCGCGACCACGCTA

Contig 2

TTMGMYGGYCGccCGGSCAGGTACTATATTGGGTGTGAGAAACAAATGACGCACTC
CACTGATCTTCAGTGGCTGTTGGTTCGCCAGAACAGTAAATTTCTGCAGAAGCGGAA
TGGCATTCGATTGAGCAGTGATCCGTTCAATAACAACGCGAATTGGACAAAGCGACA
GTCTGGTTTCCTGAACACGAAGGCCGCAGTGATAAAGACGAAGGGTGATCGCATCCT
CCTGACAACAAAGAGCGGTGATACAAACAACAAGCCCAAGTTGATGTATAAGAAGAC
TGTGATGGAACCTGGCGTGAAGTCATCAGTGGTGAAGAGAGCTGTTGCCGACATCCG
CCCCGACCTTGCTAAGATGGCATACCGTCGCGCGCGCAAGATGGCATGTACCTcGSC
CggGCGRCCRCBCKARCAGGTACTGCTGCGAGCCGCGGCTGGTCAGCGGGGCGAAGC
CCATCATGAAGAAGTGCAGACGCGGGAATGGGACAAGGTTCACAGCCAACTTACGGA
GGTCAGAGTTCAACTGACCAGGGAAGCGCAGGCAGCAGGTGACGCCGGACACAACAG
CAGACACCAAGTGGTTCAGGTCACCGAACGTTGGTGTTGTCAGTTTCAGGGTGCGGA
AGCAAATATCGTACCTGCCCGGGCGGCCSSTCGAAAA

Contig 3

GAATTCACTAGTGATTAGCGTGGTCGCGGCCGAGGTACATCTTTGATAAGATGGGAG
CACACACGCCTTCGAGCTCCTTCTGGCGGTGGTTGTATTCATCCAAACTGGCCTCCT
GGTTGTCGTTCAGCCATCGCAGTGCCTCCTCCACGGCAGTGGTGACAGCATTCTTGT
CGGCGTCGTCAAGCTTGCCAGCAACGTTGGGGTCATTGATAGTGTTCTTCATCGAGA
AGGCGTAGTTCTCAAGACCGTTCTTAGCATCAATGCGCTCACGCTGCGCCTTATCCT
CAGCCTCATACTTCGCAGCATCAGACACCATGCGTTCGATATCAGCCTTGCTCAGGC
GACCCTTGTCGTTTGTGATGACAATCTGRTTGCGCTTGCCAGTGCCCTTCTCCTCCG
CAGAAACGCTAAGAATACCATTAGCGTCTAGGTCGAAGGTCACTTCTATCTGTGGCA
CGCCGCGGGGTGCCGGGGGGATTCCAGAGAGATCGAAGGTGCCCAAGAGATGGCAGT
CCTTCGTCATGGTACCTGCCCGGGCGGCCGCTCGAA

Contig 4

TAGCGTGGTCGCGGCCGAGGTACCGCAGGCCYGCGACCGTGAAACCCAGCAGCAACG
TAAGTGCCATTAAGAATAAGTGGGACGCGTTCCGGATCATCCGCTACCCCCTCACCA
CTGACAAGGCGATGAAGAAGATTGAGGAGAACAACACCCTAACCTTCATTGTGGATT
CTCGTGCTAACAAGACTGAGATCAAGAAGGCGATTCGCAAATTGTATCAGGTTAAGA
CGGTGAAGGTGAACACACTCATTCGACCTGACGGCTTGAAGAAGGCCTATATTCGCC
TGTCTGCGTCATACGATGCGCTCGACACGGCAAACAAGATGGGTCTTGTTTAGTGTG
GTGGATCTGTAGGGGTTTCTTACTTCCTCAAGCAACTTTGTAGGGTACCTGCCCGGG
CGGCCGCTCGAAWCACTAGTGAATTCGCGGCCGGCCTGCAGGTCGAA

Contig 5

TAGYTAGCGTGGTCGCGGCCgAGGTACCGCCAGCTGTTCCACCCCGAGCAGCTGATC
TCCGGCAAGGAGGATGCGGCCAACAACTACGCTCGTGGCCACTACACCATTGGTAAG
GAGATCGTCGACCTCTGCCTGGACCGCATCCGCAAGCTCGCTGACAACTGCACTGGT
CTTCAGGGCTTCCTCGTGTATCACGCCGTCGGCGGTGGCACTGGTTCTGGCCTGGGT
GCGCTGCTCTTGGAGCGCCTCTCCGTGGACTATGGCAAGAAGTCCAAGCTCGGCTAC
ACGGTGTATCCATCACCGCAGGTGTCGACGGCTGTCGTGGAGCCCTACAACTCTGTG
CTCTCGACACACTCACTTCTGGAGCACACCGATGTTGCTGCGATGCTTGACAATGAA
GCAATTTATGATTTGACTCGCCGCAACCTCGATATTGAGCGCCCCACGTACCTcGSC
CGCGRCCRCKCKAAA

Contig 6

TAGCGTGGTCGCGGCCGAGGTACGCCACCTTCGCCCGCTTCATTTTTGTTCTGCCCT
TCATGTTATGCATTTACCAATACATCCCTGGCCATGTCGGTCTGTATATCTTCTCGT

TTCTCCTTGGCCTGACCAACTGTGTGGGTGCCATGTCGATGGTGTATGGCCCAATAA

CCCCTGGTCTTGAGACTGCGGGTCAGAAATTGATGGCTGGACAGTTGATGGGGATTT

CGCTGCTTTCTGGAATTGCCGCTGCATCTGTGCTTGCGATGATTGTGGTTGTCTTCC

TACCATGAGTACCTGCCCGGGCGGCCGCTCGAAMKSSMRKKCCCGGGCMGSYAYKGC

RRMYGGGMGCACGACTMCCCAATTGYCAGTATTGAAGACCCTTATGATCAAGACGAC

TTCGCTGGTTTCGCTGGCATCACTGAGGCTCTGAAGGGCAAAACTCAGATTGTCGGC

GACGACCTCACTGTCACCAACACCGAACGCATCAAGATGGCCATCGAAAAGAAGGCC

TGCAACTCTCTTCTACTGAAAATCAACCAAATCGGTACCTCGGCCGCGACCACGCTA

Contig 7

AATACGACTCACTATAGGGCGAATTGGGCCCGACGTCGCATGCTCCCGGCCGCCATG

GCGGCCGCGGGAATTCGATTAGCGTGGTCGCGGCCGAGGTACATGATGTCGAAGAAG

ACCATGGAGATCAACACCACGCACGCCATCGTCAAGGAGCTGAAACGACGCGTGGAG

GCCGACGAAAACGACAAGGCCGCCAAGGACCTGATATTCCTCCTCTTCGACACCTCG

CTGCTGACATCCGGGTTTACACTTGATGACCCCACGGCATACGCTGACCGTATCCAC

CGCATGATCAAATTGGGTCTCTCACTCGACGATGATGCGGAAGAGGAGGAAGCGCAG

GCACCCGTTGCCGCTGCTGCCGCCAACTCCAGTACCTGCCCGGGCGGCCGCTCGAA

Contig 8

TACCCCGTTTGCGCCACAGGTGTCATGGCCTGCTGGCCGAGTCGCAGGTTCACATCG

GTTACTCCAAGAGGAGCGCCTCCCTGGCGGCCAGACCGATTCTCTGCTGAATGCTGC

GCAGCTTCACCCATAAGTCGCTCCTGCTCCTCTCGAGTTACCTC

CCGCGTGGCAGTTCGGGTAAACATCATACTGCTACGTTGTGTGATCATGAGACGCTG

TACCTGCCCGGGCGGTCTAGCGTGGTCGCGGCCGAGGTACCTTCTCGAACCACTTCA

TGGCCTCCTCCTTACGGACACGGTGGTTGAAGCCCACACGGCTAACCCTGCGCTTCC

GGTGCGCCACGCGCTCACCGCGGCGGCCGAGAACGACGTAGAAGTCCATACCATAGA

TACCGGTAGACGGATCATATTTGATACCAAGATCAATATGCTCTGAGATACCGAAGC

CAAAGGAACCTGTGTCAGAGAAGTTGTAGCTCTTTAGCTCAAACTCCTTAACCTTCA

GACCCTTTTCCAGTAGTTCCTCAGCCTTCTTGCCACGCACAGTGCAGTGCACGGCGA

TCTTTTCGTTACGCCGGATACCAAACGTACCTGCCCGGGCGGCCGCTMGCGTGGTCG

CGGCCGAGGTACTATTTTCCTCCCTATACTCCCTCCCGGTTGCCTCGAAAACTCCCC

CTTTCTACACCGACGACCCACAGACATCTCCTCACGAGTCGTGTCCTATGCACTCTG

AGAAGAATAACGGTCATACAGTTTGGTTTGGTCATCCCGCGATATCTGAGTCCATCC

TTTCGGGTGAACATGATAAACAGTCACAATGCCACCAGAGGCACCGTCACGATACGT

CGCGTGGAAGATTGAGCGGCGCGCCAGGTCACATGCCTCCTCAACAGTGAGGTTCTT

CCGGTAGCCCTGGTCAAGCACACCATAGGCATAAATGGAACCGGATCCCACACTAAA

AATTTCATGCTTTACACGCGTGCCCTTGTCGTCCACGTAATAAAGGACGGCCCAAA
CTGATCCCAACCAGCCACCATTGTTCCCATAGATAGCCCGTGGTTGCGGTATTGGTA
GGTTATATTTGCCAGAATCTTGCTTGCGGCTGCTACAGAGATACGACAGTTGTTACG
TAGCTCCCACAGGCGGCACTCCATGCCGAGAACGCGCTCCCAGTACCKGCCCGGGCG
GCCGCTCGAA

Contig 9

TTTTCTAGCGTGGTCGCGGCCGAGGTACCCAACCTTCTTGATGTAAGCAGACACCTC
CTTCACAATCTCGTCATACCGCTCCTGTCCGTAGTTCACAGTCTTGTCGTCCATCTT
GTTGCAGCACACAACCATCTGCTTCACACCCAAAGTGAAGGCCAGCAACGCGTGCTC
GCGGGTCTGTCCATCCTTGGAGATACCAGCCTCGAACTCACCCTGCGCAGAGGCAAT
GATGAGGATGGCTGCGTCGGCTTGCGATGTGCCGGTGATCATGTTCTTGATGAAGTC
ACGGTGCCCAGGAGCATCAATAATAGTGAAGACAGACTTGGGTGACTCGAATTTCCA
CAGTGCAATGTCGATCGTGATACCACGTTCGCGCTCAGCCTTCAGCTTGTCCAGCAC
CCATGCGTACCTGCCCGGGCGGCCGCTCGAA

Contig 10

GTGWTTAGCGTGGTCGCGGCCGAGGTACCTCGTGGATCTTACTGAAGCTCTTATTCT
CATCCTCAGTTTGCCCAATGATGCGAAACTTGTCGGCCACAAGCTCCACAAGGTGCT
CCAACAGCTGCGCCTTAATGTAATACTGTTCCTCAGCAATCGCCATCTCACCATCGG
CAATGGCAACATTACGGATTTCGTTCGCCTTCTCAACGGCTGCAAGTTGGGCTTTGA
TCTGCTCCTCGTTCCCAAGAAGAGCCGTCTGCACAACAGCCACGTTCATCGTGTCCT
CAATATTCTTCAGTGCAGCCACAGGAACTTGCGCCATCTCAGAAGCGTCGAGCTGGG
ACCGGCGCAGTTCGTCTTCGATTTTCTCCAGTTTGGCGATGGCATTGTCGATCATGC
CGGTGAAAGACACCGTGCGAGTCTTGTTGTGATGTTGCTTCAGTTCGTTCATTGTCT
TCTTGAGTGCAACAATCTCCATGAACTGTGGCTTGTCCTCCGGCTTGTACCTGCCCG
ggCGGCCGCtCKAA

Contig 11

TAGCGTGGTCGCGGCCGAGGTACCCCACGACTACTTCCCCGTCTTGTCCCCAGGCAA
TTTCCTACTTTTCTTCCCTCTGTATTCCTTTGTTACTTCCAATGGTGTGCATCTCCC
CACCGCCCACCCCCACATTTCCTCCCTTTCGTAACAAGAAGTTTCCTCTAATTCAAA
TACCTAATCGGTTGATGGTTGTTTCGTCTACCACGTTGCGTCCAGAAACAAACGCGC
TGAACACCATCATTACGTATCAGATCCTACCCCCTTCCCCCGTTTGTGCCGCAGCGG
TGATTTTTACCTGCCCGGGCGGCCGCTCGAAATCGAATTCCCGCGGCCGCCATGGCG
GCCGGGAGCATGCGACGT

Contig 12

TAGYtGTGGTCGCGGCCgAGGTACCAAAGGCTGCTCACATGCCACCTTCCCCACTGG
CGAGTGCGTCCAGACTACCGGCGTCGGCTCAGTGATCGCCACGTGTGGCGCAAGCAA
CCTTACACAAATAATCTACCCACTAAGCAGGAGCTGCAGCGGTCCCTCTGTGCCGAT
TACTGTGCCACTGGATAAGTGCATACCCATTTTGATTGGGTCCGTTGAGTATCATTG
CTCCACCAACCCACCCACTRAGGCGGCCAGGCTGGTCCCACACCAGTGAGGTGGCGT
RGTCTTGGGTTGCGCTTCACCTTGTGATTGGTGTTTCTCGTTCATTATTGCTTGGtT
TTTTTTTtATTTCCTTCCCTTTTTACTCCCCCTCAATTCTATCTTCTTCGGCGAGCG
GCTGTGTAGTGGAACTGTAATACACCTGGAGGGCATGTGCGTGTGGTATGTGCATAG
CAGAATGTGCTGCGCTCCGGTACCTGCCCGGGCGGCCGCTCGAA

Contig 13

TAGCGTGGTCGCGGCCGAGGTACCGCGGTCTCTCCGTGCCCGAGCTAACGCAGCAGA
TGTTCGATGCGAAAAACATGATGCAAGCTGCAGATCCTCGTCACGGCCGCTACCTGA
CAGCGTCTGCACTCTTCCGCGGCCGCATGTCGACGAAGGAGGTTGATGAGCAGATGC
TGAACGTGCAGAACAAGAACTCGTCCTACTTCATTGAGTGGATCCCGAACAACATCA
AGTCCTCTGTTTGCGATATCCCACCCAAGGGACTCAAGATGGCTGTCACCTTCATTG
GCAACAACACCTGCATCCAGGAGATGTTCCGCCGTGTGGGAGAGCAGTTCACCCTCA
TGTTCCGTCGCAAGGCGTTCTTGCACTGGTACCTGCCCGGGCGGCCGCTCGAAATCA
CTAGTGAATTCGCGGCC

Contig 14

TAGCGTGGTCGCGGCCGAGGTACGTGCCAGTGCGCACTTCATCCACCACTGTTGGCT
CCAGGTCCAAGAACACCGCGCGGGGAACGTGCTTGCCAGCACCAGTCTCAGAGAAGA
AGGTGTTGAACGCATCATCCTCAACGCCAATCGTCTTGTCAGAGGGCATCGCACCAT
CGGGTTGAATGCCGTGTTCCAGGCAGAACAATTCCCAGCAGGCGTTACCAACCTGGC
AACCAGCCTGACCAATGTGGATGCAGATAGCCTCACGCATGATAAATAAATAGAAGT
GCTTTGTTGTTGTTGTTAGTGGTGCTAGTACCTGCCCGGGCGGCCGCTCGAA

Contig 15

TAGCGTGGTCGCGGCCGAGGTACTTCATGAACCGTGGTCGCCACTGTCTCTGCGTGT
ATGATGATCTCTCTAAACAGGCTGTCGCATACCGTCAGATTTCGCTGCTCCTTCGTC
GTCCACCGGGCCGTGAGGCATATCCGGGTGATGTCTTCTACCTGCATTCACGGCTCC
TGGAACGCGCGGCTATGCTTTCCCCTGGAAAGGGTGGTGGCTCCGTGACAGCGCTTC
CCATTGTGGAGACGCTGTCTAACGACGTGACTGCGTATATTGTGACGAATGTAATTT
CCATTACTGACGGTCAAATCTACCTCGATACGAAACTGTTTACTGGTGGCCAGCGTC
CTGCAGTGAACATCGGTTTATCTGTGTCACGTGTGGGCTCCTCCGCGCAGAACGCGG

CGATGAAAGGAGTTGCCKGAAAGCTGAAGGGGATTCTTGCGGAGTACCTcGSCCGCG
RCCRCKCKAA

Contig 16

TAGCGTGGTCGCGGCCGAGGTACTGCAAACCCTCTCACAGCAGAAGCAGCGGCAGCC
GCATCTTCACTCCCATCACGCAAGACCAAACTTCTTGAGAACCTCCACATTCTTCTG
CGGCATCTTCTTGAGGGCCTCCTTGCGGGCATTCTTCCAAGCGACGCGAACCTTTTC
CATCTTCTTGTGGTGACGCCCGGACTTTTCAATACGGCCTGCCTCAAGTTTCTCCAC
AACGTCGCTGTACCTGCCCGGGCGGCCGCTCGAAATCG

Contig 17

TAGCGTGGTCGCGGCCGAGGTGTAACAGCAAGGTTAATTATCGTTATTGAACACACC
AAACCTAAAAACAAGTAACACTACAACCATGGGCAAAGTAAGAGATACTCCGATGAT
CCTCATTGGCGTCCTCGGTCTCATCGGTGTCGTGGCCGCAGCCGCCGGCTCGCACTG
GCTGGACCCTAACCTTTCTGATAAGTTGAAGCTCGCATGGGACCACGCTGTCCAATT
TAACCTTATTCACACTGCCGCGATGATGGCTATCTTTGCTGCCCTCAAGATTGCCGA
GCCGGAAAGCAGCGCCGCCAGGTGGCTCAACCGCAGTTTCGTGTTATTGTTCGCCGG
AACAACGATCTTTTGCGGTACCTGCCCGGGCGGCCGCTCGAAATS

Contig 18

TAGCGTGGTCGCGGCCGAGGTACATGAGATCGAATTTGTGGTCAATACGAGCGAACA
CCTCTGCGATGGCCGTGGAGTTGGCGATCATGCATACCGCGCGCTGCACCTTGGCAA
GGTCACCACCTGGCACCACCGTGGGTGGCTGGTAGTTGATACCGCACTTGAAGCCTG
TGGGAGACCAGTCCACGAACTGAATCGTGCGCTTCGTCTTGATGGTCGCGACGGCAG
CATTCACATCCTTTGGCACAACGTCACCACGGTACCTGCCCGGGCGGCCGCTCGAA

Contig 19

TAGCGTGGTCGCGGCCGAGGTACGCCGTGATCTCCGCGGTAAGGGTGAGCTGGTGAT
GGGAAAGAAGACGTTGCAGAAGAAGATCGTGGAGAAGCGCGCTGAGGGCAACAAGGC
CACAGATGCTGATAAGCTGTTCCACCAGGTTTGCACCGATAAGCAGCTTCTGTGTGG
CAACACATCGAYGATATTCACGAACAGCGAAGTRKCtGWKRTCRCSaGCStGWtGGW
CAGCCACCGTGTTCAGGCGCCTGCGCGTGTTGGTGCYATTGCCCCATGTGATGTTAT
TGTGCCCGCCGGTAACACGGGTATGGAGCCGAAGGCAACGGCGTTCTTCCAAGCGTT
GAACATCGCGACGAAGATTTCCAAGGGTACCTGCCCGGGCGGCSRCWMGAAGGTACT
GAGCACTGGTGACAAGGTGGATAACTCCACGGCGACGCTGCTGCAGAAGCTCGACAT
CTCACCTTTCTATTACCAGGTCGAGGTGCAATCCGTGTGGGACCGCGGTGTGCTGTT
TACCCGTGAAGATCTCTCCGTCACCGATGCGGTTGTGGAGAAGTA

Contig 20

TAGCGTGGTCGCGGCCGAGGTACACGGACAATGTCATCCGTGTCTGGGGTGTATCGG
AGAACGCGTGAGGCAGCTAAAGGGCATGTGGTATAAGTGGTTTGATTTTGTGTTTTC
TGCGCCTTTTCCCGGGCGTGCATACAGTTGCACGTAACCTAGAAGGGTGTCGCTTGC
AATTTTCATTTGATCAGCAGTTGGAGCTGTCTTTATTTTTTCTTTTTCACCTAACAT
TTAATGGCCGCGCTTTTGGGTACCTGCCCGGGCGGCCGCTCGAA

