



This is to certify that the
dissertation entitled

CHARACTERIZATION OF FOUR MEMBERS OF THE
FERROUS ION AND α -KETOGLUTARATE DEPENDENT
HYDROXYLASE FAMILY FROM *TRYPANOSOMA BRUCEI*:
TWO THYMINE HYDROXYLASE-LIKE PROTEINS, J-
BINDING PROTEIN 1, AND AN ALKB HOMOLOG

presented by

JANA M SIMMONS

has been accepted towards fulfillment
of the requirements for the

Ph.D. degree in Biochemistry and Molecular
Biology

Major Professor's Signature

Date

CHARACTERIZATION OF FOUR MEMBERS OF THE FERROUS ION AND α -
KETOGLUTARATE DEPENDENT HYDROXYLASE FAMILY FROM
TRYPANOSOMA BRUCEI: TWO THYMINE HYDROXYLASE-LIKE PROTEINS, J-
BINDING PROTEIN 1, AND AN ALKB HOMOLOG

By

Jana M Simmons

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Biochemistry and Molecular Biology

2010

ABSTRACT

CHARACTERIZATION OF FOUR MEMBERS OF THE FERROUS ION AND α -KETOGLUTARATE DEPENDENT HYDROXYLASE FAMILY FROM *TRYPANOSOMA BRUCEI*: TWO THYMINE HYDROXYLASE-LIKE PROTEINS, J-BINDING PROTEIN 1, AND AN ALKB HOMOLOG

By

Jana M Simmons

Trypanosomes are eukaryotic parasites and the causative agents of several mammalian diseases. They are phylogenetically divergent making study of their biology pivotal to understanding their relationships to both 'lower' and 'higher' organisms.

Trypanosomes contain many open reading frames predicted to encode proteins of the Fe^{II} / α -ketoglutarate (α KG) hydroxylase family including two thymine 7-hydroxylase-like proteins (TLP5 & TLP7), J-binding protein 1 (JBP1), and an AlkB homolog (TbABH).

TLP5 & TLP7 transcript levels are 4 and 2.5 fold higher in the bloodstream form over the procyclic form parasites, indicating a role for the respective proteins in the former life stage. Protein production was verified in both life stages by western blot with a polyclonal anti-TLP5 antibody and suggested that these proteins may undergo some posttranslational modification *in vivo*. Recombinant TLP5 & TLP7 were confirmed to be members of the Fe^{II} / α KG hydroxylase family based upon formation of a characteristic metal-to-ligand charge-transfer (MLCT) chromophore when Fe^{II} and α KG were added to anaerobic samples. A homology model of TLP5 was created and a library of small molecules docked with this model, indicating a propensity of the protein to bind heterocyclic molecules resembling nucleotides. Unfortunately, no DNA binding or enzymatic activity was observed for these proteins, leaving their cellular function unknown.

JBP1 is known to be involved in the hypermodification of thymine in nuclear DNA resulting in the formation of β -D-glucosylhydroxymethyldeoxyuracil, or base J. The first step of base J synthesis involves hydroxylation of the C5 methyl group of thymidine, a reaction type known to be catalyzed by the $\text{Fe}^{\text{II}}/\alpha\text{KG}$ hydroxylases in other organisms. Recombinant JBP1 was confirmed to be a member of this enzyme family based on the characteristic MLCT chromophore formation. Unfortunately, no enzymatic turnover was observed by the *in vitro* methods employed leaving the direct detection of activity elusive.

TbABH is a homolog of AlkB from *Escherichia coli*, an enzyme that oxidatively repairs alkylated DNA, consistent with such a mechanism of DNA repair existing in trypanosomatids. Sequence analysis suggested that TbABH is a nuclear, DNA-binding protein, and electrophoretic mobility shift assays using recombinant TbABH demonstrated preferential binding to alkylated double-stranded DNA. Membership in the $\text{Fe}^{\text{II}}/\alpha\text{KG}$ hydroxylase family was verified by MLCT chromophore formation, however, attempts to detect *in vitro* activity were unsuccessful. *TbABH* was shown to be capable of complementing an *alkB* mutant of *E. coli* subjected to alkylation stress. The complemented cell line was approximately 2-fold more resistant to alkylation-induced growth inhibition than the mutant, confirming the assignment of TbABH as a functional AlkB-like DNA repair protein in *T. brucei*.

To my husband and children,
with respect, gratitude, and love

ACKNOWLEDGEMENTS

I am grateful to the many people who have been a part of my life and a part of my graduate journey in the last six years. First, my utmost thanks go to my advisor, Dr Robert Hausinger, for his ever-present and immensely helpful guidance, encouragement, inspiration, and support. I also want to thank the members of my committee: Dr. Michael Feig, Dr. Charlie Hoogstraten, Dr. Donna Koslowsky, and Dr. Leslie Kuhn. Special thanks goes to Dr. Koslowsky for her assistance with several of the studies presented in this thesis including the epitope tagging of TLP5, the DNA binding studies with TbABH, and the qRT-PCR. I would also like to thank Dr. Kuhn for her help in optimizing and validating the homology model of TLP5.

I have had the pleasure and privilege to work with many wonderful people in the lab over the years including Dr. Scott Mulrooney, Dr. Tina Müller, Dr. Piotr Gryzka, Dr. Lee Macomber, Dr. William Kittleman, Dr. Hajime Masukawa, Dr. Soledad Quiroz, Dr. Meng Li, Dr. Efthalia Kalliri, Dr. Jong Kyong Kim, Mukta Rohankhedkar, Eric Carter, Jodi Boer, Nicholas Flugga, Sandra Hempel, Greg Moyerbrailean, Mike Howard, Megan Andrzejak, Brittnie DeVries, Scott Taber, Chris Towns, Rachel Morr, Andrea Silva, Bruce Fraser, Kim Anderson, Melody Campbell, Scott Elsey, and Matt Fountain. For all of their camaraderie, technical help, and support, I am grateful. I want to especially thank Piotr for all his help with the spectroscopy studies, Greg for his help maintaining the parasite cultures, William for his help with the mass spectrometry studies, Megan for help with the EMSAs using abasic site-containing DNA, and Mike for performing the docking studies with TLP5.

I also need to thank Dr. Henri van Luenen of the Borst lab in the Netherlands for providing JBP1 protein and collaborating on those studies. I would like to thank Dr. Joe Beckmann for encouraging and motivating me and giving me excellent training at Alma College before I began my Ph.D. studies.

I cannot adequately express my love and gratitude toward my family for their support and encouragement. To my husband, Aaron, I could have never done this without you. Finally, I thank God for granting me the ability to undertake such a challenging journey, and the fortitude to complete it.