Contig 21

TAGCGTGGTCGCGGCCGAGGTACGTGAAACAAGAGCATTTGCCCATTGGAGTGAGTG
AGATACCATGTGCAACCACGCCACTCCGATTGGTATACAGCTGCTGCTTCTACCGTC
CCTTTCTTTCCCCTCTCTTTGGCCCCTAATTCCTTCCTTCATCC
ATATGACTATAGCGCCGAGACAAATTCCGTCACAGCCCCTGCGGTCGTAAGCGTGAC
RCCGCTGATTCCACATCTGCAACCTCAGTTCTTGGAAAACTTCTTCCCCATCTTATC
GGCGGCGGACTTCGAGACGACCTGCAGCACGTTATAGCGGATTGYTTTGCTCAACGG
GCGGCATTGTCCAACCACCACCTCATCACCAGGCTTCGGATCGAAGCAGGGGCTGCA
GTGGACGGTAATATTGCGGTGGCGCTTCTGGTAACGCTGGTACCTGCCCGGGCGGCC
GCTCGAATCGAATTCCCGCGGCCGCCATGGCGGCCGGGAGCATGCGACGTCGGG

Contig 22

TAGCGTGGTCGCGGCCGAGGTACATAAAAATCAAATCTCTTACGTTTTGTAGCATTC
TCTCCCCCTCCACACATTACTGCTTAAATATCAAAGACGCAAAAGTGCTCAGCGGAT
GTTCCATTCTACACATAGGGGCTTTGCCATCCCTTCTTGTCCGTCACACATATATTC
TTAATCTGGTCTGCTAAGACAGTTTGTCGTAGCAATTGGTCATTCGAAACCAAAGTG
CATTACAAAACAAGAGTACCTGCCCGGGCGGCCGCTCGAA

Contig 23

TAGCGTGGTCGCGGCCGAGGTACGAAGGCGGCGAAGGGCGCGCAAGTTCGATGAGAG
ATCCTTGCCCTTGTTCTTGCGCTTGAATTCCTCGGTAAAGTGAGCCACAAGGCGATT
GTCGAAGTCTTCTCCACCAAGGTGGGTGTCACCATTTGTGGCTTTCACCTCGAAGAT
GCCGCCGTCAATCGTCAGCAATGTCACATCGAAGGTGCCACCACCAAGATCGAAGAT
GAGCACGTTGCGTTCCTTCCCCTCATCAGCCTTGCCCAGGCCATAGGCAATTGCAGC
AGCCGTTGGCTCGTTGATGATGCGCAGCACCTCCAGACCAGCAATCGTCCCCGCGTC
CTTCGTTGCCTGGCGCTGCGAGTCGTTGAAGTATGCAGGCACTGTTACGACAGCCTT
AGCCACCTGCTTCCCAAGATACGACTCCGCAACCTCCTTCATCTTCAATAGTACCTG
CCCGGGCGGCCGCTCGAA

Contig 24

TTCGKAGCGKSSKCSCGGSCAGGTACGGAAACCTGCGCGTGCTCAACTCCTATTGGG
TAAATGCGGACTCGACATTCCTGTGGTATGAAGTGGTGGCGGTGGACCCGATGCATC

GCACCATTCGCCGTGACCCGCGCATCAATTGGATCGTCAACGCTGTTCACAAGCACC
GTGAGCAACGTGGTCTTACATCTGCGGGTCGYAAGCACCGTGGGCTACGGCAGAAGG
GCCACAAGGCCTCCAAACTCCGCCCGTCTCGCCGTGCAGCGTGGCGCCGCAACAACC
GTATGGTGTTCCTCCGCAAGCGTTAGTCGTACCTcGGCCGCGRCCRCKCTAA

Contig 25

TAGCGTGGTCGCGGCCGAGGTACAACACAGACACAGAAATGAAAAAAAGATTGCCTC
AGTGGACCGTCTGTCCACATCACGCCTGTAGCACCACAGCACCACCAGCCTTGCGGA
TCTTCTTGTCGGCAAGCTTGCTCACCCAACGGGCCTTCACAATGCACGGTGCCTGCA
AGTGGCCGTTACCAAGCAGCTTCGAATAGCCGTTGGCTAGTAGGTCAATGACGGGGA
GGACTTCACCCTTCTTGGCCTTCATGGCCTCGTCCTTCGCCATCAGCTTCGTCAGAT
TGTCCAGGTTGATTGTTGGCTTCCATGTCGTGTTCTTCTTCTTGTGGTAGTGATTCA
TCCCACACTTGCCAAAGTAGCCGGGGTGGTACCTGCCCGGGCGGCCGCTCGAA

Contig 26

TAGCGTGGTCGCGGCCGAGGTGCTGCGTRAGGCCCRGCAGWKGTAYCGKCARGTCCT
SGCCAAAACCCGCTGCATTCCCTTCAAGCGYTACAATGGAAAGATTGGYAAYACRGC
ACAAGCSAAGGARTGGGGRCAGACGAAGGGACGCTGGCCRCGCAAGTCCGTTGTRGC
RATGCTSTCRYTGTTGAAGAAYGCYGAGGCRAAYGCCATCGAGAAGGGRCTYGACCC
CGGMAARATGGTCATTAAACACGTTCAGGTGGATCAGGCCCCTCGCGTGCGTCGCCG
TACCTGCCCGGGCGGCCGCTCGAA

Contig 27

TAGCGTGGTCGCGGCCGAGGTCTTGAAGGGCTAAGCGGGGAAAGAAGACCCTGTTGA
GTTTGACTCCAGTCTGGCTCTGTGCGGCGACTTTGGAGGTGTAGTATAGGTGGAAGC
GCAAGCTCAAATGAAATACCACCACTCGGAACGTTGCTTCACTTATCGAATGAAGAG
ACCAATGGTTGTTTTGCGCAGTCTTCGGGCTGTGCGCCGTCTAGGCTGGGCATAATT
GTGGTGTTTCTTTCACCGCCACGGGAGGGGTGTTGGTTTTTATCCTTCTCTCCGTT
TCTCCTTTTCTCTCTTTCTCCCTTTTTTTGGGCGTTTGAGGGTTGGTGACCCGTGGG
CGCCCAGTCGTGCGGAACTACTGCGCCAACGGGGTGGGAACTTTTCGTTTCTTCTCC
GGAGTCTTGTTTCGAGACATCTGCCAGATGGGGAGTTTGGCTGGGGCGGCATATCTG
TTACACGACAACGCAGGTGTCCTAAGGCGAGCTCAGTGGGAACAGAAATCTCACGTG
GAACACAAGCGTAAATGCTTGCTTGATTAACGATTTCCAGTACCTGCCCGGGCGGCC
GCTCGA

Contig 28

TAGCGTGGTCGCGGCCGAGGTACGCGGCGCACCGGCACTACGCCAAAAGCAAAATAA
TAAACAACCTTTGAGAAAAGCGAACTTCCACTCACCTACCCTCACTTTTGGGGCATG

TTCCTGCTGAGATTCCCGTTCCCCACAATATCAATTCCATTGTCACCTATTCTCCGT
ATTTTTCGGAAGCGTCCTTGCAAGAATCGCCGTTGTAAATTCATGCCCTCCCTCCTG
TTGACAACTTCTCTCTCAACCACTTCCACGTTTATCAACTCAAAACCCACGGAACAA
AACAGATCCCTCAACTCAGAAACGGAGAAAAAGTGGCTTAACGTCCCATTCGTACCT
GCCCGGGCGGCCGCTCGAAAA

Contig 29

TTAGCGTGGTCGCggCCGAGGTTGTCAAACTACTGCCGCATAAACTACGGTTATCCC
AAATTTAAGAGAAAGCAATAAAGCATCAATGAGCGGAAAGGAAGTTGAAGGTGTTGT
GAGTCCTGCGGACCAGCAGCAGCCAGCCGTCCCGGAGGTAACAGATATCACGCTGGA
GGCCGCCCGCAAGCAGAAAATTCACAACCTGAAGTTGAAGACCGCCTGCCTTTCGAA
TGAGGAATATGTCCAGGACCTGCACGTATCCGAGTGGAGTGAGACGCAGAAGCAGAA
GCTGCAGGCTGCACACGAGAAAGCGCATGAATTGCTTGCCTCAGTGGAGGGTGGGAC
GAAGTGGAGCCTGACAGAGGCGTATGACATCAAGAAGCTGATGCGCGTCTGTGGTCT
TGAGATGTCTGTGCGTGAACTGTACCTGCCCggGCGRCCRCkCTaCSGMMC

Contig 30

TAGCGTGGTCGCGGCCGAGGTACGCGCCCGCCCTGTGGAGGATCGCAAACGTCGCTA
CACGTCCAAGGTGCCCTTCCGTCCAACTGGCAAGTAATAACTCCACCCGCCACTGCG
TGCCTGAAGGAGTAACCGGGGAGCATATAATAATAGTCGTGCTGGATGTGTGAAGTC
GCTGAGCAAATCGCAGTCTGTTTACCTCTTTTGTGTTCTATATCTTAGACATCAGAC
ATGTGATCTTGCATTGTACCTGCCCGGGCGGCCGCTCGAAATCGAATTCCCGCGGCC
GCCATGGCGGCCGGGAGCATGCGACGTCGGGCCCAATTCGCCCTATAGTGAGTCGTA
TTACAATTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTAC
CCAACTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGA
GGCCCGCACCGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAAT

Contig 31

TAGCGTGGTCGCGGCCGAGGTACGAGAACGTCCTTCACTCCACTGCGATTCATATCT
CCTCAACGAATCAGCCTATACATCTGCGCGATGTCATCCTCGTCAGAGTCTTCCTCA
TGTTCTTCACCATCCTTCACATTAAAGAAGTGATGATTGAGCCATTCTGGTATCAGT
GACGAGTACCTGCCCGGGCGGCCGCTCGAA

Contig 32

TAGCGTGGTCGCGGCCGAGGTACCTGTGGGGTCCACACCGTGCTCGTCACTGATCAC
CTCCCAGAACTTTGAGCCGATCTGGTTACCGCATTGGCCAGCCTGAACGCAGACGAT
TTCGCGCATAATTCGTTTGAACTATTTCTTTGATGAAAGGGTTACTGTTCTTTCTC
CTCTAGTAGGGCAATATAGTACCTGCCCGGGCGGCCGCTCGAA

Contig 33

TAGCGTGGTCGCGGCCGAGGTACGCGCCTCTCCATGGGCGAGAAGCCGACGAGGTCG
CTGATGATGGCCTTCACGGCACGCAACTTCTTGTGGGGGAGGGCGTAGCGATCGTTG
GATGACGGGCGGCGGGCGCGACGGGTTGTTTTGTAGCCCTTGTTGAAGCCGGCAATA

ATGCCCGTGCGCGGGACGGGGGCCTCTCTCTTGGTAGCTGCCATTTCAACTTGTGGT
TTCAATATAGTACCTGCCCGGGCGGCCGCTCGAA

Contig 34

TAGCGTGGTCGCGGCCGAGGTACTGGCGGGCTTGACTCGCAGAACTCCAAGGGGAAT
GAAAACTGGGTGGGGGAACGGACGGTGGTGGCAACTCAACGTGAGTTATCRGCTATG
CAGGCCAAATTAAGTGAGCTACAAAGGTGCTTGGTTGAAACCAGCGAGGGGAACCTA
GGGAGGGTAAGAGAGTTCCAGTGTCGAGAACTAGAATGGCAGAGAGAGCTTAGAAGA
AGGACTGAAGAACGAGACGTCGCACTGCAGGAATGCGAAAGACTTAGGGGTGCCACA
CCATGTCACGGATCTCAGGAAAAAGTTTCCAGCAGAGCTGTTGAAACTCCCCATGYA
CATCAGGAGCAAGAGCTCGACAATCTTCTGTCAAGGATTAGGTCTATTACGGCTGCT
CATGGAAATGCTGCCCATTCTGGGTTTCCAATGCTCTTCTCGTCGCAGAGGAGCGCA
GGAGGTGCTGGGCAGCCTCTGTCTCCGAGGGGAGACTACCGGCGATTTGCCTCCTTA
GAAAGGTACCTGCCCGGGCGGCCGCTCGAA

Contig 35

TAGCGTGGTCGCGGCCGAGGTCCGAACCAGGTATCGGACCAAACGGGGCGCCGAGGC
GGCCGCCAGCGCGCTTTATTGCACCTCTTTTCGATGCATCAGGGCGCTTACTTGGAG
CGCTGACGCATGCGTCGTCGTCGGCGCTTTAAGCGCTTTATGCGCTTCTTGTGCCAC
TTGGGCCGCATCCCTCCAGGGCGGCTGATGGTCCCCATCCTTCACCTTCTGACAATA
TAGTACCTGCCCGGGCGGCCGCTCGATA

Contig 36

TAGCGTGGTCGCGGCCGAGGTACACCGCGTGGTGCAGGAGGAATTCCAACGAGGTCA
AACTGACCCATCATTTGGTTGTCAGATGCCATTTCTCGTTCCCCTTGGAACACTTTG
ATTCCGACCTGTGTTTGGTTGTCCGCAGCAGTTGAGAAGGTTTGACTCTTCTTCGTC
GGGATTGTTGTGTTCTTGGGAATCATACGTGTGAAGACACCACCAAGTGTTTCGATT
CCAAGTGATAGTGGTGTAACGTCAAGCAATACAAGACCCTTCACATCACCACGGAGA
ACACCACCARGTGTTGCAGCACCCAGTGYAACGGCCTCGTCAGGGTTCACACCACGG
AACGGTTCGCGACCGAAGAACTGCTTCACAGCCTCCACAACCTTTGGCATACGCGTC
ATACCACCAACAAGCACAACCTCAGAGAATTCCTTCAGGTCAACAGCAGCGTCCTTG
ATGCATTGCTTGCACGGTCCAAGTGAACGTTGTACCTGCCCGGGCGGCCGCTCGAA

Contig 37

TAGCGTGGTCGCGGCCGAGGTACGGACGTAGTCCAAGGAATACACAAAACAAAGAGA
TGCAGCACCAGCGAGACCACCAGATGCCATGTTACCCATAAACCACTTGCCATAGCC
ATCCTTCTCCTTTTTGTAGTTGAACATCCTCTTAAATTTATCCTTAAAGGCGAAATT
CAGCGCCTGAGTTGGGAAGTAACGCAGAACATTTGACAGGTTACCTCGCCAAAGCGG

ATACACGCCCTCCGTGCTGATGGTGCGCCGAAAGCAGTCAACGACACCGTTGTATGG
CTTGTCCAGTCGGCCTTGCTTCATCATCTCTCCCTGGTTCTGTACCTGCCCGGGCGG
CCGCTCGAAA

Contig 38

TAGCGTGGTCGCGGCCGAGGTACCGAAATCTTTGTTTGAAGGCCCTTGAGGAGCGAC
GGAGGATAAAGGAAGCTGAAGATAACCAAACAAATGCTGAAAATGTTTGTGCTGAAC
CTACCTGTGACGATAACGCTGATGCTTGTCTGGAGGAGAAAGCTGCTGATTCAAGCA
AGTAACATCTTTTTTTtCTTTTCACTAACTTATGTTTTTAGCGGCACGTTACAGTGT
ACCTGCCCGGGCGGCCGCTCGAA

Contig 39

TAGCGTGGTCGCGGCCGAGGTACTGTCTCGTTTACTCATCTTCCACTTGCTCACTTC
GATATTAAGATGTGTGAGAAAACCAACCCAACCCAAGGGCAGCATTTCAGTCTCCTT
CACAATGACTGTCCATCCCACAACGCGCCGCTTTTCCATCCCGAAGTCCGGAAGATC
GCCGTTACTTCGTAGTGCTTATTGTGTCAACTCCTCACCACCGCAACAGATAATGTG
TAAAACGATGCCAAGGAAGAACCAACAGAGGTGTAGCGAAGTAAATTGACACACCCA
AGCACAAACATATAGCAATTAGGATCAGCTGACGGTATGGCAAACCAAGGGACCATG
GCTGCCACCAACACACTTCAACTCGTTACCAACCGTGACGACTCTTAGCGGACTCGA
GCACCACGACAGGCGTACCTGCCCGGGCGGCCGCTCGAA

Contig 40

TAGCGTGGTCGCGGCCGAGGTACCCTCGCCCGTGCTTCTGGCGACTATTGCATTATT
ATTTCGCACAACACTGAAACGGGTCGCACCCGCCTCAAGCTTCCAAGCGGCCAGAAG
AAAACGGTCCCCAGCAGCTGCCGTGCCATGATTGGCATCGTCGCCGGTGGCGGTCGT
ATTGAAAAGCCGGTGCTTAAAGCCGGTAATTCCTTCTACCGCTTCCGTGGCAAACGC
AACTGCTGGCCCAAGGTACCTGCCCGGGCGGCCGCTCGAA

Contig 41

TAGCGTGGTCGCGGCCGAGGTACCAAACACAACAGCGAATACGCCGACAATGAGCAA
CTGGTATGTGACAGCGTAGTTGAACCACCCAACCTCTTTCAGAGAGAAGTTCCCCGC
ATACATGTAATACAGTGCTGGTAGAATAAACGCAAGGAAGCCACCACAGAAACTTCC
AAGAAAGTTGAAGACAACTTCAATACTGGGTAGCACAAGTCCCAGTACCTGCCCGGG
CGGCCGCTCGA

Contig 42

TAGCGTGGTCGCGGCCGAGGTACGGAGCTTCTTGTTCACGCGGAAGTCAAGCTGCAG
ATGACGCAAGTCCTTTGTGAGCGTGCCACGCTTCCCCTTCACCGTTACGATGCGGTC
TTTTACGGAGACGGTGACATCTTCGGGGAAGGTGATTTGGTCGTGCGACTTGATCTT

CATGTCTCAGTCTTCCCCTCCCTCTTATCTTCTGCTGTTTAGTGGCTGATTTGGACG
AATCCTTTCTCAGCTCAGTATTCGCGCAGTCGTATGTAGCGGCCACGTGAGCGTGAC
CAACTGGTTCTCTGACGAGAGGAAAGTACCTGCCCGGGCGGCCGCTCGAA

Contig 43

TAGCGTGGTCGCGGCCGAGGTACACCATAAGCCTTCATGATTGATTTCGTTTTATCA
GCGAGGATGGGAATGTTCATCGTCCCCAATCCACCCTTCTTGCGTTCAACATTTGTC
CATGCAAGATGTGAAAATTCGCTGTCCATTGAGCATGCAATAACCTCACAATCGACG
TCGTTGAACTCCTTCACGCGGTCAGAGAATTGGCAGATTTCAGTGGGGCACACAAAC
GTGAAGTCGAGAGGATAAAAGAATAGTACCTGCCCGGGCGGCCGCTCGAA

Contig 44

TAGCGTGGTCGCGGCCGAGGTACTCAATTTCAATTGGCGGCGGCGTTATAGGAAAAT
GGAGAGAGCGCGCTTGTGGAACCGAGCAAGCACACGCAACGCGAGAGCCAGCGCGAC
GCAGCATACTTTTACCTTTGTTTATTAATGTCAATAACTTCGCAATATAGTACCTGC
CCGGGCGGCCGCTCGAA

Contig 45

TAGCGTGGTCGCGGCCGAGGTACTCGCGCTTGAATGCATCCAGAACAAATGGGCTCG
GTGGGAGCTCGTCCAGCGGGTTCGGCTTCTTCTTTTCCTTCGGGACATCATCATCAT
CATTATCCTTCGGAGCGGGCTTGAGAGCCTTCTCCGCCTTAGGTGCTTGCTGCTCTG
CCTTTTTAGCCTTCACTGGCCCGAAAGTTGCACCCTGGGATTTTAGTACCTGCCCGG
GCGGCCGCTCGAA

Contig 46

TAGCGTGGTCGCGGCCGAGGTACCACTGGCATTGTTCTCGACGCTGGTGACGGTGTG
ACACATACCGTGCCCATATATGAGGGTTATTCTCTTCCCCATGCCATCCGTCGTGTG
GACATGGCTGGTCGTGACCTGACGGAATATCTCATGAAGATCCTAATGGAGACTGGT
ATGACGTTCACCACCTCCGCTGAGAAGGAAATCGTGCGGAATATCAAGGAACAATTA
TGCTACGTTGCACTGGACTTCGACGAAGAGATGACGAACAGTGCTAAATCTGTCAGC
GAAGAACCGTTCGAACTTCCTGATGGCAATGTTATGCAGGTGGGGAACCAGCGTTTC
CGCTGTCCCGAGGCATTGTTTAAGCCTGCTCTCATTGGACTTGATGAGGCTCCTGGG
TTCCATGAGATGACCTTTCAGTCCATCAACAAGTGTGACATTGACGTGCGTCGTGAT
CTCTACGGCAACATTGTGCTCTCTGGCGGTACCTGCCCGGGCGGCCGCTCGAA

Contig 47

TAGCGTGGTCGCGGCCGAGGTACAACACTTCCTTGAACTCAGCAGAACCGGCAAGCA
CGAGCCCCACGATATTTGGACGGTCGTTTGTTATAAATAATTGCGTGGCGCCTTCGG

CCACCTTGCGAAGGTAGTTGTGGCGCCTTTCCATACGAATACGTGCGAAACGATTTT
TACTTTGACCACCTCTTCCGTGCTTTTTCGGTAGCTCTACGGTGAATGAACTCAGCT
TCTCCTTCACACTCCCACAAAGAGTTGCATAGGTTGTGCCACTACCGTCTACAACAA
TGAATCCAAACTTCTCATCAGACTCCAGCAT
GCGGTGGAGCTCTTCGGTGTGGAACTTGTTGTCACAGAGGTATAGGCTCCGGCTAAC
CGGCTTGAAAGGCTCAATATCAAGTGTAAGTTTTTTCTCCTTGTTTTCTGCCGTGAG
GACGGTACCTGCCCGGGCGGCCGCTCGAA

Contig 48

TAGCGTGGTCGCGGCCGAGGTACGCAATTGCGCGACGAAGAAGCGCCTTTACGTGAG
CCGGCTCGATGGAAAGCGCCGCGTTGCAGTCATCAATGACAAGTTGGTAACTGTGTG
TTTGTTGATTACAAGTCGCACGGTTCGCGTAGTAATTAGCAACTTCGGCACCACCTC
CAGTTGACAGTTCAATGGCTCGTGAATAAAAGAGTGCGGCGCGCTCGTACCTGCCCG
GGCGGCCGCTCGAA

Contig 49

TAGCGTGGTCGCGGCCGAGGTACACTCAGTTCGCGGATCATCACGAAGCAGACCAA
CACATGCTTCTCCGGAGCCACCAGAAGGCATCATGCGGCACGAAGTGAAGCTCACTG
ACCTCGGATTGCCCCGTGTTGTGCCTTGAATGCATGGAGGAGAGAGAGTAGGTTCAA
AACAAGAAAGCACAGAGGGTGAAGAAAGGATTAGATTAAAACTAAGACAGACACTCC
TGGGGAAGGTACCTGCCCGGGCGGCCGCTCGAA

Contig 50

TAGCGTGGTCGCGGCCGAGGTACGGAAACATAGAAACAGAGAGGGGGGATGACAAAT
TGCCCCTACGATTGTTCACATCCGCCTGTTACCCCGGAGGGATTTTGTTAATTGAAA
GGATCCTTATTATTGTCTCATATGATTCATCCCCACGCGCTTATACGTATACGTGAG
CAGGGGGTCTGTGTCAACACATCTGCTGTTTACTGCAACTCGTATTGACTTCCCCCC
CCCCTTCCTTCGGGGCACACTTGCACACCATTTCCCCTCCTCCTTTCTTGTTGCCTC
ACTCACGCTTCATCGATAAAACACGGTGCAAGACGTTAGAAAACTCCCAACTCAATC
GTTCTTTTCCTCCTCCCCCTGGACTTACAAACTCCTTTCGATCTTTTCTTTGAATAC
GCTCATCAGTTAATTTGAAATGCCTTACCTCTCCCACCCCACCCCGGCGGCATATGA
AGAATAAAAATAAAAAATACCCCCTTCTCTCGCTCATCCCGCTCAGCATGCATTGCT
TATACCACCGTGGATTAAACATGTACGAGGTAAAAAACCTTTCTGAATCTTCTTTAG
CTGTGGCCCAGCCCACTTACTTCCGTACCTGCCCGGGCGGCCGCTCGAA

Contig 51

TAGCGTGGTCGCGGCCGAGGTACATCACGATGGGAGGCGAGCCCCGCTTCAACCGGG
CCTTGTTTGAGAACAAGTCGATGGCACTCTTTGTCGGTGACTCCATGTCAATACATC
ACTTCACTGACGACTACGGCAGGATCACCCGCCGCTTGCTTATCCTCTATCCCATGA
CCTTCTCTCGGCGGGAAGAAATTCCCAAGCAGCAGATCCTCAAGAAGGTCGAGAAAG
CCGCGAACCAGCCCGAGGATAATTGCGGGATTGAGGTTGTCGACTTGGTGGAACTGC
AGAAGAAACACGTGTACCTGCCCGGGCGGCCGCTCGAA

Contig 52

TAGCGTGGTCGCGGCCGAGGTTGTAACCAGTGACCCAGAAGCTAATTACAGGCAGTT
CCGCTTACGACTAAGCAGATTTCATTCAGCAACTAAAGCAACACAGGTATGCTCCGC
GTTTCCATTCCCGCACTCAAGACGCTGCAGCCTCTGGGCTCGCGTGTGCTGGTGAGA
CGCACTCTGGCAGCCAAGCAAACAAAGGCCGGTGTGTTCATCCCTGAGCAAGTTGCT
GGTAAAGTCAACGAGGGTACCTGCCCGGGCGGCCGCTCGAA

Contig 53

TAGCGTGGTCGCGGCCGAGGTACTGTCTACTACGAACTCCTTGGTTCAACTTCAAAC
GGTTGCCTGCTCCTTCTTTGCACCATCATCAGCAGCCGTTTCGTTGGCGCCGGGATT

TTCATTGTTCTCAGGTCCGTCAGCACGCGTCATTTCTTCCGTTTCCTCATCAGATGC
ACCAGGGATGTGAAGCTCATCAGGACGGCGGTGTGCAACATACAATGCAATGCCAAT
GCCTACCACTGTAATACCGAGGCAAACCCATACGAGTGGGAATTCCTTTTCAGTGGT
ACCTGCCCGGGCGGCCGCTCGAA

Contig 54

TTAGCGTGGTCGCGGCCGAGGTACTACCTGCAGAGGATTCCCACGTGTAGGCATCAT
CCTCGTTGTTCTTCGACACAACGGTTACGCGGTCTGCGACAAGGTAGGCAGAGTAAA
ATCCGACACCGAACTGCCCGATCATGCTCATATCGCCACCGGCCTCCAATGCCTCCA
TGAATGACTTGGTACCTGCCCGggCGGCCGCTCGA

Contig 55

TAGCGTGGTCGCGGCCGAGGTACCGCAATAGTGCGCGTTGCATCCTCCTTAACGACC
ATCGTCCATCAACACTTCACCGAAGGCATGACTACCCCAAATCATACTGTAATTGTT
CGGTTCCTTTTCGCGAAGGGCGTTTGCGGGGGCCATGCTCTACTGGGAAACATCTCC
AAATACCGCGAACTCACGCACATCGGCCTCCCCTAATTTCCCACACATTTCCTTTAC
AAATGAATCGAACGAAGTCGTGTGTCTGCTTGCCTCATGCGGCATTTCCCTTGACCC
CTTTGCATACGTACCTGCCCGGGCGGCCGCTCGAA

Contig 56

TAGCGTGGTCGCGGCCGAGGTACTGCTCCTTCGTTAGGCATGGGTTGAATGGGTAAC
CGGCGAGACTGCGGCCGCAGCGCACACGGGTGGAAATGACATATTTACCTTCCGGAT
CAACGTCAACCAGCGTATTTAAGTCACCGAAGTCCTTTGGTGGTTGTTTGTCCGTTA
CCTTGAAGCCGTTGTGGTAATCCTCAATGACGGGATCGAAGAGGTCGGCAAACACAG
TGTAGGCTTCTGCGTCTGGTGCATATAAACCAACACCGGAATCGAGGTTCTGTACCT
GCCCGGGCGGCCGCTCGAA

Contig 57

TAGCGTGGTCGCGGCCGAGGTACTAAACACACATCAGCCTTCTGTGCACTTCCAAGG
TTTTTCGCGTGGCCAAGAATACCGAAACCTGTCACATCTGTTGCACCATGGGCACCG
TGCGAATGCATCAGCGCTGCCGCCATGCGGTTAAGACGCTTCATCCCATCTGCGGCA
GCGTTATATAGTTCTTCAATCTCCTCCGAGTCCATCTTTCCTTGAATCTGCTCTTCA
AAGATAGGGGACGGGCGGCGCAGCCACTGCTTAAGATTAACAGCGATTTGGCATCCT
AAAGGTTTCGTCAAAACTAAAATATCACCGGGTTGCAGCCCTGTTGGACGGATCATT
TGAGACTCTGCAACAACGCTCGTCGCAATGCCACCAATGAGGGGCCACGGATTCATC
ACAGTTTGCCCACCTGTAACATCAGATCCCGCTAGCCTTACGTGGTCTACAAATCCC
TTCATCATTTCCTGCGTGCACACCTCACGCTCCGTCTTATCCATGTCGGTACCTGCC
CGGGCGGCCGCTCGAA

Contig 58

TAGCGTGGTCGCGGCCGAGGTACAAGCTTTCATCTTGACCGGTGGTAAGAGCAAGCA
AACGGAAGGTCTCCTCCTTCTGGATGTCGCCCCGCTTACGCTCGGCATCGAGACGGC
AGGTGGCGTGATGACGGCACTGATCAAGCGCAACACGACGATCCCCACCAAAAAGAG
CCAAATCTTCTCTACATACTCAGACAATCAGCCCGGTGTGCACATTCAGGTCTTTGA
GGGTGAGCGTACCTGCCCGGGCGGCCGCTCGAA

Contig 59

TAGCGTGGTCGCGGCCGAGGTACCTCAGTGACCCTTCGGATCACCTGGTCGACATCT
ATAGGAGATGACGACAATGTTTCTGACACATTCGGATCGTAGCGGAGAATGATTTGA
GCCTCTTTCAGAGCATTTTGTGGGCGCATACAGTTGATCACCATAGGGGGCGCCTCT
TTGTTGTGACGAGTTACCCGGTAATCGTCGAAGATGTCATCCCTAAGACTTTACACC
GCATGGGCACCACCACTGGACGAAAACTGCTCTTTCCGGTTGTCGCTCCGAGCATTA
AGCGGATGAACGACACGGCTCATAACTCCTAACTCGTCATCAGTGAATTGTTCGAGG
GCGGCAAGCTCATTTTCAATGTCAGTAGAAAGAGACTCTAATTGTTGTATTTTCTCT
TTTTCCGTTGTGGTATCACATTCAATTCTTGCTTTTGACAGCATATCAACCATCCTA
TGCGCCTTCTTAAGTGTCTGGTTGGTGCTTACAAACGAGGATATTTGACCACAGAAC
AGATTTTCAATCAGTGGCACTACACGGACATTGTCACCTGATTTTTCTGCCAGATCC
ACAAAGGCAACGAATGCGTACCTGCCCGGGCGGCCGCTCGAA

Contig 60

TAGCGTGGTCGCGGCCGAGGTACACATATTCTTCTCACAGTGAAATGATGATGGCGA
TAACTACAAACACACGCCACCCCTACACACCACTAACACCCCGCTACCAACCATCAG
ACTATTTTTCAGAAAAATCACCCTCACCCCCCATTGGTTTACTGGGTGCATGGTGAC
AGTAGTGATCATCGTAACCATAAACGTCTCGGTTATTGAGGCCCAGCAACTCATTCT
GAGCCTTCGTCTCTCCGAGTAGCGCCACCAACTCCTGTTTATTCTTATAGCTAGACC
AGCTGTATCGCTTCACCATTGCACGATACTTCCGTACCTGCCCGGGCGGCCGCTCGA
A

Contig 61

TAGCGTGGTCGCGGCCGAGGTACCACACTTCATCGCGCTGCCCAGTCAGACCACCAA
ACACCACGTGCGATGTGTGCAGACAAAAGGGAAGTTTGTGTTGGATGCAGGAAAAGG
TATAGTGGCAAAGGGATTTAAAGGACAAAAAGAAGGGATAAAATAGAAAACACTATA
CAACACGTAATGCAACGTCTCAACCATAGGGATTAAACACCTAAAGGCTCACCCACC
CACAACCCCAACCGCAGAAAATCCAATAATAATGGTGAGGAGGAGGGAAACAGCCTC
TCTAAACAGCGTACCTGCCCGGGCGGCCGCTCGAA

Contig 62

TAGCGTGGTCGCGGCCGAGGTACCAACACATTGCCCTCGTTTCTACTGGTTTTGTTT
TATTTGATACCTTCGATACTGCGCCCCAAAACGACGGAAACCGCGAGTCAGTTCAGT
AAAAGATTTGTGCACGATCCAAGGTTTGTAAGATTTCTTTTTCATCCTTTTCTTTAC
CGTTTCTTTGGAAAGGATGGATGGAAATGCGACCGACATCTCCCTTTTTTCTGTCTC
CACGCGCTCCCGCTCCAATTTCCCTTTACTCCTCCTCTTTTTTCTTCGAGAGAGTGG
GGCATTCGCACACGCGTACCTGCCCGGGCGGCCGCTCGAA

Contig 63

TAGCGTGGTCGCGGCCGAGGTACAAATGACAACGAACTCCCTCCAGGGGCTCCACAC
TTCCAACCACGTGTTGCTGCCACAGCATTTGATTCCAGTGGTCAGAACTACTTGTGA
CGGAAGGCGTGACGCCGTTTGCTCGAGCGCTGCCTTTCCTTTCCACGTGACGACACA
TGCGGTTTCGCGTGGCTTCCAGGAACACCTGCAGCACCAAAGTGTTTCACTGATTCG
CGTCCAGACTTACGGCCACGGAGCAACATCGTGTTTTTCCCCGTCGGTGCCATCATC
GCAAGCTGATCGAAGGTTAAGCACTCGCCACCAGCACCAGTAATGCGCTCGCGGGCG
CTCTTCGAGAAGCGGAGGGCGCAGATACGGAGGGCTGGAATGCGGGTCATGCGAACA
TCATCGAGCACGTCACCAACAATGACGGCAATTGGTGACTTCTTCCCCTTCTTAAGC
CACACGGTGCGGCGCCTCATGCAGACAGCAATGCGGCTGAGTGAAATGTACCTGCCC
GGGCGGCCGCTCGAA

Contig 64

TTTAGacGYGGTCGCGGCCgAGGTACGCCGCCACAAAGTCTGCTGTCCGCATCGCAG
CCAAGAAGGCTTTTGCCGAATCGACTGATTTCGAGCGCTACCAACTGCGTGTTGCCA
AGCGGTCTCGCGCCTACTGGGCACGCAAGATCTTCGACGAAAACGACAAAAAGAACC
CCGTCTCGTGGCACAAGGTTGCACTTAAGAAGCTTCTCAAGAACGCTAAGAAAGTGG
ACAGCACGCCAGCTGCCAAAAaGCGTGTCGAGAAGGCCCGTGCTGCACGCAAGGCCA
GGGTTGCTTCCGGCAAGACAAAGGCGAAGTCCACCGCTGGTAAGAATTAGAGAGTTG
TCTCGCGACACCCCTCGACGTTCCCGATGAGTCGCCGTCCTTTCCTGCATCCGTTAG
TTTCATTCGTACCTGCCCGGGCGGCCGCTCGAA

Contig 65

TAGCGTGGTCGCGGCCGAGGTACGTCCTTTCTCTACGCCTTCAATCTCGTATCTGGA
GCAATGCAAGCGCGCTACAACCTCACACAGCGAGATCTCTCCACCATCACTACTGTT
GGTATCGCAGTTGGATATTTCTTGCTGCCCTACAGCTTCATCTATGACTATCTGGGG
CCGAGACCCATATTTATGCTCTCAGTAACGGTGTTCTGCCTGGGTACCTGCCCGGGC
GGCCGCTCGAAGAAGTGATTGAAGGATCTGTTGTGAGACTCAGTGTCTACAATGGTT
TTTTAACATTGGGATGTATGTTGTTTGACTTGGGTTCAGTGGTAACAGTACCTCGGC
CGCGACCACGCTA

Contig 66

TAGCGTGGTCGCGGCCGAGGTACATCTCTGCAACCTGTGCAGCCGATGTGGCCTCAT
GCACCTTCTTGTCCTCACGCACGCGAATAAACCGCGGAAACCGCAGTGCAATTCCCT
TTGAAGCATCTACTAACCCATATGCAGCGAAATGGGCAGGGGATATAGAAAGATCAG
CAGCCTTCACTTCCCACACTTGCGATGCCTGAAACCACACATCCGGCTTATCTTCAG
TGCGGTAGTAGCGCGGTTGTTCATCCAATATAGATGGTTGGAGTGATTTGGTGATGG
CCTCAAGCTGATCATCCTGAAACCCAGTTCCAATCTTGCATATACTCTGATACTCGT
CACCCTCACCGTCGTAGCATGCGAGAAGAAATCCACCAAAAACACCGGTGCGTTTCC
CTTTCCCGTAGTATGCACCAATGGGAACGAGGTCAAGCGTATCGGTAACACCATCCA
TATAATCTTTTTTGAGTTTCAGCCAGTAGTGCGAACGTTTTGCTGGCACATACGTCG
ACTCCTCTTCGAGCGTCTTCACCATAAGTCCCTCCGTCCCATCTCTTATGGCGCGAT
CCAGGAAGGCCTGAATGTCATCCACATCATTGCTGTCCAGGTACCTGCCCGGGCGGC
CGCTCGA

Contig 67

TAGCGTGGTCGCGGCCGAGGTACGCCATCGTGAAGGACAAGGCACGCCTGGGTGATG
CGATTGGTCGCAAAACCGCCACGTGTGTTGCCTTTACTGACGTCAACGCGGAGGACC
AGGCTGCGCTGAAGAACCTTACTCGCTCTGTGAATGCCCGCTTCCTTGCTCGCAGTG
ACGTTATCCGTCGCCAATGGGGAGGTCTGCAGCTGTCGCTCCGGTCTCGCGCTGAAC

TGCGCAAGAAGCGTGCTCGCACTGCTGGAAATGATGCCGCCGCTAAAGCCGCTTAGT
AGTTGGTGCCCCGCCGCTTTTCATCTGTCCTGTATTCTGGCATTTACTACGTTGTGT
GTGCGTACCTGCCCGGGCGGCCGCTCGA

Contig 68

TTAGCGTGGTCGCGGSCgAGGTACTGCTGAGGAAGAGGCTGCAGAATTCATAAAGAA
TCATCCCATCAAGAAACCTTTGGTTACTTTCATTGGTGGTGCCACTGCCCCCCCAGG
ACGCAGAATGGGTCACGCTGGGGCTGTTGTCTCTGGTGGGCAGGGGACTGCGAAAGG
AAAGATGGAAGCACTTGAAGCTGCAGGTGCCGTTGTGGCCCATTCACCCGCTCAGCT
CGGGGCGCTGATGGCTGAGTTAATGAAAAAGAGGGCAATCTAAAGCGCAATATCGCT
TTTTTCCTTTTCCGTTTATTTGTTTATACAGTAGGAACTTGAGGGGAAATTTGGTAC
CTGCCCGGGCRGCCGCTCGARCCCTATAGTGAGTCGTATTAGA

Contig 69

TAGCGTGGTCGCGGCCGAGGTACACCGAAAACGAACAAACAATAAAAAAATACGTTC
AAAGAAAAAAAAAAATAAAGTAATATAGTAAACGATCTACGAATGGGTGATAAGGAC
ACCGGGTTTGAGGCTCTTATCCACTGTCACCGCAACAGTCCAGTTATCAATGAGTTT
CGCAATATACACTGCTGCCCATTCAAGTTTCTTGTTTACACGCGCCTTGAGCACGTT
GCCCTTCACCAGAACACAGTCTGCTTCCACACCAACAACACCTGTCCATGGCTGGCC
ACGTCCGCACACAACCTCAGCACCAACAACGACACGATGGGGGCACTTTGGCACACG
AGCAGTGATACTGGTGGTGTATGTGTTGAGTTTATTTGTGCGGACGGCTGCCGTGTA
ACCGCAGCGCGCGTAGCGGCACGAAGCAGTCCAATCACAGTTGCCTTTTGCTCCCAA
ATTGTAGAGAGCACCGCAGCCAACGCTGCAGTATGGCCCGATAGCTGTACCTGCCCG
GGCGGCCGCTCGA