TABLE OF CONTENTS

LIST OF FIGURES	ix
ABBREVIATIONS	xi
CHAPTER 1	
Introduction.....	1
1.1 Overview of trypanosome biology.....	2
1.1.1 <i>T. brucei</i> life cycle	3
1.1.2 Antigenic variation.....	5
1.1.3 General features of kinetoplastids.....	5
1.1.4 <i>T. brucei</i> replication, chromosome arrangements, and transcription.....	6
1.1.5 Subtelomeric expression of VSG genes.....	11
1.2 Fe ^{II} / α KG-dependent hydroxylases.....	14
1.2.1 Mechanism of Fe ^{II} / α KG hydroxylases.....	14
1.2.2 Hydroxylases in nucleic acid and chromatin metabolism.....	18
1.3 Organization of thesis and open questions.....	20
References.....	22
CHAPTER 2	
<i>Trypanosoma brucei brucei</i> : Thymine 7-hydroxylase-like proteins	28
Abstract.....	29
2.1 Introduction.....	30
2.2 Materials and Methods.....	32
2.2.1 Candidate gene identification	32
2.2.2 Parasites and quantitative real-time PCR.....	32
2.2.3 Construction of TLP5- and TLP7-expression <i>Escherichia coli</i> strains	33
2.2.4 Over-expression and purification of TLP5 and TLP7	34
2.2.5 Antibody production and Western blotting.....	35
2.2.6 Spectroscopy	36
2.2.7 Enzyme assays	36
2.3 Results and Discussion	38
References.....	47
CHAPTER 2 APPENDIX	
2A.1 Introduction.....	51
2A.2 Oligomeric state determination.....	52
2A.3 Homology model.....	54
2A.4 Relocation of purification tag	58
2A.5 C-terminal truncation	59
2A.6 DNA binding.....	60
2A.7 <i>In vivo</i> epitope tagging of the TLP5 gene.....	61
2A.8 Pull-down assays with TLP5 and trypanosome cell extracts.....	66

2A.9 Conclusions.....	67
References.....	68
 CHAPTER 3	
Search for hydroxylase activity in the trypanosomal protein, JBP1	69
3.1 Introduction.....	70
3.2 Materials and Methods.....	74
3.2.1 Determination of succinate production as a measure of uncoupled enzymatic turnover.....	74
3.2.2 Oxygen consumption as a measure of uncoupled enzymatic turnover...76	
3.2.3 Spectroscopic detection of a metal-to-ligand charge-transfer feature	76
3.3 Results and Discussion	78
References.....	81
 CHAPTER 4	
Characterization of a <i>Trypanosoma brucei</i> AlkB homolog capable of repairing alkylated DNA	83
Abstract.....	84
4.1 Introduction.....	85
4.2 Materials and Methods.....	91
4.2.1 Gene identification and multiple sequence alignment	91
4.2.2 Cloning.....	91
4.2.3 Protein production and purification	92
4.2.4 Gel filtration chromatography.....	93
4.2.5 Spectroscopy.....	93
4.2.6 Electrophoretic mobility shift assays	94
4.2.7 <i>In vitro</i> enzyme assays.....	97
4.2.8 Complementation assays.....	98
4.3 Results and Discussion	101
4.3.1 Analysis of the <i>TbABH</i> sequence.....	101
4.3.2 General biochemical properties of <i>TbABH</i>	103
4.3.3 DNA binding by <i>TbABH</i>	105
4.3.4 Examination of DNA repair by <i>TbABH</i>	108
4.3.5 Conclusions.....	112
References.....	113
 CHAPTER 5	
Perspectives.....	118
5.1 Introduction.....	119
5.2 Summary of Results.....	120
5.2.1 TLP5 and TLP7 are unlikely to participate in pyrimidine salvage.....	120
5.2.2 JBP1 is inactive in its recombinant form	121
5.2.3 <i>TbABH</i> is a DNA-binding protein homologous to <i>E. coli</i> AlkB.....	122
5.3 Conclusions and Future Directions.....	124
References.....	127

LIST OF FIGURES

(Images in this dissertation are presented in color)

CHAPTER 1

1.1 Anatomy of <i>Trypanosoma brucei</i>	8
1.2 Schematic of discontinuous transcription and the production of mature mRNA by <i>trans</i> splicing	10
1.3 Bloodstream form subtelomeric expression site organization	13
1.4 General mechanism of Fe ^{II} and α KG dependent hydroxylases	16
1.5 Reaction schemes for nucleic acid modifying hydroxylases	19

CHAPTER 2

2.1 Sequence comparison of TLP5, TLP7, and T7H from <i>R. glutinis</i>	39
2.2 Quantitative real-time PCR analysis of TLP5 and TLP7 expression in BF and PF stages of <i>T. brucei</i>	40
2.3 Electrophoretic analysis of purified TLP5, TLP7, and extracts of PF and BF cells	43
2.4 Spectroscopic evidence for binding of Fe ^{II} and α KG by TLP5	45

CHAPTER 2 APPENDIX

2A.1 Native size and oligomeric state determination of TLP5	53
2A.2 Homology model of TLP5	56
2A.3 Results of docking a subset of the ZINC small molecule database into the modeled active site of TLP5	57
2A.4 Hemagglutinin epitope tagging of TLP5	65

CHAPTER 3

3.1 Synthesis of base J	72
3.2 α KG and succinate detection by HPLC	75

3.3 Spectroscopic evidence for binding of Fe ^{II} and α KG by JBP1	80
CHAPTER 4	
4.1 Reaction of AlkB with 1meA and 3meC in DNA or RNA.....	87
4.2 Structure of AlkB (PDB 2fd8)	88
4.3 Multiple sequence alignment of selected bacterial and eukaryotic AlkB-like proteins.....	102
4.4 Native size and oligomeric state determination of TbABH.....	104
4.5 Spectroscopic evidence for binding of Fe ^{II} and α KG by TbABH	106
4.6 EMSA studies of TbABH and various DNA substrates	107
4.7 Time course viability test with wild-type and $\Delta alkB$ BW25113 cell lines stressed with 0.5% MMS	110
4.8 Complementation of an <i>E. coli alkB</i> mutant with <i>TbABH</i> under alkylation stress.....	111

ABBREVIATIONS

1meA	1-methyl-adenine
3meC	3-methyl-cytocine
α KG	alpha-ketoglutarate
ABH1	AlkB homolog 1
ABH2	AlkB homolog 2
ABH3	AlkB homolog 3
ApoL1	apolipoprotein L1
ATP	adenosine triphosphate
Base J	β -D-glucosylhydroxymethyldeoxyuracil
BCIP/NBT	5-bromo-4-chloro-3-indolyl phosphate/nitro blue tetrazolium chloride
BF	bloodstream form
CFU	colony forming unit
CytC3	cytochrome c3
dNTP	deoxy-nucleotide triphosphate
ds	double stranded
EDTA	ethylenediamine tetraacetic acid
EMSA	electrophoretic mobility shift assay
ESAG	expression site associated gene
GC/MS	gas chromatography/mass spectrometry
HA	hemagglutinin
HPLC	high performance liquid chromatography

HOMedU	hydroxymethyldeoxyuracil
IPTG	isopropyl thiogalactopyranoside
J	see base J
JBP1	J-binding protein 1
JPB2	JBP1 related protein
JmjC	Jumonji C
kDa	kilodalton
LB	lysogeny broth
MLCT	metal-to-ligand charge-transfer
MMS	methylmethane sulfonate
NAD ⁺ (H)	nicotine adenine dinucleotide (oxidized or reduced)
NCBI	National Center for Biotechnology Information
Ni-NTA	nickel-bound nitrilotriacetic acid
OPDA	<i>ortho</i> -phenylenediamine
ORF	open reading frame
PCR	polymerase chain reaction
PDB	protein data bank
PF	procyclic form
PVDF	polyvinylidene fluoride
RNAP	RNA polymerase
RNAi	RNA interference
RT	reverse transcribed
SDS-PAGE	sodium dodecyl sulfate - polyacrylamide gel electrophoresis

SL RNA	splice leader RNA
ss	single stranded
SWI-2/SNF-2	switch/sucrose nonfermentable
TauD	taurine/ α KG dioxygenase
TbABH	<i>Trypanosoma brucei</i> AlkB homolog
T7H	thymine 7-hydroxylase
TLP5	thymine 7-hydroxylase-like protein (from chromosome 5)
TLP5 Δ C13	13 residue C-terminal truncation of TLP5
TLP7	thymine 7-hydroxylase-like protein (from chromosome 7)
UTR	untranslated region
VSG	variable surface glycoprotein

CHAPTER 1

INTRODUCTION

This dissertation focuses on trypanosomal enzymes of the non-heme iron (Fe^{II})- and α -ketoglutarate (αKG)-dependent hydroxylase family, with particular emphasis on proteins that act on nucleic acids and those catalyzing the modification or repair of DNA. To provide a framework for understanding these studies, I divide my introduction into two major sections; first I discuss trypanosomes and their unique biology including key features of gene expression, and then I describe the $\text{Fe}^{\text{II}}/\alpha\text{KG}$ hydroxylase family emphasizing its members involved in nucleic acid modification.

1.1. Overview of Trypanosome Biology

Trypanosomes are eukaryotic unicellular parasites that are the causative agents of several diseases affecting mammals including Human African Trypanosomiasis (commonly called Sleeping Sickness), Chagas disease in humans, and Nagana in cattle (1). These diseases affect millions of people each year and pose a heavy economic burden to the endemic countries due to the loss of livestock. The World Health Organization estimates near 100,000 new cases of Sleeping Sickness each year with only a fraction of those being reported or treated since the most affected populations are mainly rural and among the poorest in Africa. Illustrative of the severity of the problem, the BBC World News produced a documentary aired on October 11, 2008 calling Sleeping Sickness “The Deadliest Disease” on Earth (2) (http://www.survival.tv/documentaries/sleeping_sickness.php). Given the prevalence of trypanosomal diseases, there is great interest in the development of drugs to combat infection. Still, treatments are sorely limited and continued active research is needed to better understand these organisms and to develop improved methods of control. For

example, detailed biochemical analyses of enzymes specific to trypanosomes could suggest targets for future drug development.

Trypanosoma sp. are members of the order kinetoplastidae which also include organisms of the genera *Leishmania*, responsible for additional vertebrate diseases (3), and *Crithidia*, responsible for various invertebrate diseases (4). The Kinetoplastids are a phylogenetically distinct group that diverged early from the central eukaryotic domain before the branching of fungi, plants, and animals (5,6). Therefore, study of their physiology and biological processes provide useful insight not only for the purpose of disease treatment and prevention, but also for understanding the development of this ancient branch of life. Because my work involves *Trypanosoma brucei*, the following sections focus on this microorganism.

1.1.1. *T. brucei* Life Cycle

T. brucei has a complex multi-phase life cycle including a morphologically distinct proliferative stage in an insect vector (Tsetse fly) and another in the targeted mammalian host. The insect stage or procyclic form (PF) of the parasite begins in the gut of the fly when the insect takes a blood meal from an infected mammal. Most parasites are cleared from mature insects and very few actually end up transmitting the trypanosomes to a new host (7). Surviving parasites migrate to the salivary glands of the fly over the course of several days where they transform into a non-proliferative metacyclic form awaiting transfer to a new mammalian host.

When the *T. brucei*-carrying fly bites another mammal, the trypanosomes are introduced into the bloodstream and another major life stage of the parasite begins. The

bloodstream form (BF) is the stage where human disease is manifest. There are several forms of disease caused by the strains of *T. brucei*. *T. brucei brucei* is the strain that infects livestock since humans have an innate immunity to this particular parasite due to apolipoprotein L1 (ApoL1), also known as “trypanosome lytic factor” (8). *T. brucei rhodesiense* is most prevalent in eastern Africa and causes an acute form of the disease in humans that often is fatal after only a few weeks of infection (1). *T. brucei gambiense* is most prevalent in central and southern Africa and causes a chronic form of the disease where the human victim may go undiagnosed for weeks to months (1). In a chronic infection, the parasites eventually cross the blood-brain barrier and colonize the cerebrospinal fluid while also invading other internal organs. Personality changes, impaired mental functions, and severely disrupted circadian rhythms (hence the term Sleeping Sickness) are all symptoms of this second stage trypanosomiasis. If left untreated, the infection leads to coma and death (9).

The ability of *T. b. rhodesiense* and *T. b. gambiense* to effectively colonize humans relates to a serum-resistance-associated gene (10,11). In particular, humans produce ApoL1 which is taken up into the trypanosome cell as part of a high density lipoprotein complex with haptoglobin and hemoglobin (12,13). ApoL1 associates with the cell’s lysosome, forming pores in the membrane and leading ultimately to cell lysis (8). *T. b. rhodesiense* and *T. b. gambiense* produce the serum-resistance-associated protein from an active variable surface glycoprotein (VSG) expression site which binds to the C-terminus of ApoL1 and neutralizes this effect. Recently, it was shown that minor mutations to the C-terminal helix of ApoL1 can ablate this resistance in *T. b. rhodesiense* making it sensitive to recombinant ApoL1 *in vitro* and in transfected mice (14).

1.1.2. Antigenic Variation

In order to maintain an effective colonization of the host, the strictly extracellular parasite must continually deal with the immune system responses to which it is exposed. BF *T. brucei* utilizes a process called antigenic variation that allows the trypanosome to evade the host immune system by periodically changing its expressed VSG coat glycoprotein in a coordinated fashion (10). The genetic arrangement of the VSG-encoding regions and process of expression will be discussed in section 1.1.5. In the PF trypanosomes the surface coat is dominated by proteins called procyclins (or procyclic acidic repetitive proteins). Unlike the BF protists, the PF cells can express several different procyclins at one time and these proteins remain invariant through the course of this life stage. There are two major forms of procyclins termed EP and GPEET. These designations refer to the single amino acid code for the residues found in the repeated regions of the glycoproteins (15).

1.1.3. General Features of Kinetoplastids

T. brucei's surface coatings in its two life stages are characteristic of this species, whereas other features of their biology are shared more broadly by the order kinetoplastidae. Namely, members of this order contain a single mitochondrion (Figure 1.1.6) containing a DNA compartment termed a kinetoplast (Figure 1.1.7) that has a unique genetic content made up of several maxicircles and several hundred concatenated minicircle DNAs. Through elaborate mechanisms beyond the scope of this introduction, these serve to generate the required mRNAs to allow expression of mitochondrial

proteins. In *T. brucei*, the function of the mitochondrion varies greatly throughout the life cycle. In PF parasites, the mitochondrion is large and active with a functional tricarboxylic acid cycle responsible for generation of ATP. In contrast, nutrients are plentiful in the environment of the host bloodstream where the parasite's mitochondrion is greatly reduced in size and quiescent. Glycolysis alone provides ATP to the protist and pyruvate is excreted from the cell (9).

All kinetoplastids are characterized by a single flagellum that extends down the length of the parasite and provides for motility (Figure 1.1.2). The flagellar pocket includes a portal providing the sole means for import and export of cellular metabolites and proteins across the cell wall (Figure 1.1.4) (16). This exclusivity is due in part to the previously mentioned dense coating of glycoproteins covering the rest of the cell surface. In addition to the dense protein coat precluding access from the outside, the subpellicular microtubules are too closely spaced to allow transport vesicles access to the plasma membrane from the inside (Figure 1.1.3). *T. brucei* undergoes very high rates of non-specific, exclusively clathrin-dependent endocytosis through this pocket so that the surface proteins are constantly recycled. BF parasites also express several transferrin receptors in the flagellar pocket which are responsible for uptake of host ferritin and provide the iron needed by the cell (16).