Contig 70

TAGCGTGGTCGCGGCCGAGGTACTCCACCATCTTGCTGCGGCGATCCATGTTGCCGG
ACTCAACCTCCTCAGCAGGGTGAACAAAATCGATACCAGCCTGGTTCAGCGCATCCT
CAGTAGGTCCAAACATCTCCAACGCCTGCGCCATCTTGTCCTTC
AGCATCTCCAACTCTTCCTCCACCTGCGCATGAAGTTTGTATAGGTCTTTCTTCTTG
TCGGAGTGTAGTTTCGCGTTGGGGTCAAAGGTCTCAATGGCAAACTCCAGTTGAATG
TGTGTGGTGCGGATGTTGCGATCAATCTCCTCCAGGCGCTTTTC
TTTCTTGTACCTGCCCGGGCGGCCGCTCGA

Contig 71

TAGCGTGGTCGCGGCCGAGGTACATAACTCTGGCTAAAGCTTGCCCGTGGTGCGGTG
CGGCTGATTCTCCCGTCTTCGGGATTTAGGGGTTCCATGTTTGCAGTAGCTAAACCA
CGCCTGCCAATCACGTATATGAGCGTAGAGAGAGAGAAAGTTCAAGGGAAGAGAAAA
CCACGAAAACACCAACGACCGAAATAGCAAAGAATATAGAGAAATTCCTATGTTATT
ATTGCTGACTGCTGCTCCTGTGGGTCACAGCCCCCTTTCCTTCCTTAAACGGTGCTC
CAAAAATTGAAAAAAAAAGCAAGATTGATCCAACCGTATGCTTCACGGACGGGCATC
CAAGTGCATGCACCAATATGTCTACAAACATATATATCTGCCTCTATACTTTTTAAA
ATTTACTCTGCATATGTCTGCTTCATCATGCGAGTGCTTGTGCGAAACTGGGAGGC
GTGAGTACCTGCCCGGGCGGCCGCTCCGAAA

Contig 72

GCCGASGTMGCATGTTGCCGGCCCAACATGGCGGCCGAGGGAATTAGATTAGCGTGG
TCGCGGCCGAGGTACTTCATCGGCCCAGAGATCACCAGTGGCCCATTCTTGGACAGC
TGCTTCAGAATGACAGCACGGCGGCCACGGAAACGACCAGCAAGAATGATGGCAATG
GTCCCGGGCTTGCAACTCTTGCGCACCGTCGAGTATTCGGGGCTCTTGCGGCTCACA
CCTGCCCGGGCGGCCGCTCGAA

Contig 73

TAGCGTGGTCGCGGCCGAGGTGTGTATTGCCCGACACCCCAACTGAAACGGCTGATG
TTGATAAAGATCCCACCACTAGCGATGGGTCGGTGCATTATGATGGAAAGTAGTTAT
TCGGATCTAAGTGCGCTCCCCTTCTTCCCAGCATTTCCTCATCGGCTTTTCACACTT
CACTGCTATCATCATTTACTCATTGAATGGGTTTACTCACGTARTGGTGGACACCTC
TCGATGTTTTTTTTCCTGGTTATGAAGTGTACCTGCCCGGGCGGCCGCTCGAA

Contig 74

TAGCGTGGTCGCGGCCGAGGTACCCATCGATAGACAGGTTGGTCAAGCCACACCTTT
TTGTTGAAGCGCTGCGCTGACACCCACGGTTTAGTAGTCCACAGCACCAAATTGTAG
AGTGTAATATCCGGGTGTAGGAAGGTAAGGCGATATGTGAACCACAAGGTGTATGCC
ACCACGATAAGCGTCAGGACTGTAGCGGGAATTCCCTCCATGGAAGTGAGTTGCAGG
CGAATCTTCTCAATTGTGCTCACCTCCAGCCAGTGAACGTAGCGACCATATGGGTTG
CGCGAGAAGAAACCACATGCGAACCCTTGTGGTAGCATATTGTAGCGAAATCGATAG
AGCCAATGCATTGGTTTCGGGCTGAAACCCTCAAGGTAGCCGGTGCTGACACCGTCA
TGAACGGCGTAGATGTACCTGCCCGGGCGGCCGCTCAAA

Contig 75

TATTYGTGGTCGCGGCCGAGGTACTCACGCCGCTGTTGAAAAGTATGACGTGAAGAA
GGCGTCCAGCAAGCTGAAGCGCAAGTGGGAATACCGGAGGAAGCATCACAAGATTGA
GAAGGCATTGGCTGATCAGTTGCGTGAGGGTCGTCTTCTCGCCCGCATCACGAGCCG
CCCAGGTCAAACTGGCCGTGCGGACGGTGCGCTTCTGGAGGGCGCTGAGTTGCAGTT
CTACCTGAAGAAGCTCGACAAGAAGAAGCGTTAGGAAGCGTTATGAAACCATGGTAT
CTTGTTTGGGCGCTCATCCGCTCGTATTTTTGTTTTTTTCTCTTTTCGTAAAAAAAA
AAACCGGTACCTGCCCGGGCGGCCGCTCGAA

Contig 76

TATCGTGGTCGCGGCCGAGGTACATGTATACATTCCACGCGTTATGCCGCTTCACAC
TGTGACAAGATAAGTTGCATCATACATTGCAACTTAGTGGTCACATTATACCGCTTC
CTCTTCCAGCATCAGTGCAACCTTCCCCTCCAAACGATCCCATA
TATCACGTTGGTTGTGCCCAGCGTTCCACTGCTGAAAGAGGGGCTCAAATTGATGCA
AAGCGTACCTGCCCGGGCGGCCGCTCGAA

Contig 77

TAGCGTGGTCGCGGCCGAGGTACACATAGGAAGAGAATAACCACTGAGACGCACGGC
TGTGTTGACCGACCGTTAAGCTTCCTTTCCTGAGTCGAATGCCCTGAAGGCACACGT
CTCCGCGGCGAAAAAAACAGACAAACAACCGGCGCAACTGGGACTAAGCAACGCCGG
AGCGCACTATTTCCAACTTACTACTGAGTGGAAGGACTTCAGAGTGAATCAAGGAAG
GCTTTGAGCTCGCCTTCAGTACCTGCCCGGGCGGCCGCTA

Contig 78

TAGCGTGGTCGCGGCCGAGGTACACTTTATTCTCATTTTTGCCAGATACGGGAAATA
CAACATGTGCGGCGCGTGTGCCCAAAGTGAAACCGGTGGTGACCCTCTTTACAGGGC
CTTGTAGCGCACCATCCATGCCAAGGTATCCGTCACGCCAATGCGTGCGGTTGGCCA
CTTCAGACCAAAACCCAACGAGGCCCACCGTTGGGTCAAACTTTGGCGCTGACGTGC
TCTCCTCGCCCGTAGTGCCATAAAGCGCCATCACACTCCCGTCCACCTTCCTCCACA
AACCCACCACTGCCTTCAGTTCATGCACACGGTCGGTGAGATCGGTAAAGAAAATGC
TTTTCAACTGGCTGAAACCCTCCTCATGCAGGGAGTACCTGCCCGGGCGGCCGCTA

Contig 79

TAGCGTGGTCGCGGCCGAGGTGTTTGCCTTCATTGATCCCCCATCGCAGCAGAGTGT
TATGGATGCGTATTGTCAACTCAGTCTTTTCGGGGCTGTAGACGACGACTTACAACT
CACGGACTTTGGGCGTCTGATGGCCGATTTTCCCGTGGACGCCTGTTTGGCCCGTGT
GCTTATGCGTTCCGCGCAATACGGCTGTGCAGCAGACGCAGCGGTTATAGTAGCGAT
GTTAGAAACGAGGAATGTGCTGGTGCGCCCAGCATCCCGTGGGATCGAGGCGGAGCA
GAAACACGTGATGTTTCGTCATCCTGATGGCGACCATTTGACTCTCTTTAAAGTTTT
TCACGCTTTTTGGCGTAGCGGCCAGTCCTCGCAGTACCTGCCCGGGCGGCCGCTCGA

Contig 80

TCGACCTGCAGGCCGGCCGCGAATTCACTAGTGATTAGCGTGGTCGCGGCCGAGGTA
CCATTACCAACAACAGCAACACCATTACAAAGTGGAAGGGGAATATGTTCGTGATAA
CCGACAAGGGCTACCATGCACTCCGTATGATATCAGTGGTTGGATGAAAATGTTGTT
GCTGGCGTTGCTGCTGCTGCTGTTGTTGTTGTTGTTGTTGTTGTTTTCTTCGCCTTC
GTGACTCTTCCTGCTCCCGCATTTCAGGTAACGGTGTGCGTGGAGGAGGCGGAAGGC
GGTGGGTGATATTCCATTTGCTTGCGTAGTGATACCACAGAGCCGCACCTGCCCGGG
CGGCCGCCCGGGCAGGTACCGGGAATAAACCGGCCACCGAGGAACGATGTCCCAACG
AGTTGGGAAAACTTGAAGATGGCGCGCTGCCCGAATAGGCGTGTGGAACAGACGGTG
ACATCCTGTGGGTTCTCCA

Contig 81

TATTNGTGGTCGCGGCCAGGTACGATGCCTTCTTGTGTTCGGAGTCTATCATTAAGA
CAGTCCCGCGTCTTGTCGGCCCTCACATGCATCGTGTAGGTAAGTTCCCTACGGTGT
GTGCGCAGAATGAGAGCCTGCCGGACAAGGTGCTCGAGCTTCAGTCCACAGTGAAGT
TTCAGCTCAAGAAGGTGCTCTGCCTTGGAACATGCGTCGGCCATGTCGATATGACTG
AGGATCAGGTACCTCMCGGGCGGCCGCTCGA

Contig 82

TAGCGTGGTCGCGGCCGAGGTACTCGCCCTCTTCGTTCACACCCTCACTGAGAAGCT
CGTCTTCACACCCCTCTTGGTCATAACTACCATCATCATCATCCACGATCACGT
TTCCGTTGCGATTGCCTTCGTCGTCAGCGTCGTCGCTACCCTTTGCGAGTGCTCGTT
GACGCTTCTTTAAGCGCTCAGCTAAGGAAAGGCCCTCCTCTCCATCTTCGTCACTCT
CGGTTGCCGCTAGCTTTGCAAAAAACTGGTCTTCATCCTCGTGGTAGCCCAGTACCT
GCCCGGGCGGCCGCTCGAA

Contig 83

TAGCGTGGTCGCGGCCGAGGTACCGTAGCCAACAAAACCACCCATCGGAGTGATGGT
CTTAGCGGTAAGATCATAAGCAGTTGTGGCTTGGTTGGGTTCCATGGAAACTGCACG
ACCGAGCTGGTAAATCTTCTTGTTGAGGTGTGTACCTGCCCGGGCGGCCGCTCGAA

Contig 84

TAGTCGTGGTCGCGGCCGAGGTACATCAGTGATATTGTGGGCACGGCACATATAAGG
GAGGAATAAACACACACACACATACACACATACACACACCGGGATAACTTTCCATGT
GACCTCTTTCTGTTTGTTGCTTAGCTCACTCGCGCATCTCAACCAGTTTACCCACGT
ACCTGCCCGGGCGGCCGCTCGAA

Contig 85

TAGCGTGGTCGCGGCCGAGGTACTTTCCGTGGCCTTCTCTCCCGACAACCGACAGAT
TGTCTCCGGTGGTCGTGACAACGCTCTGCGCGTATGGAATGTCAAGGGCGAATGCAT
GCACACTCTGAGCCGCGGTGCTCACACCGACTGGGTGAGTTGTGTCCGCTTCTCCCC
GTCTTTGGATGCCCCTGTCATCGTCTCTGGTGGTTGGGACAACCTGGTGAAGGTCTG
GGACCTCGCTACCGGGCGACTTGTCACGGACCTTAAGGGACATACAAACTATGTCAC
ATCTGTGACCGTTTCCCCTGACGGTTCCCTC
TGTGCATCTTCCGACAAGGACGGCGTCGCTCGTCTGTGGGACCTGACTAAGGGTGAG
GCCCTTTCTGAGATGGCCGCTGGTGCGCCTATCAACCAAATCTGCTTCTCTCCCAAC
CGTTATTGGATGTGTGCGGCTACCGAAAAGGGCATCCGCATCTTTGACCTGGAGAAC
AAGGACATCATCGTCGAGCTTGCGCCGGAGCATCAGGGCAGCAAAAAGATTGTCCCG
GAGTGCGTTTCCATCGCGTGGTCTGCGGATGGTAGCACCCTGTACCTGCCCGGGCGG

CCGCTCGGGTAGGTACAACCCAGCACAGAATATGCGTTCGAGTTGGAGGGATGCTGT
CCTGTTCTCCTATACGATACACGGGAAAACCGAGGCCTTCGTGGTGTATTGGAACCT
GTGGCAAGGAAAGGTGACCCGTCTTGCATTGTTGAGTATGGCGGTTGCTACTACGCC
TTACTTGATGAGGAAGCAGTCCGAAGGTTCCTCATGCGTCCATGGCAATATGTGGAC
GGAGCGAAGTTGCCGCCCTCGCGGAAGGTGCCACTACCGGAAGGTAAGACGATGTCT
ACCATTGATGAAGAGGAGTTCATCCGCCGTATACTTTACGACCCTGTTGCGCACGCT
CTGATAGCCGTGGCGGAGGTGCGCCCAAAGTACCT

Contig 86

TAGCGTGGTCGCGGCCGAGKTACCCACGTGGCGCCAAGGCTCAGGTTTACTTCCTAC
CCAAGAAGCTTGACGAAAAGGTTGCTGCATTGCACCTCGGCAAGCTGGGCGCTAAGC
TGACAAAGCTCACACCGAAACAGGCARAATACATCAACTGTCCTGTCGACGGTCCCT
TCAAACCTGACCACTACCGTTACTAAGATTYCCACTTCAGGGGCTGCTGGGATGCAC
CTACGCGTCTGTTCAAGTATTTAAAAAATGGTACCTGCCCGGGCGGCCGCTCGAA

Contig 87

TATTCGTGGTCGCGGCCGAGGTACTTCTCCTGTTTGGTGTGTGCCTTTAAACCCTGC
CACGGCAGTGAGCTCACGGAGGAAATACATGAGGATATAACCGATACCTTCCAAATC
ATCACGTCGGCTCTGCTCAATGCCAATATGTGTGTTGATGGAACAGTAACGTGCCGT
CCCAGTTAAACTTTTGCCTTCCTTGTATGGAATGTGCTGATGTGTTCGTGCATCACG
GTACCTGCCCGGGCGGCCGCTCGAA

Contig 88

TAGCGTGGTCGCGGCCGAGGTCGAGCTTGAGACGATACGAACTAAAAAGACAAAAGA
ATTGGAGCGGTTGAACGGCGAGGCAGCGGCGTTGAAGAAGGAGCTTGAAGAAAAGGC
CCTGAATAGCGCTGTGCTTCTGAGGGAGGTGCTCCTAACAAAGCAAGTGGCCCGCTA
TTCACCACGTGCAGATGGGGGCAAGGAGCGTATTAGAGGGCGGTGTAGAGCGCTTTT
TTTTTAAATCCTGCAAGGGGTTCTGTTTCCCATATCCCTTCCTTTCCACACCCTCAT
GTGTGGTGAGGCGGTTGGTCACGCCCATGATCATATGCGCTGTAATTTGAGGTAGGC
CGAGTGGTGGCTGCGTGTCGCGATATGTAGGATCAGATTGAAGAAATACCAATGACC
ATCACGAGGGTCAATGTATTGTTGATATCATTATTATTTGGTGTTACCGCAAACGCA
GGGGAGAGGTATTTGATGGATATGTAACGCTACACCTCGTCAGAGGCTTCTCGGGAG
CAAGGTCTATACAACTGCCGATGTGGTTCTGGAGGCGGCGTAGAAGGCGACTCCATT
TGTATTTTACGTTGAGTACCTGCCCGGGCGGCA

Contig 89

TAGCGTGGTCGCGGCCGAGGTTACAACATTCGTGATGAAACGGCAATTGGTGATCAC
ACCGTTGTAATGGACTGTGTCTCGTTTCTTGGTCAAGTCCGCGTAGGTGGTGGCGTT

TATATTGGTCCCGGCAGCACGTTGGATTGCTGCACCGTTGGCGACAACGCGTATATT
GGTGCTGGTGCTTCCATTGCATTGGGTGCAGTGGTTGAAAATAATGCCATCATCGCC
GCGGGTTCTCATGTACCTGCCCGGGCGGCCGCCTCGA

Contig 90

AGGTACATCACCTTTTTCGCGTCCCTCCCTAGTCCTCCTCAGAAACAAAGGCGGAAT
TAGCGTTTTAAGTCATTCCCCCCTTGTTTTCTGTTTGCGCTACCCTCAACCGCGTTC
CCTCCCTCGCTTATTTTTTCCTTTACCCATTTCACCACACCCTTCTCTTTCTTCCCC
CCATGCACTCTACACTCACACTTTTTCCTTCTCAGCCTCTTTTGCCTTTATATTTAC
CTCCCAGATCGACGACGACAACACATATTCATATCGTGCGCCCCATGAAGAAGATCC
ACAATGATAATATCAAGGGAGAGGAAAAAGGGGGAATATAAACGCAAGCATCGTCTT
CATTAGCTTATTGGCCTCCCACCGTGAGTTTTCGGGGCATACGTAGGGCCGGCATGC
TGCAAGAATAGTAATAAAAAGGACAAAGGTTGAAAGATTCCACGCGGTACCTGCCCG
GGCGGCCGCTCGAA

Contig 91

AGGTACACGAACTCGGCCAGGAACACCCTTCACGCCCTTATGCCAAATGAAGGTGTT
CAGGGATGCATCAATGCGGTTGTCCTTCGTCTTCATCAGTTTACCCACAAATGAGCG
AATGCGTTTAATAGCKATGGGTGCGCGTTTGTGGAATGTCTTCTTCTTCAGAAGCTT
TGAGAGATGAATGGACGCTTCCATTACGATGTTATCGGGTTTGCGACCAGTTCTCTT
GAAAAGTGATGCACGTGTGCAACCGCGGACACGAGAGTGTTCAGCCGTATTCCCCAC
GCCGTGGTACCTGCCCGGGCGGCCGCTAGAAA

Contig 92

TAGCGTGGTCGCGGCCGAGGTACGTCCTTTGTGTGGCCAAGAAACTTGTATTGACAT
TGTCCGTTCTGCAGGTTCCACAAGCGCAGGGAATGGTCCCATGACGCAGACACAGCG
AAATTCCCGTTGTTGGAAAGGGCAACGTCAGACACAAATGCGCTGTGCCCCTCCAGA
CGCCTGTCGGGAAGGCCATAGCTGCACTCACTGCTGTGCCGGTCAGGGTTAGGGCCC
CATGAAAGAAGGGTCTTGTCCCGGGACGTGGATACCACCTTCGTTGCGGTTTCAGGG
GTTTGTGGGCAGGCGAGGGATGTAACCCAGCCACGGTGACCCGTCAGTTGTCCTTCG
TAAGCGACAGCCATGATTGTGGGCAGAAAATAGAGCGGTTAGCCTTGCGCTACTCAG
CTTGCAATATAGTACCTGCCCGGGCGGC

Contig 93

AGGTCAGAGAACAAGGAGGACTACAAGAAGTTCTACGAACAGTTCAGCAAGAACGTG
AAGCTGGGCATCCACGAAGACAGCACAAACCGCAAGAAGCTGATGGAACTGCTGCGC
TTCCACAGTTCGGAATCCGGTGAGGAGATGACGACATTGAAGGACTACGTGACACGC
ATGAAGGAGGGTCAGAAGTGCATCTACTATGTGACGGGTGACAGCAAGAAGAAGCTC
GAAACGTCTCCGTTCATCGAGCAGGCGAAGCGCCGCGGCATGGAGGTACCTGCCCGG
GCGGCCGCTCGAAATCACTAGTGAATTCGCGGCCGCCTGCAGGTCGN

Contig 94

CGGCCGAGGTACTAATAAACCAAAAACTACGATATAATTCCTACAGTAGTGACCAAC
TTACTTCATCTCCGCATATTGAGGAGCTTGACCATATCTGTACCTGCCCGGGCGGCC
GCTCGAA