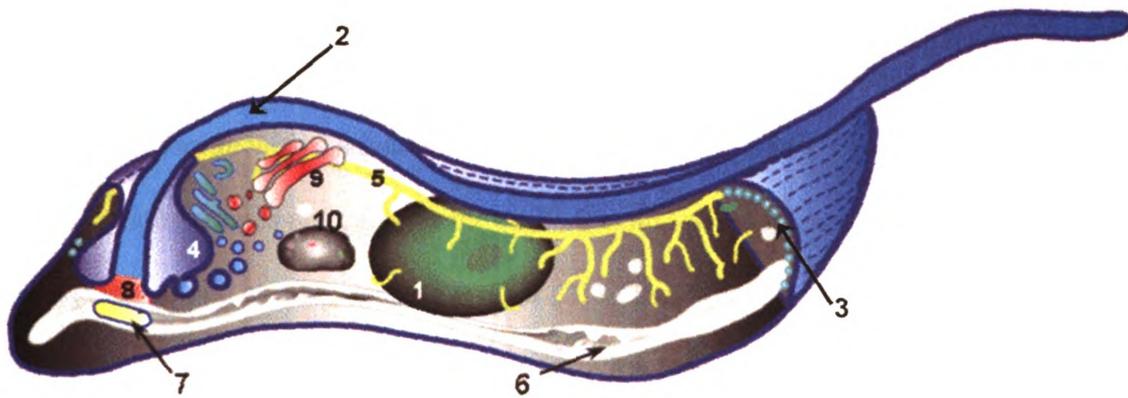
1.1.4. *T. brucei* Replication, Chromosome Arrangements, and Transcription

T. brucei replication exhibits several unique properties compared to more typical eukaryotes. They contain single copies of their Golgi complex (Figure 1.1.9) and lysosome (Figure 1.1.10), thus creating a challenge for replication as each of these

organelles needs to be copied and dutifully partitioned to the daughter cells (17). The nuclear genetic arrangement in these cells is very streamlined, and the kinetoplast nearly lacks intronic regions and possesses some overlapping genes. *T. brucei* contains 11 linear megabase-sized chromosomes containing the vast majority of the protein-coding genes and gene arrays, several intermediate chromosomes that contain repetitive elements and occasionally a VSG expression site, and up to hundreds of mini-chromosomes containing mostly pseudogenes and repertoires of VSG-encoding genes amongst highly repetitive sequences (18,19). The latter regions are believed to provide a repertoire of nearly inexhaustible options for recombination to create new varieties of VSGs.

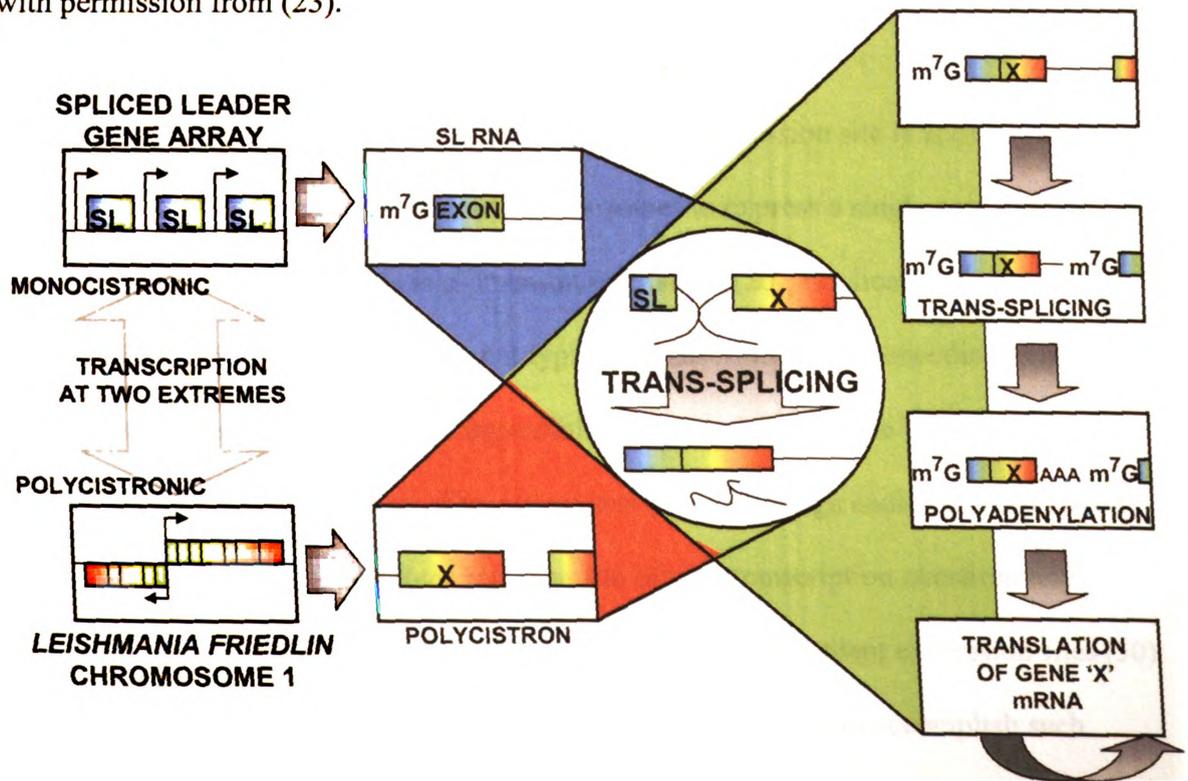
Trypanosomes are diploid for their large chromosomes. These contain most, if not all, protein-encoding genes except for some VSG-encoding genes. The exact sizes of the chromosomes vary greatly between trypanosome populations due to large repetitive elements and tandem gene arrays, which give rise to high rates of recombination and unequal crossing over events. Mini-chromosomes also vary greatly in size as they are made up of typical eukaryotic telomeres linked by large tracts of a 177-bp repeat with usually one VSG-encoding gene or pseudogene. The ploidy of the mini-chromosomes is less certain due to a lack of distinct chromosomal markers (18).

Figure 1.1: Anatomy of *Trypanosoma brucei*. 1. nucleus 2. flagellum 3. subpellicular microtubules 4. flagellar pocket 5. endoplasmic reticulum 6. mitochondrion 7. kinetoplast 8. basal body 9. Golgi body 10. lysosome. Figure adapted with permission from Dr. Markus Engstler.



One thing markedly lacking from the nuclear genome of *T. brucei* and other trypanosomatids is the presence of typical RNA polymerase (RNAP) II promoters. Conventional eukaryotic protein-coding gene transcription involves the formation of initiation complexes that congregate at an RNAP II promoter with many genetic regulatory elements affecting the transcription initiation of each open reading frame (ORF) (20). In contrast, trypanosomes have long poly-cistronic genetic arrangements on each chromosome with a strand switch region somewhere in the middle from which transcription initiates and radiates outwards toward the telomeres, with one set of ORFs on one strand and another set on the other (21). This leaves only the 5' and 3' untranslated regions between the ORFs to provide regulation, almost all of which occurs post-transcriptionally (22). Therefore, gene products needed in abundance, such as tubulin, are arranged in repetitive gene arrays in order to obtain greater amounts of the mRNA. That mRNA is then stabilized to allow for greater levels of translation. Due to this unique mechanism for generation of transcripts, trypanosomes process pre-mRNAs by *trans*-splicing. As depicted in Figure 1.2, a 39-nucleotide spliced leader sequence (SL-RNA) is transcribed from the RNAP II promoter (using an RNAP that is altered in its subunit composition when compared to that involved in poly-cistron transcription) in great abundance (100-200 tandem copies per genome) and is *trans*-spliced onto the 5' end of every mRNA produced in the nucleus. These are then polyadenylated based on the location in the intergenic region of the 5' splice acceptor site (23).

Figure 1.2 Schematic of discontinuous transcription and the production of mature mRNA by *trans* splicing. Transcription of short discrete tandemly repeated genes (SL RNA) and polycistronic protein-coding regions (exemplified by *L. major* chromosome 1 (24)) by RNAP II is indicated at the left. The primary transcripts from the two classes of gene are shown in the center intersecting with steps in the *trans* splicing pathway (right panel). The bimolecular *trans* splicing reaction resembles the unimolecular *cis* splicing pathway in many aspects, including the initial cleavage of the SL RNA splice donor site, the formation of 5'-2' phosphodiester bond on a branch-point A residue, cleavage at the pre-mRNA splice acceptor site and ligation of the two exons. Differences between the two processes include the formation of a Y-branched intermediate (rather than a lariat-like intermediate) and the coupled polyadenylation of the upstream mRNA. Figure adapted with permission from (23).



1.1.5. Subtelomeric Expression of VSG genes

Another unique genetic feature of *T. brucei* relates to its expression of genes in subtelomeric regions of the megabase chromosomes. These regions, spanning up to 2 kb upstream of the telomeric repeats, include the VSG expression sites (Fig. 1.3).

Trypanosomes contain approximately 20 of these sites which are made up of various repetitive DNA satellites, the VSG gene, and several other expression site-associated genes (ESAGs). These genes encode variations of transferrin receptors (ESAG 6 and 7) for iron sequestration from the host, a homolog of a bipterin transporter (ESAG 10), several proteins likely involved in the differentiation between life stages (ESAG 4 and 8), other protein-encoding genes of unknown function, and the serum-resistance-associated protein responsible for the ability of *T. b. rhodesiense* and *T. b. gambiense* to effectively colonize humans (10,11,25).

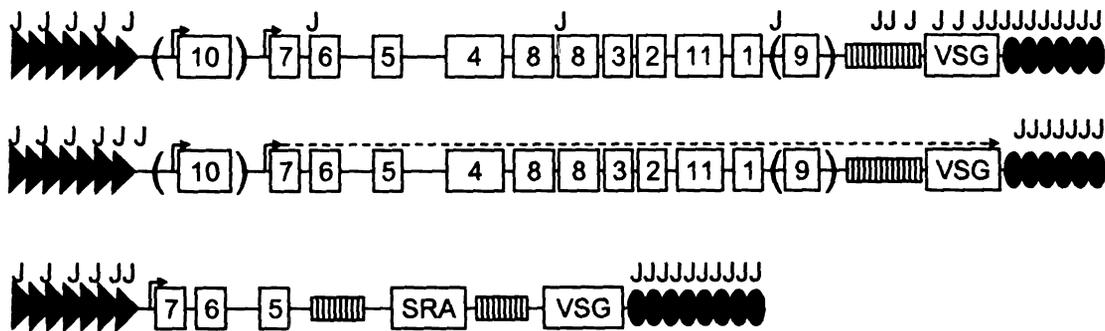
During trypanosome infection, only one VSG expression site is active at any given time, allowing the population of trypanosomes to express a single coat protein. Transcription of the VSG gene is α -amanitin insensitive (26), implicating the involvement of RNAP I which does not typically transcribe protein-encoding genes. RNAP I was shown to be directly responsible for transcription of the VSG and procyclin surface protein-encoding genes of trypanosomes (27). Although each VSG expression site has its own RNAP I promoter and seems to initiate transcription constitutively (28,29), transcription aborts before reaching the VSG gene in silent expression sites (30).

The silencing mechanisms employed by these parasites to accomplish such coordinated gene expression have been studied for many years, though they remain

largely a mystery. Positioning effects mediated by the subtelomeric localization have been investigated with some contradictory results, but it does seem that there is some promoter restraint, especially in BF trypanosomes (31,32). In addition, a modified base found in telomeric and subtelomeric regions has also been implicated in gene silencing. β -D-glucosylhydroxymethyluridine, known as base J (33), replaces up to 1% of the thymidine in trypanosomal DNA (34). This novel base is localized to telomeric repeats, subtelomeric repeats surrounding the VSG expression sites, and sequences within VSG expression sites not being expressed (Fig. 1.3). However, the one active VSG expression site is void of the modification (35). Further, the active VSG expression site is sequestered in the “expression site body”, an area of the nucleus that is separate from the nucleolus; this subnuclear localization may also play a role in its discriminatory expression (36).

Trypanosomes can switch the expression of their VSG by several mechanisms. The most common mode of switching is by keeping the same active VSG expression site and swapping out the VSG gene with another by homologous recombination. Alternatively, trypanosomes can splice together new VSGs by recombining part of the currently expressed VSG with a pseudogene from a minichromosome or intermediate chromosome. Another mechanism is to silence the current VSG expression site while activating another. This approach may be less common as it would require more chromatin rearrangement, though it would also allow the expression of variant transferrin receptors and other ESAGs.

Figure 1.3 Bloodstream form subtelomeric expression site organization. Figure adapted from (25). ESAGs are depicted by numbered open boxes (those genes present in only some expression sites are shown in parentheses). Relative location of base J is shown above the expression site diagrams and is based on immunoprecipitation of fragmented DNA with an anti-J antibody (35). Filled triangles represent 50 bp repeats, hatched boxes represent 70 bp repeats, and filled ovals represent telomeric repeats. The top diagram depicts a silent expression site with base J found throughout the coding sequences while the center diagram depicts an active expression site, devoid of base J and fully transcribed (dashed arrow). Finally, the bottom diagram depicts a specialized expression site of *T. b. rhodesiense* containing the serum resistance associated gene.



1.2. *Fe^{II}/αKG-Dependent Hydroxylases.*

The Fe^{II}/αKG-dependent hydroxylases use a mononuclear ferrous ion to catalyze the oxidative decarboxylation of αKG while activating molecular oxygen in order to hydroxylate a target substrate. This broad enzyme superfamily has representatives throughout the domains of bacteria and eukaryota, with other members found in viruses, and they catalyze a wide variety of reactions (37),(38). These enzymes are characterized by a double-stranded β-helix core fold, often termed a β-jellyroll structure that is made up of 8 β strands within which lies the active site (39). Section 1.2.1 will discuss the overall reaction mechanism and Section 1.2.2 will focus on those members whose primary function is the modification of nucleic acids or chromatin.