Contig 95

TAGCGTGGTCGCGGGCGAGGTACACCACAACCAGCCACTTTTGGTTCTGTAGATGGA
GCAGCTGACATTTGGGATCTTAAAGATATTCAACTGCTCAATGGTCAAAAGCGGAAA
CCATTACTCAAATATCATAACGGAAATGCAACATCTGCAAACTCTAAGCCGCGTGTG
GGGGATTTGTTGATTTACCCACGGCAACCCAATGGTTTTCCATGCGGTCATGTTGCC
GTTGTTGTGGGTGTGACAGGGGATCGTATGTTTGTAGCGGAACAGAACTGGGAAAAC
GCAGCATGGCCAGGCCCGTATCATAATTACTCGCGTGTGCTTAACTTGTCATGTAAT
CCCAATGGCACGGCATGTACCTGCCCGGGCGGCCGCTCGAA

Contig 96

AGGTACATTAAGCATGATACAACTGCGTGGGACGACCCCAACAGGTGTGCGGATCCC
GATTATGTGACCCCAGACGATGTTCAGCAGTTCGTTGAGGGAATGATGAAGTCCGCC
TTGCTTTACCGTTGGTTTCTCAAGGATCAGCGTGTGCGCATTATGCCCCCCGGGCGC
CCGTACCTGCCCGGGCGGCCGCTCGAAATCACTAGTGAATTCGCGGCCGGCCTGCAG
GTCGA

Contig 97

GCGGCCGCCCGGGCAGGTACACTCTTGCGTGAATTGGCATTGCGAATCTCTACCGCT
TCGTTAGGGTCCAACCACACGCGGCCGCGGCCGCAGCGCAGGATGTCCGCTGCCAGG
CTAGCGTGGTCGCGGCCGASGTACGTTGCAGTTTTCACGAAGGGCACCTCAACGTTG
TGAGTTGAGGGGGCCTGATCTTCCATCTTGTTTCCGGTGAAAATGTCAAGAGTAACG
ATGCTCACCTTTGCATGACCGTGTTTTCCAGTTTTTGAGACTGAAAGATCGATTACC
TTGCGCGGACGTCCATTAATGCAAATGTMKCCGCCCTTTTTCARAGCGCCTGCCTGC
ATCGGATACGTCWCAGAGCCCACCTGCGCTCCGCCTTCGGCAAACTGTCCCTCATCG
TCAGACATTGTGTAGGATTTAAAGCAGAAGAACTCACGGTCAACCTGCCCGGGCGGC
CGCTCGAA

Contig 98

TTCGAGCGGCCGCCCGGGCAGGTACAGCTAGCTGACCTTGCTGCCGTTGAGGCTAAG
GGTTAAAGGAAAAAATGGTTGTTAAAACTAAACGGTAGACGAAGAGATACTAGTAGA

Contig 99

TTCTAGCGGCCGCCCGGGCAGGTACATATACCGGCGCTCCGCAAGCACCATGAAGAA
GTTAGGGTTGGAGCTTGGTGGAAACGCACCATTTATCGTGTTCAGTGACGCTGAGGT
GGATCGCGCGGTCACTGGACTCATGAATGCGAAGTTTCGTGCTGCAGGACAGGCATG
CATAAGCGCCAACCGTGTTTTCGTGCACTCCACGGTTTACGGTGACTTCATGACGAG
GTTGCTGGAGAGGGTGAACTCAATCCGTGTGGGGAATAACTTTGACCCGTCTTCAAC
GATGGACGCACTTGTCAGCAGTT

Contig 100

TTTCNANCGGCCGCCCGGGCAGGTACTCACGTTAACTGCAGATGTCACCCAGGGCAA
CAGAGCCAATAATAAACGAGGACGAACACCCAACCTTAATACCCAGCTTGTCTCAGG
GTAATGTGTTGGAATTACCTCCAGACAGCGATAAATCTACAGCAGCTGCTGACGGGG
AGCACAGGGGCTGCCTGAACACGGTATTTGACCCCATCAAGGGAGTTGTTCCGTCTG
GTGGAATGGCGTCAAATGTGTTTAACCTGGAGAGTGCTACCCTTGGTGCTGGAATTG
TGATGCTGCCTTCTGGTTTCCTCAACTCTGGCATAATAATTGC

Contig 101

GKTTTTTTTTTTCRGTAAAAAAGAAAAGAAATAATATGAAACACGGCAGTTTTCTT
GGCAGCCTTTCCTAAGGTGGAGTTCCTACGCCTACTTGGTGCGGRACACACCGACAG
TCTTACCAYGGCGGCCAGACGTGCAAGTATGYTGACCACGAACGCGCAGACCATAGG
CGTGACGAACACCACGGTGCGCACGGATCTTCCTCAGCCGTTCCAGGTCCTCACGCA
AGCGTGTATCCACCATCGAGCTAGTCAGGTGCTCCGTCTTTCCAGTTTTAGGGTCGC
GCTGACGATTCAGGAACCACTCGGGGATTTTGAACTTCGCGGGATCGGCGATCACCT
CAGAGATTTTCTCAAGTTCCTCAGGTGACAGCGTACCTGCCCGGGCGGC

Contig 102

TAGCGTGGTCGCGGCCGAGGTACTCTGAGGCTTGTAGGAGGGTAAAATAGAGACCCA
GTAAAATTGTAATAAGCAGTGCTTGAATTATTTGGTTTCGGTTGTTTTCTATTAGAC
TATGGTGAGCTCAGGTGATTGATACTCCTGATGCGAGTAATACGGATGTGTTTAGGA
GTGGGACTTCTAGGGGATTTAGCGGGGTGATGCCTGTTGGGGGCCAGTGCCCTCCTA
ATTGGGGGGTAGGGGCTAGGCTGGAGTGGTAAAAGGCTCAGAAAAGTCCTGCGAAGA
AAAAAACTTCTGAGGTAATAAATAGGATTATCCCGTATCGAAGGCCTTTTTGGACAG
GTGGTGTGTGGTGGCCTTGGTATGTGCTTTCTCGTGTTACATCGCGCCATCATTGGT
ATATGGTTAGTGTGTTGGTTAGTAGGCCTAGTATGAGGAGCGTTATGGAGTGGAAGT
GAAATCACATGGCTAGGCCGGAGGTCATTAGGAGGGCTGAGAGGGCCCCTGTTAGGG
GTCATGGGCTGGGTTTTACTATATGATAGGCATGTGATACC

Contig 103

TAGCGTGGTCGCGGCCGAGGTACGAGCTTATCAGCAAGGGACTCAAACTTGCTGCGG
CTCACCATCATCTGCACGTGCTGCGCTCCGTCCTGGTTAGCTGTGATAAACGGAAGG
TTTACCTCCGTCTCCATCGTCGTCGAAAGTTCACACTTCGCCTTTTCTGCAGCCTCA
CGAATACGTTGCAGTGCCATCCGCTCCTTGCTCAAGTCAATTCCAGATGTCTTGCGG
AATTCCTCCAGAATGTGATCAGAGAGGCAGAGATCAAAGTCTTCTCCACCAAGGTGA
GTGTCGCCATTCGTTGCCTTTACCTCAAAAACACCTCCGGCAATTTCAAGTACCTGC
CCGGGCCGGCCGCTCGAA

Contig 104

TAGCGTGGTCGCGGCCGAGGTCTCGCGAAGCACGCGATGGCAGAGGGGACGAAGGCT
GTGTCACACGCTTCCAGCTAAACACCCTACATCAGGAGACCTTTGGTTTGCCACCTT
GACATTAGATCCCTCAGTGCCCCTCAGGTGAGGCATTGATATTCCTTTTCTTTTTTA
TTTTCCCCCCACCATACCCTCTCCTTGTGGTGGTGATAGATTTCACTGTTTGCCGGA
ATGTGCCGCTGCCTTCGTTCTTTTCGTGTGTGTGTGtgtgtgtgtgtgtgtgtgt
gtgtgtgtgCGTGTGGTCCACGCGCTGACACGAAGGCATTCGCATTCGCTTTTGGCA
CCGCTAGCGCATTTCACAGTGTGTATGTATCGGTGTGGCCGCAGGCGCTATGTACCT

Contig 105

TAGCGTGGTCGCGGCCGAGGTACCCTCATACTCTGCTGTACCCTTCCTTCATTCTTT
CGGTCCTTCCTCTTCTCTTTTGTGCCTTCACTCGCCCCCCGCCTCGTCCCGCCACCC
ACTTGTCTACGGCGACAGCAGTGGGGATAGAAGGAAGACAGGGATGTAATGGGGGTA
AAGGGCCGATTCCCTTCATGCCCGCATTGCACAGACACCTCCCGTTACTCACCAAGG
TAGGCAGCAAAGTCAACTGGAAGCTCTTCAATTTGCGTGTGATAATGCGATTCAATC
TCCCGCAGGACCTCCACATCCTTCTGAGTGACGAAGTTGATTGCCACACCCTTACGT
CCATACCGACCACCACGACCAATACGGTGAAGGTAGTTTTCTTTGTTCGTGGGGAGG
TCGAAGTTGATGACGATGTTCACGTGGTGCACATCGAT

Contig 106

TAGCGTGGTCGCGGCCGAGGTACATGACACGAGACGGATGCCATGCGCCAATGCACG
CAACCTTACGTAGACCACGGTGAGTTTTACGTGGTAGGCACGCAACACCCCAACGTT
TCACGACACCAGTGAAACCATGGCCCTTAGTCACAGCGCACACGTCACATGCCTCCG
ATTGTTGGAACACAGAGTCGATGCGAACTTCCTTCTCAAGTAGCGATTTGGCGAGTT
CGATCTTCTGCGCAATGGTGCCGCCGTTAATTTGGATCTCGCTTACGTGAGCCTTCT
TCACACCAACGCGGTTGTTGCGCAGCTTGCGGAGTTGCGTGTGTGCCACCACACGGA
TGATGTCAGCCTTCTTTTCGAATGCCTTCAACGTCCGTGCTTCAGCGAGACGTCCCT
CTGTTGTGCGAGCGAACTGCTTGCGCTTCGTGAAGGCGAGCTGTGCGGACTGTTTCC
AGTTCTTGTAGAAGCGCCGGCGGAATTCCACGCTGGTGTGGTGTGCCCACACGGTGC
CAATCGTCTTGTGGCCCACAGGTGTCTGCCGATAGCCAACAA

Contig 107

TAGCGTGGTCGCGGCCGAGGTACCACTTGTTTTTTTCTTAATTTTTTTGTATTTAT
TTCTTTTGGAGCATCTTCCGAGATGTTTATTGTGTCCACAGAGGACGGAAATTTTTA
ATCCTTCGACCCGAGTCTATGCGCACTCTTTTGTGCCTCTTCAGGCGCATCGGCGGT
GTGTTAGCTGTGTGAAAAATAACGTAGAGAAGGGACTGTGCTAGTGGACGGGTAGGG
AAGCATAAGTTACTGCGCAAAATGGGAAAGGGTATGACTAAAATAACGTATACGTAA
TAAACGTAGGAAGCAAAACGGGGACGGGAAGACACCGTAAGGATGCTGTCATCTTTT
GCCATTTATTTACTTTTGTTGTTGTACCT

Contig 108

TAGCGTGGTCGCGGCCGAGGTACGTGTTTGTTCCCTCAAAGACGGCATTAGGGCGCG
CGGCGCAAGTGTCACGGAATGCGGTGGCTCTAGCGATCCTCCAGATGGAGAACAGCC
CCATTTCGGCTAAGATACAGGCGGTTAAATTGGAGATTGAACGGCTCCTGTAGTGAT
GCGTCTTCACTGTAAAGGGGATGCAGACTGAAAAAAGGAAAGCAAAGTACCTGCCC

Contig 109

TAGCGTGGTCGCGGCCGAGGTACAAGAAACCTGAGGTTACCCAGAAAAAGGCAAAGG
CTTCTGCTGCAGCACCATCACCCTTCGTGGCGCGCCCGAAGGATTTTGGTATCGGGC
GTGATGTCCCTTATGCTCGTGACCTTTCACGATTCATGCGGTGGCCCACGTTCGTGA
CGATGCAGCGTAAGAAGCGCGTGCTACAGCGACGCTTGAAGGTACCTGCCCGGGCCG
GCCGCTCGAA

Contig 110

TAGCGTGGTCGCGGCCGAGGTACAAGCTTTTTTTTTTTTTTTTTTTTTTTTTTG
KTGCTGAATGATATTTTATTARCTTGATAATTTGGGCCTGCCCTTAGCATTAATAAG
CTTCAGCACTAGTCACAAGACTTTCATTCACTGGTGGGGAAACTTTCTTGTTTTAAA
AAATGCAATTCAAGAAAGGGCATCTATTTCTTGGGGGCTGCGGTGACAGCAGGCTTC
TCTTCACGGGTGATGGGAATGGTGCGCTCAGGGCCAGAGACCTGTTTCCTTGGTCCA
TTCACAGTGAGGACCCCATCAGATGACAGGGATGAAGTAATGGTGAGAGGGTCTACA
TCAGCACCTGCCCGG

Contig 111

TAGCGTGGTCGCGGCCGAGGTACTGMTGMGGAAAAGGCTGCACAAKTYATATARAAT
CATTCYTATCAATTAAAYCTTGTGGCTTASATTTCATTGGTGGAGC

Contig 112

TAGCGTGGTCGCGGCCGAGGTACTAGTTGAGAGGTGYAACKCGTCGTGCTAAAGGTA
ATGTCTATATKASATCACAGKATATTTCTTTTGAGCGCGCACAAAGAAAAASC

Contig 113

TAGCGTGGTCGCGGCCGAGGTACTCCTCCTTTGTTACATCTTTCGGGTCGCGCGTCC
ACAGTGGCTTGTGCTTGTTCTGCACCACAAACTCCTGCTTCACTTCCTTCACCTTTT
TGGTTTTCTTTTTCTTGGCATCGGCATCATCGCCGTCCTTCACCTCTTCCACCTTTG
GCTCCTCACCTTCCTCAGCCTTTTTGGCGGCCTCTTCGTCTTCGTCTTCATCTGTCA
CCTCCTTCTCAGTGGTGTTCTCCACCATTAACTCGATGTCGTAGCCGATAAATTCGG
AGTGTTTTTTGATGAGGTCCTTCAGCCGCCGCTCCTCCAGGTATTCCTGCTGGTCTT
CCTTCAAATGCAAGACGATACGCGTACCTGCCCGGNN

Contig 114

TANTCGTGGTCGCGGCCGAGGTACTTCATCCGTCTTCTAGCCGCTCCTGAATCAAAG
CTATCTTTCTCGCTCCCCCGTTTCTCTCATACTTTCCTTTAAAAAAAAGACACTGCG
ATGTCTTATCACACACACACACGGCGTTTCACTACTCCTTGCGCTTCTTCTCTAGCG
AAAACGGAAGCTTGCTCGTTGTCAGCAGCAATTCGTCGGTGTATGGAGGAAGCTCTG
GCGCCTTCTGCATTTTCAAGCCAGTCACAACTTCAATCTTTTTGGCTGTCACGAGCA
ACG

Contig 115

GGTGTAAAGCTTTTTTTTTTTTTTTTTTTTTTTTttkTGAAAAGAAAAAATAAAGTA
AAATAAAATTATTCCTTCACCACATAATGGGCATTATTGGTTTTCTCCTCTCTTAAC
GCTTACTTCTTCTTGAGTGCTTTCGCGACATTTTCCTCGCTTTCGCCAACCGAACG
ACCTTGGCGGCGTGGCGAGCCTCACGGTTCTTCTTGTTCTGCATGCCCTTAGCCATT
GCCTTCTCCTTCTTCTGGAACGGGTTCAAACGCAGGCGYAGGCGACGGTTGCGAATA
CCATTCGTTGGCTGCCTATAGCGGCTTGGCCTCTTGGGTTGCAGTTTCTTGGGTTTC
ARAACGCGACGCACCTCTTCGGACTGCATGATGCGCGTCACATCAGTACCT

Contig 116

GTGGTCGCGGCCGAGGTACACTGCTGTTTCTAGAAAAACGAGACGGAGAAGGTTTCC
ACAGTAATGATAAGCTAAACCATAGTTTGAATAGGTGGGACGAAAATACCGATGAA
AGGCGTTCTCTGAATCATGCGGATCTCCGCCACCAAAGATGCAAAGCGGTTCGGTTA
CATCTGGTGATTTCCCTCAAAAGTAGACAAATTACTGTCACCGGACCTTCCCGTTTC
TTCACCACCCACCTTAGCGTGCCGTATTTGTAGGTGAATGTTAACTTTATAAGGAAC
CTCCTTCTGGGATGCTTC

Contig 117

TATTCGTGGTCGCGGCCGAGGTAACTCATCAGGAAGGCAAGCCACTGGATGGGAGAG
GGATGCCCAAAAAAGAAACACTGCAGATACACATCAAAAACGCTTTTGCGCCGTCCA
CACCGTCCTCAATATTGTGAATTGCCGCCGCCGATCTGTCCTTCAATTCCTTTTCAC

ATTTCTGTGTAACTGCTCGACGGGTTTTGTGGAGATGTGCGCCCATTTGACGGTCTC
AACTAAGCATTGAGCCGCCCCTCGAGTAC

Contig 118

CAACGAGNACGAAGGGCATTCCTCACCTCTCCCCTCCTTCAATACGGCCATCCAAGC
TGCCCGTGATCACCTTGTTTTCTTCCCAGTGCCGCAAAGGCGCTTTGCATCGGTATT
ACGCTCTTGTTCCTCTTTGTATGGTGCTACGGGCGCCCCTATAAGATGACAGACTCC
GACGCACCCCTGCTTGCGTTAACGGCGCGCCCATAGTACCTGCCCGGGCGCGCTCGA
A

Contig 119

GTTTTTTTTTTTTTTTTTGACAAAGTAAAAGAGAAAAAGAACAAACAAGTGGGTCA
GGTAACCTTAGTTGAAGACTCAATCCTCCCCTCTGCCCAGCTCAATATGGAATCATT
GCCTCGTAGTCTGCCGTGCGGCGCGCCAGCACCTCCACTTTCTTGGGGCAAACAAAT
CCGCCGTTGCGGGTGCTCACCTTTTTGGGTGGAAGTGGGGTTCCCGCCTTCTTAGCA
GCCCGAAATGCTTCAAGGTCCTTTAGCTTCTGCTTGAATGCCTCCTGGCAGCGAGAC
TTACGCACGTGCTCAAAACGCACGCAGATGCGCTTGCGAACTGTACCTGCCCGGGCG
GT

Contig 120

TANTTGTGGTCGCGGCCGAGGGACAAGCGCACTGCTAAACCGAAAGAACAAATCCAA
AAGAACAAACAAGACATTCAACTCATATTGTAATTGACGATGTCTGGCCTCGCTAAG
TATCTTCCTGGCGCAACCAACCTGCTGTCCAAGTCGGGTGAAGTTTCACTGGGATCC
CTCGTTGGGAAAACTGTGTTTCTTTACTTTTCTGCCTCCTGGTGCCCCCCATGCCGG
GGTTTTACACCGGTCCTCG

Contig 121

TTTTTAGCGGCCGTCCGGGCAGGTACACAACGATACAAAAAGTGATACGTAGCGACC
AAGCCGCAAGTGAACCAACACAAAGTGTTGGGAAAAAATAATACAAGTTTCAACAAG
ATGCACTCCCCACTGCCACTCCGGCTATCCACGAGCAAGCTACGCAACTGTAGAGTC
AACGGCCAG

Contig 122

TATAAGTGGTTTGATTTTGGGTTTTCTGCGCCTTTTCCCGGGCGTGCATACATGTGC
ACGTAACCTAGTAAGGGTGTCGCTTGCAATTWTCATTTGATCAGACGTK

Contig 123

GGTACAGGTATGGGGACCAGTGCAAGCGGCAACAAAGAGGAAAGAAGCCAAAAGAGA
GAGAGAGAAGGTAGGGAGAGAAAAAAAGGGAAGGAACGTTCAGTGATTGGTATTCCT
CCTCCCCCCCCCCCAAACACACACACACACACGCATATTCCTTACCTATTTCGCTTA
CCCATTTCTTCTTCTTCTCATTATTATTGTTTTACCTTTCCCCTTCCTCTTTATCCC
TCTCTGGCCATTCCATGTATGTATGCCCAACAGCGCCTTCTTTTACTTTGCCTTTGT
CTCAGCGCCCATCTTGCGGAGGGCAATCTTGAAGGCAATGGTGTCTGCCACCTTGCG
ACCGAGTTTAGTA