1.2.1. *Mechanism of Fe^{II}/αKG hydroxylases.*

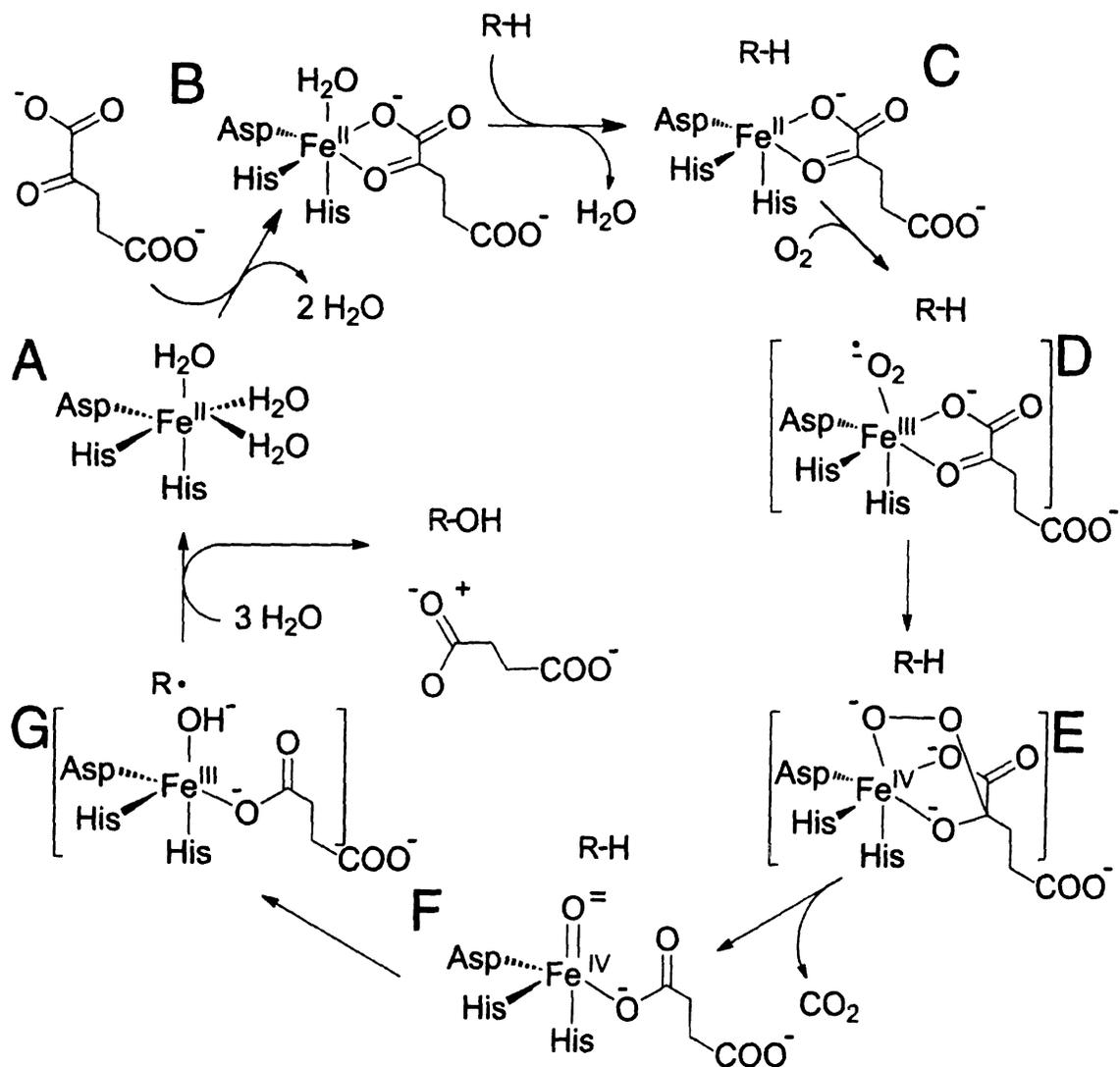
The current view of the enzyme mechanism for these enzymes, summarized in Figure 1.4, is a modification of that first proposed over 25 years ago (40). Two histidines and one carboxylic acid residue (usually an aspartate) coordinate one face of the iron atom (41). This “facial triad” usually occurs in a His-X-Asp/Glu-X_n-His motif where n is typically around 50 residues, with the remaining 3 coordination sites on the metal atom taken up by water molecules in the resting state (Fig. 1.4 panel A). Upon addition of the cosubstrate, αKG, two waters are displaced by the bidentate binding of the αKG molecule via its C1 carboxylate and C2 keto groups (Fig. 1.4 panel B). The C5 carboxylate of αKG often is stabilized by a conserved arginine located approximately a dozen residues C-terminal to the second His residue of the facial triad (38). This enzyme state leaves

one iron coordination site occupied by a water molecule that is displaced by the binding of the primary substrate (Fig. 1.4 panel C).

The two α KG-bound states can be visualized spectroscopically. Thus, in the absence of primary substrate and oxygen, but the presence of iron and α KG, these enzymes exhibit a metal-to-ligand charge-transfer electronic transition that is observed at around 500 to 530 nm (42). For example, taurine/ α KG dioxygenase (TauD), a well-studied member responsible for the oxidative cleavage of taurine to release sulfite during sulfur starvation in *Escherichia coli*, forms this characteristic chromophore at 530 nm with an extinction coefficient of $140 \text{ M}^{-1} \text{ cm}^{-1}$ (42). This feature can be used to diagnose proper protein folding and cofactor/cosubstrate binding. It is also used to confirm membership in this enzyme family for proteins without a verified function or a known primary substrate. Once the primary substrate binds to the protein, near the active-site but not coordinating the iron atom, this feature often shifts. In the case of TauD, taurine binding causes a 10 nm shift in the absorbance to higher energy and increases the intensity of the chromophore to give an extinction coefficient of $180 \text{ M}^{-1} \text{ cm}^{-1}$ at 520 nm (42).

Displacement of the final water molecule leaves an open coordination site on the iron atom that provides a site for molecular oxygen to coordinate the iron. This process likely creates two very short-lived species including an Fe^{III} -superoxo species and a bridging Fe^{IV} -peroxo species that have never been observed (Fig. 1.4 panels D and E). Decomposition of the latter species leads to cleavage of the cosubstrate into the products succinate and CO_2 as well as the formation of a highly reactive Fe^{IV} -oxo species which

Figure 1.4 General mechanism of Fe^{II} and α KG dependent hydroxylases. Primary substrate is generalized and abbreviated by R-H.



was directly observed first for TauD (43-45), and later for prolyl hydroxylase (46) and a related halogenase, CytC3 (47) (Fig. 1.4 panel F).

The highly reactive Fe^{IV} oxo species then abstracts a hydrogen atom from the primary substrate generating a likely Fe^{III} -hydroxo species and a substrate radical (Fig. 1.4 panel G). The traditional view is that the hydroxyl radical then rebounds with the carbon radical, thus hydroxylating the substrate (which spontaneously degrades into sulfite and aminoacetaldehyde in the case of taurine) and regenerating the ferrous ion at the active-site. Very recent studies suggest this step is more complex, involving an Fe^{III} -oxo species and an alkoxo species (not depicted), ultimately yielding the same products (48). The products are then released from the active-site and are replaced by three waters to regenerate the resting state for the enzyme (Fig. 1.4 panel A).

This reaction cycle can be tightly coupled, resulting in the utilization of stoichiometric amounts of primary substrate, αKG , and O_2 for each turnover (38). However, in the absence of the primary substrate or the presence of alternate substrates or inhibitors, these enzymes can decarboxylate αKG and activate oxygen to generate reactive oxygen species that can pose a threat to the cell. Some of this damage seems to be avoided by “fail safe” mechanisms involving side chain self-hydroxylation (49,50) or reaction of the activated species with a cellular reductant rather than releasing deleterious oxygen species into the milieu of the cell (51). This “uncoupling” of oxygen activation and substrate hydroxylation can be observed by monitoring oxygen or αKG consumption and succinate production by various methods and has been a useful tool to aid in the characterization of proteins of unknown function.

1.2.2. Hydroxylases in Nucleic Acid and Chromatin Metabolism.

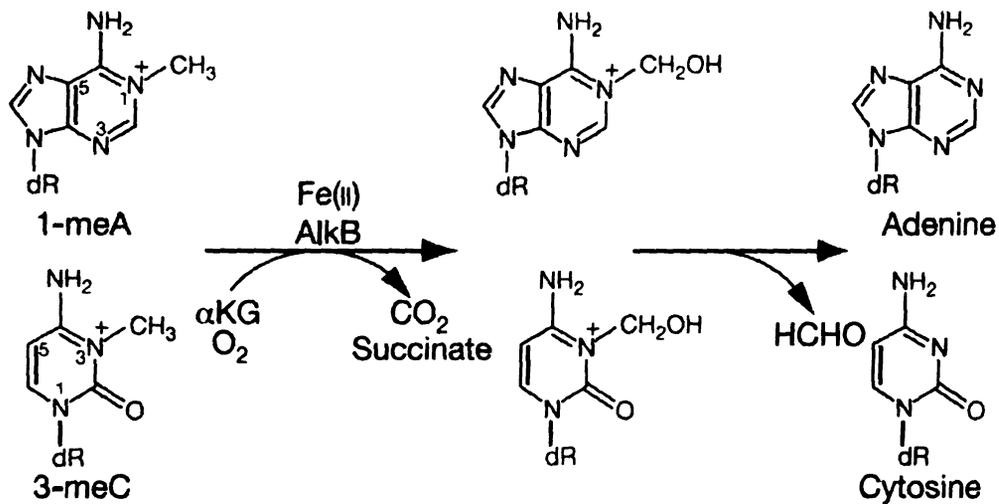
Several members of the Fe^{II}/αKG hydroxylase enzyme family utilize this powerful oxidizing reaction to modify nucleic acids. One of the best studied enzymes in this subgroup is AlkB, an *E. coli* protein, that repairs alkyl lesions on DNA and RNA (52,53). This enzyme follows the same basic mechanism described above to oxidize methyl adducts on the N-1 position of adenine (1meA), the N-3 position of cytosine (3meC), and selected other lesions in the context of DNA (Fig. 1.5 panel A). The hydroxylated methyl groups are spontaneously released from the molecule as formaldehyde, regenerating the unmodified bases.

Multiple complexes of AlkB along with its cofactor, cosubstrate, and primary substrates have been crystallized and show that single-stranded nucleic acids are preferred by this enzyme with the strand bent into the active site, thus flipping the alkylated base deep into the pocket where chemistry takes place (54). Further, this protein has a 'nucleotide recognition lid' that alternates between an open and closed state upon binding of the substrate, providing more stability for the transition state and facilitating the enclosed environment needed for the oxygenation reaction. AlkB has been demonstrated to function in the repair of both DNA and RNA, with methylated, ethylated, and 1,N⁶-etheno modified bases. The physiological relevance of reactions involving more bulky lesions is not clear given other pathways in the cell, such as base excision repair; however, the substrate versatility (plasticity of the active site) suggests that AlkB may be able to serve as a back up system of repair for these lesions (55).

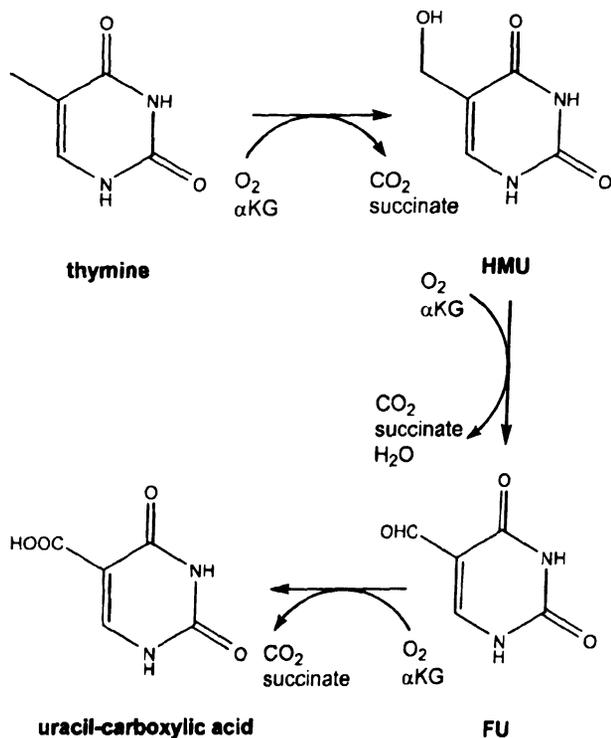
Another example of nucleic acid hydroxylation is found in a fungal enzyme, thymine 7-hydroxylase, which participates in pyrimidine salvage. This enzyme catalyzes

Figure 1.5 Reaction schemes for nucleic acid modifying hydroxylases. **A.** Predominant reactions catalyzed by *E. coli* AlkB. **B.** Reactions catalyzed by thymine 7-hydroxylase. HMU is 5-hydroxymethyluracil. FU is 5-formyluracil.

A



B.



the 3-step oxidation of the free base, thymine, to uracil-carboxylic acid, through hydroxymethyluracil and 5-formyluracil as illustrated in Figure 1.5 panel B (56). The kinetics of the later two reactions are very slow, calling into question their physiological relevance (57). Further studies show that the three reactions are not processive, but rather the product leaves the active site and must rebind along with another molecule of α KG and O_2 to allow a new reaction cycle to commence (57).

Other members of this enzyme family that participate in nucleic acid hydroxylation include xanthine hydroxylase (58,59), which converts the free base xanthine to uric acid, and deoxyribonucleoside-2'-hydroxylase, which converts thymidine or deoxyuridine to the free base (60). In addition, the chromatin modifying JmjC domain-containing histone demethylases remove methyl groups from lysines in histones (61). The aforementioned enzymes encompass a substrate range from free bases to chromatin and highlight the versatility and utility of a well controlled highly reactive oxygenation system in the cell for maintaining and modifying nucleic acids and proteins.

1.3. Organization of this Thesis and Open Questions

Chapter 2 of this thesis explores two thymine 7-hydroxylase-like proteins (TLP5 and TLP7) in *T. b. brucei* including sequence analysis, cloning, and expression of their genes, protein purification and characterization, and the *in vitro* search for their activity. Chapter 3 describes my part in the characterization of J-binding protein 1 (a thymidine hydroxylase) of various kinetoplastids, again involving protein characterization and an *in vitro* search for enzymatic activity. Chapter 4 includes the characterization of a *T. b. brucei* homolog of AlkB including sequence analysis, cloning, and expression of its gene,

purification of the protein, and both *in vitro* and *in vivo* searches for activity. Finally, Chapter 5 summarizes my overall conclusions and describes the remaining open questions in the area of trypanosomal Fe^{II}/αKG hydroxylases.