Contig 124

AGGTACTATRKTGTGGAACGAGCACATTTCACTCACACAAGGATTTGATTAAAGGAA
GAAATGTCGCAGCAGCTTGCATGCACATACGCCTCCTTGATACTCTCTGGCAGCGGG
AATGTTGACGCCGCGAAGCTGCTTGCCGTAACAAATGCTGCTGGCGTTACTGTCAGT
AAAGGTATGGCAGAAGCTTTTGCCTCAATTCTTGGTGGTATTTCCATTGATGAGGTT
CTTGGCAATATTGCGTTCGGTGGTGGTGCGCCTGTTGCGTCTAGTGGAGGCGCCGCT
GCTGCTGCTCCTGCCGCCGGTGGCGCACCTGCTGCCGCTGCTCCTGCAAAGGAGGAA
GAGGAAGATGATGATGATATGGGCTTCGGTCTTTTTGACTAAGCGCGCTCCTCGCAG
TGTGATTGTTGCTCGCACAGAAAGAGAGTTTGTTTCATTCCTTTTTTTAATTTTTAT
TTGTTCTTTGTTCTCAAGTAACTGGTACCTAAAGCCGCGACCACGCTAATYT

Contig 125

CCGGGCAGGTACCAGGTGCTTTTTGGGAGAAGCTTTCAGGAGAGCTGCCTTTTGAA
GCTTCTTAACTTCATTCTTGATTATTCTGTTCCTCATTTTCTTTGCCTTCATAACTT
TAAAACGATCAAAATCTGTCATCTTGGCTTTCCTTTCTCTGGCTTCAATCTTCTTGG
CCCATCGTGTGGCTGCCCATTTTGTATTGATGTCTGCCTTCTGCCAGGCTTGTCGGA
CATACTTCTGGTGGGCACTGTGCGGAAACTTGAGGATGAAATCAGTGAGCTGCATGC
ACTTGAAAGGCATGGCCTGTCTCCTCACTTGAGTGCAAGGTCCATCGACCAAAGCCC
TGTTCTGATCAATAACATCTACAATCGCGACCAATTTTCCGGCATGAGGTCCAAAGG
AGACATAGGCCACCCGGCCAACCTCCACGAAGCGCCTGAACACCAACCT

Contig 126

TACAACTGGTGACAACGTTGGGATCAAGACGGACGCTCCTTCGGCGATGGCGTTGGG
GCCACTCAATACAAAGAGACTGAGCACCATCCCGGCACCGATAGACGGCATCCACCA
GTTAACCTTCGTTACGAACCAATCAGCAGCATGGTCGAAAAGATGGTGAAGTTTTTG
ATAAAAATACTTCTCTTCCCAGGTTGCCTGGGGGACGGCGGGGCGACCGAACATCAT
CACTGCAGAGGACTTGTTACACTAACGAGAAGGTAGTCAATATAGTACCT

Contig 127

GGTACGCGCCTGCACAAGGACATCTTCACCACGCATAATAGCTTCGAATGTTAGTGC
CTGCACGGGAAACAGCGAAACAATACCTTGAGCCTCCAGTGCCTTCACCATCCCGCT
GGAAAGATTAAACTCAGAGAACGGCCGACCCACCGCGGCAGAGCCGTTGTTGGGCTC
CTGATCAGTCTCTTCATTCCCAGTCATCTCTTCGCGGGAGCGTTTCGTTTTGGTGAC
GTGCCCTTCATGCCTGTGAGTCCCTTCCTCAGCTCTGGAAGACATGCTTGTCAGTGC
CTTTTATACCTTAGTTTAGTGTTACTGTTGTTTTGTTTTGTTTTTTAAACCCACTCC
TTTCACCACTTTATTCCTAATTCGTCTCCGACGCACCCTCTTTACGTCCGACTCTTT
ATGTCACTCCTCACCACTTCTTTCTGATTAACTTGACCGCTGTAACAAATACAGGGT
AGTCGCGTGGCACCGGGAACCTCAGCCTACAGACCAATATAGTACCT

Contig 128

GGTACATAATGCCACATATCGGAAGGTTCCATCCCCTGTAAGTTCGTAACAAAAGGG
GATGGCGCTCTGTTGCTCCACCGCCGTGGTTAAAACAAATTCTGGCATATCCGGTAA
CGCCCTAACAAAGTAGCCAAGGAAGGGTGAAGGGCCTTCGCCTTCTGTGGTGCCCGA
CCTAAACGGAGAAACATAAGACATGATTGATCGACCCACACGATCGGCTATCAACAG
TAGATCATGTGAAGGATTTTCATATATACCGCTGGCAGCATGCATGAAAACCGGTGA
TGTGACCGGATAACTCGGTTGTCGCTGATGAGGGGTGTTAAAGTCACACCT

Contig 129

GGTACTCTTCATCGTCACAGTCCGTAGCGAGAATACACATGTGCGCGGTGCGGCGGT
CCAACGCCCGAGCAACCTCAGAAAGGCCGCGGACAAGTCCGTTTACCTCGAGCGCCT
TACGTATGACGATGCGAATCGCGTCTTGCAGTGTGTCGGGCTCCACATCAATCACCT
CCACGACAGGCGTTGGTTCCACCTGTTTTTCAACCTGATTCTCGACCTGGTGTGCAC
TCACTTGGATATCACACCT

Contig 130

GGTACATTCACGATACCTCTGATGCACTTCAGTGGTTTCGGTCAGATGTATCTATGA
CCAAGAGCGGTGTGTTCCTACTTTTTGGGGACGGTTAGAATAACCTAGCACACGCAT
GTTACAATAAAGGCATGCGCAAAATGAAATGCAAAAGAAATGCGTTTTCACTTTACA
CGCACGTTTATTGTTTCATTTCCCTTACCTACTTTAATACTCATTTTCTCGTTTTCT
CACAGTCCTTCACCTGTTGATTGCTGACCTTTGACCCTTTACTTCTTTCCTTTTATT
TTATCTTCCCTTTTTGTTGGTTCCCCTCACCATATCACCATCACCTCGTTGTTGCCG
TTGGTTTCACTCGTCCCTCTTTAAGTGTCTACTTTGATCTGCTGGTGGTCTCTGGAC
ACTCGTCAGGTTGACCTCGCGCAGCATTCCAGAACCACCATCTTTTCACTCCACACT
TCTCCCCACATCTGCTTCACTTTCCAAATTTCACATCTCTTCGCTCCCACACCCCCC
TCTCTTCATCCACTTTTCTTGCAACTAAAGTGGGGTCACTATACCT

Contig 131

GTACTGACTCCCTTTTGACACCCCAATTGACGTGCATAACTGACATTCAACAGTTTG
AAAGGTTTCAGAGCCTCCGGTAATAATTTTTGACTCAATATTTAGCGGCCGTCTCCA
CCCAAAAGAGAAGCCTCACTTTCAGGAATGTAATCGAGAACATCGATCCCATCTTCA
AAGAACTCGTCGAGGAATTCGAAGTCTGAGTCAAATAATGCACCATAATCCTCATCA
TCATTCGCTGGTTTTTTTTTGTCCATCACAACCATTAGTCCCACACCTCACGCTAGT
GGCAAGTAGTGGCCGTGACATGCTACGCATATGGACCAAGAAATTACTTGTGATTGG
GACGTGTAACCTCCGGAACATTTTCCCTGTTTCTTTTTTAGACAATTAACAATATA
GTACCT

Contig 132

ACAGCCACAGAACTAATCATATTTTATATCTTCTTCGAAACCACACTTATCCCCACC
TTGGCTATCATCACCCGATGAGGCAACCAGCCAGAACGCCTGAACGCGGGCACATAC
TTCCTATTCTACACCCTAGTAGGCTCCCTTCCCCTACTCATCGCACTAATTTACACT
CACAACACCCTAGGCTCACTAAACATTCTACTACTCACTCTCACTGCCCAAGAACTA
TCAAACTCCTGAGCCAACAACTTAATATGACTAGCTTACACAATAGCTTTTATAGTA
AAGATACCTCTTTACGGACTCCACTTATGACTCCCTAAAGCCCATGTCGAAGCCCCC
ATCGCTGGGTCAATAGTACCT

Contig 133

GGTACTACACGCACAAAAGCACACTTACAAAGCCTACAAACCCACTTGCCCCCTTTT
AGAGAATACCACTCTCTAGTCAGTGATAGGAGAGGTGATATTAAAAGTTTTCCCTTT
ATAATTATTTCTGACTCCTCAATTCGTTGCRACATAACATAACATGCAAAGATGTTG
CATCAGGCAGCGGGGCCACCAGCAATAGGGTTTGGTTCATCCTGGGTGATCTCCTTA
ACAATATTCCCTTCCATCTTCACTTGAAACTGGTTGATTCTATTCCACAACTCTGTA
GTATAATCGCCTCTGCCCCCAACTGGTTCGCACTTCTGCCTCTCCTCGATTGGCTTG
CTGTCACTGTACCT

Contig 134

GGTACAAGTAAATATTGTAACTAGTTAAAAAGGTAAGAAATGAAAAAGATAATCAAG
GCGGGGTTTGAAGAGTAGTCAAGTGAAGACCAGCCGGCTGAACTCATACGATCAACC
ATCTTCCCCGACACATCATTCTCGCGTGCCCTTTGTTTTTTCATTTACTACTATGCC
TTCTTCGATACGTCGTCCTTCGTTATAATGTTTGTGGTGTCCTTCAAGTAGTCCTTC
TCTGCCTCAAAGAGCGGGCCGAGCTGCGACCT

Contig 135

GGTACGCCGTATGGAAAAGCGAGAGCGAAGCAAGGTATTCAATATCAACATCTGCGC
GATTTGGCACGACGTCCAAATTTTCAACATCCGTCACAACATGTGTGGCCCATGTGC
CCACCGTTGGATTATTAACCCATACAAGGTCGCCTTCCTGAAGCAGCAGTGAGGCCC
CTTTACCAACTTCTTCAACAACTCCAACACCCTCAACACCCGCAACCTGCGGAAATG
CGGTGGGTCGAAGCGGCCCACAGTACCT

Contig 136

GGTACGTGAAGCACGGCCTCTGGCGAACTAAAACGGCGAGAAAAAACTGAGCGGCCC
ACGAGAGTGTTGTCAGAAACATGCTGTTCGTCGCTTGGCGTCTGATGACGTCGAACG
AAGCATATTGTTCAAGGTTCCTAGGATGCCAGAAGCTCCTTGAAGTAACGCACTGCA
TCACATTTTTCAAACAAAACACCCGTCTCCTCCCGCAGGTGGCGGTTCCATTCCGAA
GTTCATAATCACACTTGGGTACCT

Contig 137

GGTACCGGTGCGACGGCGTTTGATGGCCTTGACAGRCCAGTTGTAACGTCGGCGCTG
GGCACGAGGGTAAGCGCAGGCAGCGCATCGCTCCCACTGCACATGGTAGGCGTTGCG
CCCACAACGGCGGCACAAAATGTGCGTGCGCCCGTGGCGCTGCCCCATCGACGTGGT
GCCCTTTGTCATTTCGTTTCTCTTCTCACAGCTACAATATAGTACCT

Contig 138

GGTACCCACGCATCACCATTATTGCTCCGAATAATCTTGTAGGGCACATTCTTGATG
TCGTGTTGGATGTGCTCATCATCGAAGCGACGACCAATCAGCCGTTTCACCGCGAAG
AATGTCGACTGTGGGTTCGTAATCGCCTGACGTTTCGCTGCAAGGCCAACAAGTTTT

TCCTGTCCTTTGAACGCAACCACAGACGGTGTCGTCCTGAAACCCTCTGTATTCTCA
AGTACCT

Contig 139

AGGTACAAGAATAAGAAAAGACTGCACAAGACATTGGCATCAAATAACTAGTTGCAC
TCAATTTACATAATAATTGCGGCCACACTATTCGAGGCGTGCCGGTGCAGTGTTGTT
TCCGATGCGGACCTGCTCGCCACCGGCACCAGAGCTGCGCAGTAAGCGCGCCCTCTT
GATTTCCTCGCGTTCGGCGGCAATCTTCACCTCCTCCTGCTTCGTCAGGTGCGAGCG
GTACCT

Contig 140

CGGCCGASGTACATTTTCTTGGTGGATTCCTCCTCCGTTACCTTGCGTGGGGCCTCC
CGTGCGGGGGTAACCAAACCTCCAACAGCAGCTGTGCCATACCGAGAGCGGKCCTGA
GTTGCAGCTTTCACCTCCTCTTCAGTTGCATCTCCCTTCTGTGGCTTCTTGCGGTTC
AGGGGGAACAGCACCAACTTGCTCATATATGTCTTCAGGCGTTGCACGTTAATATTC
ATACCCTCCTCGCTTTTATTCTTTCGGCGACGGTCAACACGAATGCCAATCGTCCTC
GCATAGCGGGGTTTCACACCAGCAGCCTTCAGCTCCTCCAACGAGAAGCCACGTCCC
AAGCGCCTCTTCATGTTGTAGCGAACTGTAGGGCAATTCACCTGCGGGCGAAGTGCC
TTAAGCGGACGCGGGAAGATCTTCTTCGCCTTCAACAAACGCAAG

Contig 141

GGTACCAGAGAGGAGCCAGTGCGAGCTTAGAARCGAAGGCGCTGGAAGAACCAACGG
TTGTTACCGGCGTTGTAGCGGGCCTGAAACAGACGTTTCACCAACTGCTTCGACTTC
GCCTTCTTCGATGCGTCGGAAACATTGATGCGACCGCGCAGTTCACGCGAGAGGTCC
ATGTTGTATCGGGTTGGTAGAAAGTGCTTGTGATTCACAACGCGGAGGAACACACCA
ACTTGCGAGCGGCGTGTGATGGAGCGCTTGCTCATCCCGCGCACAACCTTTCTTGGG
TATTTTTTAATTCCGGCGAGCAACGAATGACCGTACCT

Contig 142

CATAACAATTGCTTATGGYGTTGGCGTTGMCACTACACCGCACAAATTWTTTCAGAT
GTCKGCACGGCCWGATGCGGAGCTTATCCCTCTCGTTTCCCTTTCCTGCGTTGCTTT
TACTACAWGTART

Contig 143

GGTACCTGAAGCGGAGAGCGCTAAGTTGCATGCAAGAGGGTGATGGCGGTTTATCGG
CCGTGTGCAGCTTGTTGATTCCTGAGGTCTTAACGTGGATGAAGTGGGACAAAATGA
GAGTGTGTAAATGACGTGTGGCGAACCGACGTATCGGATTATTTAGGACTACTTGTA
GTAAAAGAAACGCAGGAAAGGGAAAGGAGAGGGATAAGCTCCGCTTGTGGCGTGCCG
ACATCTGAAAAAATTTGTGCGGTGTAGTGGAACGCCAACACC

Contig 144

GGTACGAGTGAACAGTGGTCATGAGGCCAGTGGAGATGCCGAAGCCCTCCTTCACCA
ACACGTGCACGAGTGGGGCGAGGCAATTAGTTGTGCATGAGGCGTTCGACACCACAT
GGTGTTCACGAGGGTTGTAGTCGTTGTGGTTCACGCCCATCACGAACGTCTTGGCGC
CACCAGAGGCGGGGGCACTGATGACAACCTTGCGAGCACCACCACGGAGGTGACCCT
CGGCAGCAGATTTCACTGTGAAGAGGCCAGTTGACTCAATCACATACTCCACACCAA
GCTTTCCCCATGGGAGGTCCGCAGGGTTCCGCTGCGCTTTCACGCAAAGGATGCGGT
GGCCGTTGACGACGAGAGTATCATCCTTCGCGACGGATGGCTTGCTCTTCGTAGTCG
ACACAGAGTGCTTGAACTTGCCGTGCACGGAGTCGTACCT

Contig 145

GGTACGCGATTTTACACACTTATTCGTGTGCATTTGTCGGTGTGGTGGTCGTGTTGC
CGGTTAGGGGTTTCAGTTTTTCCATTATTTTAACGCTCCGCATTTAGAGCGTTCGGA
ATACTCATCCCGCCCCGTTTTGCGTATCTCTCTTGCCTTATTTTCATTTCCATGGAA
ATGTTGAGGGTTCTTCGCTTACACAAGTGGTACCT

Contig 146

AAACATGATGAGGAGTGGTAATAAGCAATCGTAATGCATCATGATAAAATATTTGTG
GGTAGATGACTGCGGAAAAACGTAACACAGCTGGGTTGGTGCTTTTGAAGGAATGCT
GTTGGCTGCTGTAACGTGAAAATAATGGTGCACCGAGTACCT

Contig 147

GGTACTATATTGGTTGTAGTGGAGTAGGTGCACAAACCATGTCGCACATCGTCAAAA
TCCGTGACCTGAAGGAAAAGGGGAAGGATGACCTGCTGAAGCAGCTTTCTGAATTCA
AGAAGGAGTTGTCGCAACTTCGTGTCTCTCAGCAGATGAACGTTGGTGCTGCTCGTC
TTGGTCGCATTCGCACCATTCGCAAAGGCATTGCCCGCATAATGACGGTGCTTAACA
AGAACGAACGTGAAAACCTCCGCAAGTTTTATTCGGACAAGAAGCTGCGGTCGGCGA
AGCCCAAAACACTTCGTGCAAAGCTGACACACCGGCGTCGCCTAGCACTGAAGGCCA
ACGAGAAGAACCGTAAGACCCGCCGCCAACTCCGCATGGCCCACAAATTCCCGAGGC
GCATATACGCCGTGAAGGTTTGATGCGACATATTGTGACATACTGTTT

Contig 148

GGTGCACAACCCGTGGGTCTCCCTGAGGTTGCGACAAACCACCTCCTTGCGGTGCCC
CTGTCAAGTTCGGGATTGGCTGTCCAAAAAAGGGAGTTTGAGGGGCGGAAAGGGGGC
CTCCTCCGGGCACCATTCGCGTCATCTCCTGCAAAAATTCCGGATTAGACGCCAACT
GCATGACGTATTGCATCATGAGAGGGTCGCTTAACAGTTGCTGCATTCCTGCATCAT
TGCCCCTTGGCTGGTTTCCAGC

Contig 149

GGTACAATCGAAGCGAGGGCAAATGGCACCACACTTGTTTAGACGACCATTGAGGTT
CACGACAATCTTCCCCGCACGGTGGTCATCGACAATCTCGAACTCGCCAATGTATCC
GTGTTTTTGCATGACTTGCAAGAACTTCACCACAACCTTCGAGGAGGGTCGGATGAG
AACCTGACGCTTGCCGCGGCGCTCTGCGCCCGCAATGCAGCGGAGCGCATTCGCCAG
AACGCTCATCATTGTCATCCTTGCGCTTCCTTGGCTCCAATACCT

Contig 150

GGTACAAGCGAAATAAGAACAAGGGGAGGCAAGTTCACAAGAGTATCACAACCAAAA
AGGTGCGTAGAACAGCACCAAACGCTCCAACAACAAAAGCAACAACAATATGTTAAC
CGGTTCAAGTCCCAACGAGAGCAACACCACCAAGGAGGGTCATTTAGCGCTACGATT

TCCACATAGAAGAGACCCAACCTGTCAAATTTTCCTTCCAAGACGGGCCACCAGCAC
CGCTAAATTTGTAGCGTTCACGCAATTTGTTTGCATCCTCACGACTCCCCGACTTTA
ACCACGCGATTGGTTGAAATCCAAACCGCACCT

Contig 151

GGTACCGAAACGTCTGTCACACCCTCGATTGGTTGCTCACCGACCCTCTCCACAGTG
ACGTTCAGCACGGCAATGTCGCTCCCAAGAATGCAGCCACGGAGGGACTTGCGGCGA
CGCTCACCCTGATAACCACGGAAGGTGTTGAAGCCAACAGCGCCACGCTTCACGAGA
AGTGACACACGACTTGGGGCCATTACGCCTTGCACCATTGGAAATCCCTCTTTATCG
GAACCGCCACGAAGTTTGAAGGTATAACCCTTGAAGGGTTCGCCAAATATGGCACCG
TCCACCTCGTTGCCAAGGCGGTAGTCGCCCAAGTTGACACGGCGGAGGACCTCATCG
GTTACTTCAACCTGCTTGACGGTGCCGTTCCGCGGATAGGCGACATTCAACTTCATC
GGAGCCACTCTTTCAATATAGTACCT

Contig 152

GGTACCTGCCGGCGTGTCCCCGCTGCGCGGTTGGGGAAGAGACAAGTTTAGGGGGCT
GTTGTGACACAGTTAGATTTTCCACGTCACCTCCAGCGGTGCGGATGCCTCCACTTC
AGGTGTGGTTCCCCATTCCCCCGTATATCTTTCTTTCTTTCTTTTCGTCTGCTCATT
CCACTCTGGGTGCAATATCCGGGAGTGTGGAATAAGAGTTGAGATTGGGGTTAAGAT
GAGGGTAAAGGGGAGTGG

Contig 153

GGTACCTATCCATTTAGATGAGGAAACCTGTGTGTGTTTGTTAAGCAATCACGAAAA
AAAAAGAAGGCCACACTGTTGGAGCTACACGCCCACAGGAACATGTCAAAGGAGCGC
CACGCTAATTCTTCACATAGAGAAAATGTCATCTTCGATGTACCT

Contig 154

GGTACGAGCATGTCGGGTGCTGGTGCATTCGGTAACATGTGCCGGGGTGGTCGTATG
TTCGCACCCACAAAGATCTTCCGCCGCTGGCACCGTAAGATTAATCTGCACCAGAAA
CGTTTCGCTGTGGTGTCCGCCCTCGCCGCCTCGTCACTGCCAGCACTTGTAATGTCA
CGAGGACACAAGATCGAGAATGTCGCTGAGGTGCCGCTCGTGGTGGAAGATGGTGTA
CCT

Contig 155

GTACACATCAAGGTCGGGGTTGTAGCCGTAAACGTCAATTCCTACCACCTTGGGTCC
ACGGATGCCGTTCATTATGTGAAGCAAATCGCGTGTCGAGAGCCCTCCAGCAACGGG
GCTGTCAACGGCGGGGGCAAACGCGGGGTCTAACACGCTGGCATCAATGCTTATGAA
AACGGGGTAGTCATTCCTGATATCACGTATGCAAAAAAGTCCCTTGGAGTATATGGC
GTGCATATCCATGTAAAACATCTTATGAAGCTTCCGAACCTTCCTGTCCTCCGATGA

CACTTGACGGTTACCCACGCTAACAACGCCCTTTAGCAACCCCTTATCCAAAAGCAC
CCTCAGCGGCGACCGTGGGTCGCTCACAGATGGACGTGCCGAAAAGTGAACAATAAC
AATGTCGTCAGATAGGAACAGTCGCTTGTACCT

Contig 156

AGGTACCATTCACTACCTTCCCCTTAGGTCGGTGCACCCCAACAGTGACACGCCACG
CATACGCCAATAAGCGTCGTCCTCAGCCACGGGGAGGTTCACGTCAAATATCCAGTG
TAAATCAGACCGTGTGGGATACTTATAAACGGGGCAACGGTAATATTCCTGCTTCTT
CTGCCTCTCCTTCAACTGTCCAGGTTTCATCGGTGCATTCATGTGCGGCGGCGGTGC
ATCTTTATTGATAGCGGAAATGAGCAACATGGGCAATTCACGGAACAGATCACCTGG
CGGTGGATCCTTCAGCCGCTTATTCGCTTTA
TCCCAC

Contig 157

GGTACTTTCCGTTGATGGCAGCTTTGGGCGTGGCGAAACCGAGACCGATCTTCTTGG
AGTAACGAATGTGCCCACTGCGGTTCACGTGCTTCTTCGAGGCCGTGCGAATGTTTT
CATTGACGGCCGTCTGGCGCTGATAGGCCTTTTCGTGCTGCACCGTCAGATCGACGG
TGGGCGTGCGCGCATAGTTCTCCTGCTTGCTCGTCATTTGCTCTTATCCCAATATAG
TACCT

Contig 158

GGTACATCTCCACAGGATTGGGAACACCGCGTAACGTTACAGGGCCCAGTGCAGTTA
CATCAAACTGCTCACGCTCTACCACACTCAGCGACATGTAGGTCGCACGTGTCATCA
GCACTTGCCCGCCGTTCGCAACACTCTCTGTCCTTGCTGCCATATTCGCTGTCCGTC
CATAGTAGTCATATCCCTTCGTCACCTCATCGTGTCGGATGTCGCACAACCCGGTGT
GGATTCCAACACGTACCT

Contig 159

GTACACGATTTCCTGGATCGTTTCCTCTGCGCATATGCGCGGTCCTTCGCCAACATT
TGCCGCGTTTCCCACGAAATATTTAAGCCAGGGATGGATGCTGTTGTGGGACCGTCG
CATGCGCATCCGGTGCAAACTCCGAGATACATCAGATGATCCGCCACAGATGGATAA
TCCAAGCTGTCACTGCAACAATAGTCTTCCCCTACACTTCCCGTGGAAAAGTTAATA
TCGTCCATATCGCTGCAGTTCAGGCACTCCAACGGATCATCATACTTAAACCACAGT
TCGTGTGGCGCATGCACATAACCCATGAACAATGGTGGAAGGTGGGGAACAATGTCC
TTTCGGTGCACGGCGCGATAGGACTCATGGCTCCCGCTACGGAACAATGTTTTCACC
CACTGTACCT

Contig 160

CAGGTACATCGCACACAAAAGCTACGCCTCGTCTTTTGAGGAAACTGAGCACTGCAT
TATCACACTTGCTAGTTTCTTCACCGGTGGGACAAAAGGAGAAGGTGTGTAAAACGC
AGCAAGTACCT

Contig 161

AGGTACCGCGGTAGTTGGGGGGATGGGCTCTCACCTCACTTTGGAGTGTGTTTTATT
GCTCTATCCTTTACTGATGCAGTGACTTTTTAAAATCCTTCCCCTATCGTTTTTCTT
TTCGTTGCAGTTAGAGGTGAAGTCGCAGTCCTTTCGTTTGTTTTTGCCTTCACGTGG
GGTAAACATGAGAGAGCTGCGCACGGAGAGCTTGGTTGATGTTTGCCACGTGTTGTT
TTCCCATCTGCGTGGGAAGCTAATACTTGGTGAAAGTGGGGTACCT

Contig 162

GGTACTCGACGAACTTGTGCGTTCATATACATGGAAAAGTCTCATGCACGCGTTGGA
CTACCCTTCATGTGTTTTTTCTCTTTTTTAATTATTTAATTTTTTACTGTCAGAAGA
TCGATTGTATGACCGTTGCCTCCATCATTAATTAATCTGACAAAAAGTAAATATATA
TATATGTAAAAAAAAAAACCT

Contig 163

GCCGACGTACACYGTGATCTCCKTGKYAA

Contig 164

AGAGCAAGTTGAAGTCCAGCTGTTGAAGTCCGGCRTCGTCTGATGTCWGTGCTTCAS
TGKTA

Contig 165

AGGTACTTTGTTAGCGGTGGATTTGCGCATATAAACAATGAGGGATCGTGTGGTGTC
AACACTGTGGAATGCACTCTGCTGGATGATTTGGACCTTGCCATCGCTGAAAAGGAG
CTTGCGGCACAACAGGCAGCGTTAGGGTCGGCGAAAGATGACAAGGCGAAGTCTGTT
GTTGAAATTCGCATATCTGTCATTGAGGCTGTTATTGCTGCCCTAAAACACCACTAG
TAAAAGGCTGCAAACAGGAGGAGTGGGACAAGAAAAAGAAATGAGGGACTGACCGTA
GTTTGCGAGCTTTAGTTTATTGCATGTCTCTTGTTGTA

Contig 166

GGTACCTCCTGCACGACGGCAATATGAGACCTTCCAGTTTGGAAACATTCCAACATA
GAGCTAAGCTTTGTTTCGGAATGAACGATATGGCAGCTTCTGTTGTAAAACTTTACT
AACACAAGCACAGGTGTCTCTTCGTCGGGGTTGACCATCAAGAGGTCTCTTGTGAAT
AGCGCGCCAACGATGTTATTCCTATTCCCCTGGTATACAGGAATGCGGCTGTGGCCC
CTTTCACAGATAAGCTGAATGGTTTCTTCATTCAGAGGTTGACTGGCCTCCAGCATG
AACGTCTCCCATATGGGTGTAAGGACATCCATCACAGTTTTCTCATGTAGTTCCATA
GCGCCCACCATGAGATCCGCCTCCCGCTCGCCAAGGCCGGATTCGGCACCACGCGCA
GCGTGCATAAACATCAGTTTCTTTAACTCGTTTCGCTCATATATCTGCCCAGGGTCC
ATACCAATAAAACGGTCCAGAACCATACTGAGGGGCTTACAGACTGGGTAGAACAGT
AACACGAAGAATTTTACAAGGTACCT

Contig 167

GGTACACAGCGGCATAGATTGAAATACTACATAAAGGTTATACATTGATTATAATGA
AAGACATTTAAGCGCAGCGCTACTGAGTGATCTGCGGCATCGT

Contig 168

AAACCATCCAAATCCTTTACCGGTTGGCCTTGGCGACGCGCTCCACCTCATCCTTCT
TTTTGATGGCATAGGAATTGGAGCTACCCTTCGACGCATTCACGATCTCGTCCGCAA
GGCACTCAGGAAGTGTCTTGAGGTTGCGAAACGCCGCCTCACGCGCCCCCTTGCACA
TGAGGTAGATTGCCTCATTAACACGGCGCATCGGTGACACATCAACAGCCTGCCGCC
GCACAACACCACCAGCGACCTGCC

Contig 169

AGGTACTCACTATTGCCCCGCTAATCGCTACTGTCGTGTTCAAACATACCCAAGATG
CCAGTTTTGTTCCTTTGTTCGAACAACCCAAGCTTAGGGAGGGAAAGGATCCGTCG
ACTCTCCATAAAGTGTTAACTTTTTTTTCCCTTTTCCACTGTGTCGATGTTGATAGG

TATAAGGGAGGTGGATGCTCAGTGGCGAGTGTTTCGTGGGAATTTATGTGTAGGCCG
CTTTTTTTGACCACGGTACCT

Contig 170

AGGTACGCACTCGCCGCGAGTGACCATATGGTTGCGGGCGACGTGCATGCAAAATGT
GTTCGGGGAAACGAGATGCTTTTTGTCGCTGACTGGAATGCTGAGGAGCTTGCCGCG
GAGTTGAGAGAACTCGAGGAAATAGTGCAAGGTGGGGGTGGCACCCGCAGTGACGGA
TCGTCCAGTAGCAAAAAGCAGACAGGCGCCCTAAACAGCGCAACCGGCGTTTCCACG
TCTTTTTTCTCTGCGAGAAACGCAGGAAAGACATCGGCAGCGGTAGCAGCCTTACTT
AGCGCCCCGAAATTCTCCAGAAATATCGGAGAGGGGGGCACCGGCAATGCCGACACT
GTCACCACAAGCAGCGGCATCCCAAATATAAATGGCACCTCAAACAACACCGATAGA
GATGTCTTACAGCAATCGATTGACGTCGACGTGCTGAGTGCCGCAAACTTCACCTTT
CATACTGTTTCGACATCAGCACTAAAGGGTGAATCGATGGTTCGTCGGGTTGGTAAC
ATCACTGGGGACACCTTCGAGCGTGGTTCGTACCTA

Contig 171

TTTTGTCGATAAATACAACCGAGCCACACAACAATTATTGAGAAAATGGTGAATGTG
ATTTGTGGACAAAAGTGCCAACAACACATACAAACAAGAAAAAAAAAGAAAAAAAAA
GGGAGTGAAGTAAACAGTTTGCGGGCTCACTCGCAACAATCCCGTAAGGAATGCAGT
GAGGGATCCGCATAGGCATCTAACTATTTCCAGGTCGCGGGATATTCACCAGACCGA
CTGTGAGTTACTCTATCTAGACGCTTGGCG

Contig 172

GGTACGGCACAAGGTTTGTCTGGAACTCTGTCAGATCCACGTTCAATGCACCGTCGA
AGCGGAGGGACGCTGTCAGCGAGGAAACCACCTGACCGATGAGGCGGTTCAGGTTGG
TGTACCT

Contig 173

AGGTACATAAACAGCACCAAACAATAATATTGAGTGCCCACAAAGACACAAACCATT
TCCACCACGTGATGCCGACAGAAACCAAACACGAATAAGAATTTAGACCTACCCGCG
GACACAAAAAACTCTGTGGCACATCACAAGAACAAAGAAATGAGGTATTTGTCAGCT
CTATAACCCCTCCTCGAATACGCCGCGCTCTTGCTGGATCCACCGAGAAAGAAACAC
ACCATCACACTGCTTTCTGGATGCTTCACCCCGCTGGGGACGAGAAGCGCTTCCCTC
CACAGACAAATTCGTGCCGTCACTGTTGTTATCATCAGCATCACCGCCCTCCCCACC
TGAAGCCGCAGACCT

Contig 174

GGTACCAATAACAATATGTCTCGGTGTTATCTCAGCTTTTTCCACCCACCTTCCTTG
CAGCCCTCGTTGGATTATATACCACACGCGCATAATTTCACGACGGATGTATACATG

TTTAAAATTTTCCCCGCACATCTCAGTGTGGCCCAATGCAAACCTGTCGTATGGCCA
ACCTTCCTCCGAAGTGAGAACAAAAATCTCCAAACCATAAGGTCTCTCCGGCACTCG
TGCATCAAGATTCGCAGGATCAGGATTTATATCCACCTCCGCTTCGGTCCATATGCG
CTGTGGTTTTCCATCGAATACTCTCATTCCAAGTGGCTCTCTACGATAACCCGACAT
TACATAACTCCATTTTGCATTATAAATGGAATCATACACTCCTTTAATTTCGACAGC
ACTAGCAGCTGTTTGGGCAGCTCGTTCCAATGCTACATTGAGTGCATTATTCAACAT
GTTAGTGACAATAATCCAAAGCCGCTCTCCATCTCTCAGTCCACGAATATTCACCTG
TAACCCGGCCTTTACTACCTCATCTGCTTGTTGCCACTGCTGAAGATCCAATATTCC
CTCTC

Contig 175

TAGCGTGGTCGCGGCGGAGGTACACTGCACTTGTGATTGAACTCACCCGTCGATACG
GCTACGGTGGAAGTCGTTTCGGCGGGGCCACGTCCGTTGCTCCCACTTGATGGAGGC
CGAACATCGAAGAAGACAAAATCTCCTGCAAGGTTCCCCGTTCCACCGAGAGTGCAG
AAACGGCCATTCGGTGACCACAAAATAACATTCCTTGGTGCTTCACCAAATGTCATC
AGCGCAACACCCGCCTTGTCGTATAGTGTTGTCTTGTTGACGGGCATCTTGCCGTGC
ACGACCATAAGCTCGTCAACCGTCGGGCTCCATTGACAATCATGAACTGATTCCCCG
CTAGGGAACTTTATCTGTTTTATCTTCCGCCCGCCTACGTCAACGAGATAAAGTAAC
ACCGTACCTGCCCGGG

Contig 176

GGTGCGGCCTAAAAACTGAAACTTTCGATGCCCACCTCCAGACAGCCCCTCTCACTC
CTTTTTCTTCGGCTTGATTGCACCGGTCCCAACGGCAAGATGCTCCGCAGGAACTC
CAAGCTTTTCCGCCCGACAGAAGAACAATGGTAAGTGCTAATTAGCTGTTAC

Contig 177

CAGTGCCGACTGTCCAAAACASAAACGCTAACAATAYTATTAGCGTYGAGGTCAAAG
GTCATTTCTATCTGTGGCACGCAGCGAAGAGCC

Contig 178

GGTACCACAAATTGAGGTGACATTTGACATTGATGCTAATGGTATTTGCCACGTAAC
TGCGAAGGACAAGGCAACTGGGAAAACACAGAACATCACCATTACTGCTCATGGTGG
GTTGACGAAGGAGCAGATCGAGAACATGATCCGCGACTCTGAGATGCACGCTGAAGC
TGACCGTGTGAAGCGGGAGCTTGTGGAAGTGCGTAACAATGCTGAGACTCAGGCTAA
CACTGCTGAGCGACAACTGACGGAATGGAAGTACCT

Contig 179

GTTACKTTCCATACAAGTGCCGGKTCGGCCGKGGWAATAC

Contig 180

CCATTGGATTATCGTATTTCTTGTGTAAGGGTCTTGCCAACGGACTTGTTGGTGGAT
CCATTATTGCGATGTTTTTGAACCGTTTCACTGTCGAGGGTCTTGTCTACCAGCGGC
TAACTAACATCGCCGGCATGGGGTGGTCCGTGAAGCCTCTTTTGGCTGCTATCAGCG
ACATATTCCCGTTCTTCGGTTATACCAAACGGTGGTACCT

Contig 181

GGTACTGAACAGTTTTCTCACGAAGGACGCGCTGTGCCTCGGCGAATTCCTCGTCCT
TGGTCCTCTGTTCTTCCAGCCGCTTTGTGTAAAACTCAGCCTGCTTGTCCAATAGCA
CTGTCAATGCCTTAATGGCGTTCTGTTCATCCTGAAGCTCCGTCAAAACCTTTTTC
TCATTATCCACTCGCTCGACACTCCAACTACACTGCTTGATTGACTGCCGGAACTTA
TACACCCGGTTACCTGCTTCACCAGGGCTGTACCT

Contig 182

GGTACTTGCGGTTTTGCATTAGGAGCATATAAAGATCTTGCACGTTGTTGACGGGAA
AGTTTTTGAAACCTGACGGGGTGATATACTTTGTTCTGCGGTCGCTTCCATAACCCT
TGTTTGGCATTGCCTTCTGTCCCTTGTAGCGGCGACGAACCGGGGAGTCCTCACCAC
GGGGTTTACGCCAACTGCTGCTCAGTTGAGGGAAGAGTTCACAACGATGACGTGTGA
ACTTCTTCGTCCGCTTCTTCACGATGTTCCTCGGCACAAACGGCTTGACCATCTTGA
ACTGGCACCT

Contig 183

GGTACCCACCACCAAGGGAGTGGCAGATCTGGAAGCCCTGGAGGCAGTCACAGCTCT
CCGCCTCCTTGCAGCACACATCGAGCACAGAGTCGATCAGTTCCGCACCCTCCGTGT
AGTGGCCCTTTGCCCAGTTGTTGCCGGCGCCAGACTGTCCAAAGATGAAGTTGTCGG
GGCGGAAGATCTGACCATAGGGGCCAGCACGTACCT

Contig 184

TCTCATTGGCCCCCATGACGCTGGATATGGTGTGGACGGTAAAGGCAAAATTTTACC
TGTTGTGATTCCCCATGATGACAAAAGTGGGCACAACCCTGGTCCCTTACCTACACT
TCGTTAAGTGACTCAAATGAGGTGTCCTCATCGCTTACTGTAGTAACCTTCGCCATC
GCTCATTACTATTATCACTGTTGCTGCCGTCACGGTATCTCTTCTTCATTCACTAAC
ATGCCTCTCAATTACGCGCGGAAGCTGCTGCTTACGTTATTCGCTTTCCGTGGGAAG
GGGGAGGGGTATCTGCCCGACAGGGTTGTTAATGGCGGGGGGAGTGGAGGAACGAGT
GCCCGCTTTTTCTGGTAATGCAAAAAAAATAACCAAATAAAATAAATGGATGAATGT
GATGCAGTGATGCAGATATGGGAAGAATGGGTGTGATGAATGATGTACCT

Contig 185

GGCGCGCGCAATTCACTAKATGATATATTCG

Contig 186

NGCGCGCGCAATTCACTAATGATTTTTAGCGGCC

Contig 187

CTGAACCACGTTGGTGTCTGCTGTTGKGTCCGGCGT

Contig 188

AGGTACAAGGCACCGATCTGCAAGACCTGCGGCCACCAGAGACTTCGCAATCCAGCG
AGCAGCATACGCTGCCGACCGGTCCACCTTCGAGGGATCCTTGCCAGAAAACGCGCC
ACCACCATGGGCACCCCAACCGCCATACGTGTCAACAATTATCTTGCGACCAGTTAA
CCCAGCGTCACCATGTGGGCCACCGATGACGAAACGCCCAGATGGGTTCAAGTGATA
TTTGGTGTTCTCATCGAGGAACTTCGCGGGAATGGTGGTGTTGATTACCTTCTTCAC
CAGGTCCTCGCGGATGGTTTCATTTGACACACCGTCATCGTGCTGGGCAGAAACAAG
AATCACAGCCACTCGCTTAGGTGTCAGCAACTGCTTGCCACCGCGGGTGTCATATTC
GTACCT

Contig 189

NGNCGNGGGGGCTCCCAAAAGGGTTTACCGGAAGGCGCCCGGAAATYCCAAGGRAT
TAGAGTTGCGCGC

Contig 190

GGGTGCGGATACAGCATGTGGAAACATTCCTTGTT

Contig 191

TCTTGSCAGCCTTTCCTAAGGKGGAGTTCCTACGCCTACTTGGTGCGGGAC

Contig 192

GCGGCCGTTTGTACCTCAGCACGCGAACACGCTAATCGAAWTMT

Contig 193

GTTACGGGCTGATGTAACCGCGGTCCACACTCATGCCTTCCACAACTTCCAGCTCAG
TCGTCAATGTCTTCCCATCTTGTGTTGTGATCACGCCATCCTTCCCCACCTTCTCCA
TCGCCTCCCCAATCAGCTTGCCGAGTTCAACATCACCGTTCGCGGAAATCGTCGCAA
CCTGCACAACGTTCTCCGTATTTGTTACCGTTCGGCTCTGAGATTCGATGTTCTTCA
GGATCACCTCCACGGCGCGGTCCATACCACGCTTCATGTCAATGGGATTCGTCCCGG
TTGCAATCGATTTGATACCCTCGCTAAAGATGCTTGCAACGAGGACAGCCGATGTCG
TCGTTCCATCACCCGCGAGGTCATTTGTCTTATTGCACACCTGCCGCACGAACACCT

Contig 194

AGGTACGATGCTGGGTCCCGACTCGTCGTATTCACTCTTCGTTATCCACATCGACTG
GAAGGTTGTTAGTGATGAGAGGATGGAACCGCCAATCCACACGCTATACTTGCGCTC
CGGTGGTGCCACAACCTTAGGCTTGATCGATGACGGTGCAAGATTGCTGATCTCCTT
TCCAAGTCGCTCAGGTAGGTTCTTGAACATCGTGGTACCT

Contig 195

TATTCGTGGTCCCGGCCGAGGTACC

Contig 196

GGTACTTCCTGAAGTTGATGCGGTGATGGTGCTCGCCACCGGCGTTACCACGACCGG
AAGGGTGCTTGCGGTGCTTGCCGACGCGACCATAGCCGCAGAATGTCGAGCCACGCT
GGTGGCGTGTCTTCTTAAAGCGGGTCGGCATGTTGAAGCAATACCT

Contig 197

GGTACAGGTCTCGCTGCCATTGCTCTTGCAGGTGTGGGTCTCGGCATCGGAACCATT
TTTGGTAACCTGCTGGTTGCTTGTGCTCGGCAACCCAATTTGACAAAGATGCTCTTT
AATTATGCAATTCTTGGTTTCGCGCTCACGGAAGCTATCGGCCTATTCGCACTCATG
CTTGCGTTTCTGATGCTCTTCTCGTAGAGTAGTTAGTTGAAACAACGAGTTGCATGG
GAAATGTGCAAGAAAAAAAGTGGCGGAGGGGAGAGGGAGAGGGAAGTGGAGGTGCT
GTTGGGTCTCGTTGGTGCATTTATGGTATTGTAACCAACATCCCCCACCGAGGCTGC
CTTTTCTAGTTGTACCT

Contig 198

GGTACTACACTTGTGGTCCACTTGTTTAAACCCTTTCGTGTTTTGTTTTGTTGGATC
GTTGCCGTGACTGCCCCGCTTCTGGAAGGGAGAGGGACACTGTAACCATCCTATTCC
TCAATCTTTCATTGTTATCCGGCTGACATAAATGGATTCTTCAGTGCAAGGGCTCTC
TGATGAAAGATCCTCCTGATTCCCGTTCCACCTTTCTTTGATACTTTTCCTTGTTTC
TCTTACTTTTTTCCCTTTCTTTCGGACTACGCAACTCTTCTATTATCTTTTACTTTG
TCCCGTCGGTTTGTCTCCCCTTCATTCCTTCCTACGCAACTCACGATGGGACAGGTA
GACGGGATGGTTGCCCTTGATGCTATCAAGATCATGGGACCCCACGCAGGTTGTGGA
AACGGGAGAGGTGCTACGCGTAACGGGAAAATGGGATGGGATTCGTACCT

Contig 199

CATGGTCAAATTCTGGAGCAYTACCGATGTTRCTGCSATTCTTGAMAATGAANCTTT
TTAT

Contig 200

AACATGTGTTGGAAGGWTYCAGCAGTCTCTCCGG

Contig 201

TTCYTGCCGGCAAGYTKGCTAACCMAACGGGCCTTCACAA

Contig 202

GCTCACCAKGGGTSTAAAGATAGTACTAGCSGATGACGCTACTCTT

Contig 203

GGTACCCCTATCAGTGAAAGAAAAGACAAAAAGAGAAGGGCCTTCCATACAGCCAAT
CACGGGAGTCGTCAGTCCCTCCATTTCGCTCATCATTCGATGACGTGTCAGCGATCT
TAACGCAGCTTGACGAAACCAATGTTCTCGGCATTCTCACGGAAGCACTGACGACAG
ACGTTCAGTTCATACTTGCGGATGAGAGCCTTTTGGTTGGAGCAAATGACGCAGTGG
CGGGATCCTTTGCCCATGCCAATCTTTTGGCGGGAGCGCCATTGGTGGAGGTGACCC
ATTTGTCTACTTCCTTTTGCTTTGTGCTGCTACTAAGCAGGAGGTCTTTCCTTCTCA
ATCCTTCTTGGGGAAACAATATAGTACCT

Contig 204

GACGGTTCATACATAGTCACGTAAGGACAGACAGGGACCCCAATACATCCCACATCA
TGGTACCT

Contig 205

WKACGCGYGGTCGCGRCCGACGGACSTCCTTTCTCTACACCGACAATCTGG

Contig 206

ACAGGCACCAACCCATACCCGCAGACTTGCCGGGAGCGTGGCCCGTAATTGTGTTGA
CATCTCCAGATCCAAATAACTACAAAGACTGGAAGGAACCACGGGGTTGCGAGTTCC
TTTGCATCAATTGCTATGAAGAGGTGGAACTGAAGGCTGCGGTTGCGTGGAGGCGGC
TGTCGAAGTTGGACGAAAGTAAAATCACCGAAGCTCACATCATAAACCTTGAGAATG
AATGGCAGAAAATAGAAGGATGGATTAAGAAGGTGGGCCCACTGCCCCGCCAAGTAC
TT

Contig 207

GGTACGCAAGTTTCTTCAGGGCAGCCGCCGCCACTTTGTTCGCAGCGATCGCCTTCT
TGAAGTTCTCCGTCTGCTGCTGCTTGAAGGCAGCGCGCTTGATCGCAGACTCCGGGG
CGGGAACAGCTTTAGCGGGCATAACTTCTCTACCTTTCCCAATA
TAGTACCT

Contig 208

CATTTTTGATAATGATGAGGAGCACACACCCCTTSCAGCNTCCTTCTGGCTGGAGAG
TTGTACATCACTCCAATRCTGCTCTCTGGHGGTCGTCC

Appendix III

Blast search result

Name	Blast Result	Species	E-value	Size
Contig 68	succinyl-CoA synthetase alpha subunit	*T.brucei*	$9.2e^{-40}$	338
Contig 44	cytochrome c oxidase subunit 10	*T.brucei*	$8.2\ e^{-16}$	188
Contig 99	aldehyde dehydrogenase	*T.brucei*	$9.4e^{-47}$	308
Contig 95	D-alanyl-glycyl endopeptidase-like protein	*T.brucei*	$2.8e^{-63}$	383
Contig 48	stress-inducible protein STI1-like	*T.brucei*	$1.2e^{-32}$	242
Contig 150	GTPase activator-like protein	*T.brucei*	$3.9e^{-24}$	318
Contig 133	retrotransposon hot spot (RHS) protein	*T.brucei*	$1.1e^{-27}$	356
Contig 132	unnamed protein product DC24 Putative protein STRF6	*Homo sapiens* *Homo sapiens* *Homo sapiens*	$4.e^{-59}$ $6.e^{-44}$ $2.e^{-42}$	363
Contig 174	retrotransposon hot spot protein, RHS4	*T.brucei*	$2.3e^{-100}$	575
Contig 197	ATPase subunit 9	*T.brucei*	$2.2e^{-29}$	359
Contig 37	ADP, ATP carrier protein 1, mitochondrial precursor	*T.brucei*	$9.2e^{-56}$	352
Contig 165	ATP synthase, epsilon chain	*T.brucei*	$4.9e^{-34}$	323
Contig 114	aminopeptidase	*T.brucei*	$2.8e^{-19}$	288
Contig 47	eukaryotic peptide chain release factor subunit 1	*T.brucei*	$5.6e^{-72}$	459
Contig 95	D-alanyl-glycyl endopeptidase-like protein	*T.brucei*	$2.8e^{-63}$	383
Contig 148	ubiquitin-like protein	*T.brucei*	$3.0e^{-41}$	250
Contig 112	proteasome regulatory non-ATP-ase subunit 3	*T.brucei*	$1.2e^{-10}$	100
Contig 55	RNA-binding protein	*T.brucei*	$2.5e^{-39}$	320
Contig 45	elongation factor 1 gamma	*T.brucei*	$2.4e^{-32}$	241
Contig 66	DNA ligase I	*T.brucei*	$7.0e^{-106}$	634
Contig 59	spliced leader RNA PSE-promoter Transcription factor	*T.brucei* s	$1.0e^{-95}$	612
Contig 79	ATP-dependent RNA helicase	*T.brucei*	$1.7e^{-61}$	399

Contig 105	eukaryotic initiation factor 4a	*T.brucei*	$1.4e^{-36}$	483
Contig 97	eukaryotic initiation factor 5a	*T.brucei*	$2.9e^{-43}$	470
Contig 170	5'-3' exonuclease XRNC	*T.brucei*	$1.0e^{-88}$	549
Contig 3	heat shock protein 70	*T.brucei*	$1.5e^{-85}$	550
Contig 36	heat shock 70 kDa protein, mitochondrial precursor	*T.brucei*	$2.7e^{-79}$	512
Contig 54	heat shock protein 83	*T.brucei*	$3.4e^{-26}$	209
Contig 23	heat shock protein 70	*T.brucei*	$3.4e^{-72}$	475
Contig 103	heat shock 70 kDa protein, mitochondrial precursor	*T.brucei*	$1.8e^{-52}$	360
Contig 52	10 kDa heat shock protein	*T.brucei*	$1.6e^{-19}$	269
Contig 58	heat shock 70 kDa Protein (C-terminal fragment)	*T.brucei*	$4.2e^{-35}$	261
Contig 93	heat shock protein 83	*T.brucei*	$1.3e^{-44}$	332
Contig 113	heat shock protein 83	*T.brucei*	$6.7e^{-60}$	379
Contig 7	heat shock protein 83	*T.brucei*	$7.6e^{-45}$	400
Contig 193	chaperonin Hsp60, mitochondrial precursor	*T.brucei*	$7.3e^{-63}$	399
Contig 138	heat shock 70 kDa protein Mitochondrial precursor	*T.brucei*	$2.6e^{-37}$	235

Name	Blast Result	Species	E-value	Size
Contig 178	heat shock 70 kDa protein Mitochondrial precursor	*T.brucei*	$5.4e^{-43}$	264
Contig 15	ATP synthase alpha chain Mitochondrial precursor	*T.brucei*	$2.4e^{-71}$	469
Contig 92	guanine nucleotide-binding protein beta subunit-like protein	*T.brucei*	$2.0e^{-60}$	427
Contig 87	casein kinase	*T.brucei*	$2.6e^{-26}$	253
Contig 122	activated protein kinase C receptor homolog [ξ]	*T.brucei*	$1.4e^{-15}$	106
Contig 158	receptor-type adenylate cyclase GRESAG 4	*T.brucei*	$2e^{-38}$	246
Contig 127	nucleolar RNA helicase II	*T.brucei*	$3.3e^{-44}$	503
Contig 32	beta tubulin	*T.brucei*	$2.1e^{-15}$	215

Name	Blast Result	Species	E-value	Size
Contig 46	actin	*T.brucei*	$6.2e^{-82}$	509
Contig 172	alpha tubulin	*T.brucei*	$1.7e^{-15}$	121
Contig 183	beta tubulin	*T.brucei*	$1.2e^{-35}$	207
Contig 156	dynein heavy chain	*T.brucei*	$8.7e^{-52}$	322
Contig 194	actin	*T.brucei*	$2.7e^{-33}$	211
Contig 172	alpha tubulin	*T.brucei*	$1.7e^{-15}$	121
Contig 41	amino acid transporter	*T.brucei*	$1.2e^{-32}$	239
Contig 100	amino acid transporter	*T.brucei*	$2.8e^{-47}$	328
Contig 180	pteridine transporter	*T.brucei*	$5.0e^{-33}$	211
Contig 29	69 kDa paraflagellar rod protein	*T.brucei*	$1.8e^{-57}$	451
Contig 70	69 kDa paraflagellar rod protein	*T.brucei*	$2.3e^{-50}$	346
Contig 158	receptor-type adenylate cyclase GRESAG 4	*T.brucei*	$3.2e^{-38}$	246
Contig 10	69 kDa paraflagellar rod protein	*T.brucei*	$3.5e^{-79}$	530
Contig 139	73 kDa paraflagellar rod protein	*T.brucei*	$1.1e^{-19}$	236
Contig 49	hypothetical protein, unlikely	*T.brucei*	$1.1e^{-18}$	261
Contig 73	hypothetical protein, conserved	*T.brucei*	$5.8e^{-12}$	282
Contig 82	hypothetical protein, conserved	*T.brucei*	$4.8e^{-43}$	304
Contig 28	hypothetical protein, conserved	*T.brucei*	$4.6e^{-38}$	363
Contig 51	hypothetical protein, conserved	*T.brucei*	$1.5e^{-48}$	323
Contig 74	hypothetical protein	*T.brucei*	$8.9e^{-74}$	438
Contig 31	hypothetical protein	*T.brucei*	$3.8e^{-18}$	202
Contig 53	hypothetical protein, conserved	*T.brucei*	$5.1e^{-39}$	308
Contig 76	hypothetical protein, unlikely	*T.brucei*	$1.8e^{-18}$	245
Contig 88	hypothetical protein	*T.brucei*	$1.4e^{-20}$	603
Contig 34	hypothetical protein, conserved	*T.brucei*	$5.8e^{-86}$	543

Name	Blast Result	Species	E-value	Size
Contig 60	hypothetical protein, conserved	*T.brucei*	$3.8e^{-34}$	343
Contig 89	hypothetical protein, conserved	*T.brucei*	$1.8e^{-36}$	265
Contig 38	hypothetical protein, conserved	*T.brucei*	$2.0e^{-23}$	252
Contig 65	hypothetical protein, conserved	*T.brucei*	$1.8e^{-44}$	359
Contig 80	hypothetical protein, conserved	*T.brucei*	$6.0e^{-18}$	475
Contig 96	hypothetical protein, conserved	*T.brucei*	$8.2e^{-32}$	233
Contig 104	hypothetical protein, conserved	*T.brucei*	$2.1e^{-07}$	401
Contig 134	hypothetical protein, conserved	*T.brucei*	$7.9e^{-11}$	260
Contig 160	hypothetical protein, putative	*T.brucei*	$5.4e^{-12}$	125
Contig 175	hypothetical protein, conserved	*T.brucei*	$8.5e^{-69}$	415
Contig 126	hypothetical protein	*T.brucei*	$2.5e^{-39}$	278
Contig 166	hypothetical protein, conserved	*T.brucei*	$3.6e^{-93}$	539
Contig181	hypothetical protein, conserved	*T.brucei*	$1.7e^{-38}$	263
Contig128	hypothetical protein, conserved	*T.brucei*	$2.4e^{-57}$	336
Contig 152	hypothetical protein, unlikely	*T.brucei*	$1.8e^{-34}$	246
Contig 173	hypothetical protein, conserved	*T.brucei*	$9.4e^{-26}$	357
Contig 184	hypothetical protein	*T.brucei*	$8.9e^{-19}$	449
Contig 17	hypothetical protein, conserved	*T.brucei*	$4.0e^{-46}$	396
Contig 69	hypothetical protein, conserved	*T.brucei*	$1.6e^{-74}$	526
Contig 207	60S ribosomal protein L7, putative	*T.brucei*	$4.5e^{-22}$	166
Contig196	60S ribosomal protein L27A/L29, putative	*T.brucei*	$3.8e^{-25}$	160
Contig 182	ribosomal proteins L37,	*T.brucei*	$5.6e^{-49}$	295

Name	Blast Result	Specie	E-value	Size
	putative			
Contig 168	40S ribosomal protein S5, putative	*T.brucei*	$3.3e^{-35}$	252
Contig 157	ribosomal proteins S17, putative	*T.brucei*	$2.1e^{-33}$	233
Contig 154	60S ribosomal protein L4	*T.brucei*	$5.2e^{-37}$	231
Contig 151	40S ribosomal protein S6, putative	*T.brucei*	$1.3e^{-65}$	425
Contig 149	40S ribosomal protein S15a, putative	*T.brucei*	$1.0e^{-40}$	273
Contig 147	60S ribosomal protein L35, putative	*T.brucei*	$5.1e^{-62}$	447
Contig 141	ribosomal protein L27, putative	*T.brucei*	$1.5e^{-48}$	324
Contig 140	60S ribosomal protein L13, putative	*T.brucei*	$1.0e^{-72}$	446
Contig 137	ribosomal protein L37, putative	*T.brucei*	$1.0e^{-31}$	218
Contig 129	RPS12 40S ribosomal protein S12, putative	*T.brucei*	$1.6e^{-33}$	251
Contig 124	60S acidic ribosomal protein, putative	*T.brucei*	$2.6e^{-51}$	510
Contig 123	ribosomal protein S19, putative	*T.brucei*	$2.4e^{-09}$	355
Contig 119	60S ribosomal protein L21E, putative	*T.brucei*	$9.6e^{-38}$	344
Contig 115	60S ribosomal protein L4	*T.brucei*	$2.9e^{-43}$	393
Contig 109	60S ribosomal protein L7a, putative	*T.brucei*	$5.0e^{-32}$	238
Contig 108	ribosomal protein S6, putative	*T.brucei*	$2.6e^{-19}$	227
Contig 106	ribosomal protein L3, putative	*T.brucei*	$1.9e^{-94}$	555
Contig 101	40S ribosomal protein S18, putative	*T.brucei*	$1.9e^{-48}$	396
Name	Blast search	Species	E-value	Size
Contig 91	60S ribosomal subunit protein L31, putative	*T.brucei*	$5.6e^{-49}$	319
Contig 83	ribosomal protein L3, putative	*T.brucei*	$5.9e^{-19}$	171

Contig 81	60S ribosomal protein L10a	*T.brucei*	$1.1e^{-36}$	261
Contig 75	40S ribosomal protein S8, putative	*T.brucei*	$5.8e^{-38}$	374
Contig 75	40S ribosomal protein S8, putative	*T.brucei*	$5.8e^{-38}$	374
Contig 72	60S ribosomal protein L6, putative	*T.brucei*	$9.2e^{-24}$	251
Contig 67	60S ribosomal protein L7a, putative	*T.brucei*	$8.1e^{-41}$	370
Contig 64	40S ribosomal protein L14, putative	*T.brucei*	$3.8e^{-50}$	434
Contig 63	60s ribosomal protein l18, putative	*T.brucei*	$3.3e^{-67}$	528
Contig 42	60S ribosomal protein L9, putative	*T.brucei*	$1.2e^{-23}$	335
Contig 40	60S ribosomal protein L2, putative	*T.brucei*	$2.4e^{-39}$	268
Contig 35	Similar to Human 60S ribosomal protein HG12	*T.brucei*	e^{-161}	256
Contig 33	ribosomal protein L36, putative	*T.brucei*	$1.5e^{-30}$	262
Contig 30	ribosomal protein S26, putative	*T.brucei*	$5.0e^{-09}$	506
Contig 27	rRNA large subunit (28S beta)	*T.brucei*	$2.3e^{-115}$	576
Contig 26	60S ribosomal protein L17, putative	*T.brucei*	$6.6e^{-43}$	310
Contig 25	60S ribosomal protein L27A/L29	*T.brucei*	$4.2e^{-51}$	359