REFERENCES

1. Barrett, MP, Burchmore, RJ, Stich, A, Lazzari, JO, Frasch, AC, Cazzulo, JJ, Krishna, S. 2003. The trypanosomiases. *Lancet* 362: 1469-1480
2. Bain, R. 2008. The Deadliest Disease. In *Survival*: BBC World News
3. Herwaldt, BL. 1999. Leishmaniasis. *Lancet* 354: 1191-1199
4. Otterstatter, MC, Thomson, JD. 2008. Does pathogen spillover from commercially reared bumble bees threaten wild pollinators? *PLoS One* 3: e2771
5. Stevens, JR, Noyes, HA, Schofield, CJ, Gibson, W. 2001. The molecular evolution of Trypanosomatidae. *Adv Parasitol* 48: 1-56
6. Pace, NR. 2009. Mapping the tree of life: progress and prospects. *Microbiol Mol Biol Rev* 73: 565-576
7. Roditi, I, Lehane, MJ. 2008. Interactions between trypanosomes and tsetse flies. *Curr Opin Microbiol* 11: 345-351
8. Pays, E, Vanhollebeke, B. 2009. Human innate immunity against African trypanosomes. *Curr Opin Immunol* 21: 493-498
9. Molyneux, DH, Ashford, R. W. 1983. *The Biology of Trypanosoma and Leishmania, Parasites of Man and Domestic Animals*. New York: International Publications Service. 267 pp.
10. Horn, D, Barry, JD. 2005. The central roles of telomeres and subtelomeres in antigenic variation in African trypanosomes. *Chromosome Res* 13: 525-533
11. Hertz-Fowler, C, Figueiredo, LM, Quail, MA, Becker, M, Jackson, A, Bason, N, Brooks, K, Churcher, C, Fahrenberg, S, Goodhead, I, Heath, P, Kartvelishvili, M, Mungall, K, Harris, D, Hauser, H, Sanders, M, Saunders, D, Seeger, K, Sharp, S, Taylor, JE, Walker, D, White, B, Young, R, Cross, GA, Rudenko, G, Barry, JD, Louis, EJ, Berriman, M. 2008. Telomeric expression sites are highly conserved in *Trypanosoma brucei*. *PLoS One* 3: e3527
12. Widener, J, Nielsen, MJ, Shiflett, A, Moestrup, SK, Hajduk, S. 2007. Hemoglobin is a co-factor of human trypanosome lytic factor. *PLoS Pathog* 3: 1250-1261
13. Vanhollebeke, B, De Muylder, G, Nielsen, MJ, Pays, A, Tebabi, P, Dieu, M, Raes, M, Moestrup, SK, Pays, E. 2008. A haptoglobin-hemoglobin receptor

- conveys innate immunity to *Trypanosoma brucei* in humans. *Science* 320: 677-681
14. Lecordier, L, Vanhollebeke, B, Poelvoorde, P, Tebabi, P, Paturiaux-Hanocq, F, Andris, F, Lins, L, Pays, E. 2009. C-terminal mutants of apolipoprotein L-I efficiently kill both *Trypanosoma brucei brucei* and *Trypanosoma brucei rhodesiense*. *PLoS Pathog* 5: e1000685
 15. Roditi, I, Furger, A, Ruepp, S, Schürch, N, Bütikofer, P. 1998. Unravelling the procyclin coat of *Trypanosoma brucei*. *Mol Biochem Parasitol* 91: 117-130
 16. Field, MC, Carrington, M. 2009. The trypanosome flagellar pocket. *Nat Rev Microbiol* 7: 775-786
 17. Hammarton, TC, Monnerat, S, Mottram, JC. 2007. Cytokinesis in trypanosomatids. *Curr Opin Microbiol* 10: 520-527
 18. Ersfeld, K, Melville, SE, Gull, K. 1999. Nuclear and Genome Organization of *Trypanosoma brucei*. *Parasitology Today* 15: 58-63
 19. Horn, D. 2001. Nuclear gene transcription and chromatin in *Trypanosoma brucei*. *Int J Parasitol* 31: 1157-1165
 20. Sikorski, TW, Buratowski, S. 2009. The basal initiation machinery: beyond the general transcription factors. *Curr Opin Cell Biol* 21: 344-351
 21. Palenchar, JB, Bellofatto, V. 2006. Gene transcription in trypanosomes. *Mol Biochem Parasitol* 146: 135-141
 22. Clayton, C, Shapira, M. 2007. Post-transcriptional regulation of gene expression in trypanosomes and leishmanias. *Mol Biochem Parasitol* 156: 93-101
 23. Campbell, DA, Thomas, S, Sturm, NR. 2003. Transcription in kinetoplastid protozoa: why be normal? *Microbes Infect* 5: 1231-1240
 24. Myler, PJ, Audleman, L, deVos, T, Hixson, G, Kiser, P, Lemley, C, Magness, C, Rickel, E, Sisk, E, Sunkin, S, Swartzell, S, Westlake, T, Bastien, P, Fu, G, Ivens, A, Stuart, K. 1999. *Leishmania major* Friedlin chromosome 1 has an unusual distribution of protein-coding genes. *Proc Natl Acad Sci U S A* 96: 2902-2906
 25. Vanhamme, L, Lecordier, L, Pays, E. 2001. Control and function of the bloodstream variant surface glycoprotein expression sites in *Trypanosoma brucei*. *Int J Parasitol* 31: 523-531
 26. Lee, MG, Van der Ploeg, LH. 1997. Transcription of protein-coding genes in trypanosomes by RNA polymerase I. *Annu Rev Microbiol* 51: 463-489

27. Gunzl, A, Bruderer, T, Laufer, G, Schimanski, B, Tu, LC, Chung, HM, Lee, PT, Lee, MG. 2003. RNA polymerase I transcribes procyclin genes and variant surface glycoprotein gene expression sites in *Trypanosoma brucei*. *Eukaryot Cell* 2: 542-551
28. Pays, E, Coquelet, H, Tebabi, P, Pays, A, Jefferies, D, Steinert, M, Koenig, E, Williams, RO, Roditi, I. 1990. *Trypanosoma brucei*: constitutive activity of the VSG and procyclin gene promoters. *Embo J* 9: 3145-3151
29. Ansorge, I, Steverding, D, Melville, S, Hartmann, C, Clayton, C. 1999. Transcription of 'inactive' expression sites in African trypanosomes leads to expression of multiple transferrin receptor RNAs in bloodstream forms. *Mol Biochem Parasitol* 101: 81-94
30. Vanhamme, L, Poelvoorde, P, Pays, A, Tebabi, P, Van Xong, H, Pays, E. 2000. Differential RNA elongation controls the variant surface glycoprotein gene expression sites of *Trypanosoma brucei*. *Mol Microbiol* 36: 328-340
31. Cross, GA, Wirtz, LE, Navarro, M. 1998. Regulation of vsg expression site transcription and switching in *Trypanosoma brucei*. *Mol Biochem Parasitol* 91: 77-91
32. Glover, L, Horn, D. 2006. Repression of polymerase I-mediated gene expression at *Trypanosoma brucei* telomeres. *EMBO Rep* 7: 93-99
33. Gommers-Ampt, JH, Van Leeuwen, F, de Beer, AL, Vliegthart, JF, Dizdaroglu, M, Kowalak, JA, Crain, PF, Borst, P. 1993. beta-D-glucosyl-hydroxymethyluracil: a novel modified base present in the DNA of the parasitic protozoan *T. brucei*. *Cell* 75: 1129-1136
34. van Leeuwen, F, Taylor, MC, Mondragon, A, Moreau, H, Gibson, W, Kieft, R, Borst, P. 1998. beta-D-glucosyl-hydroxymethyluracil is a conserved DNA modification in kinetoplastid protozoans and is abundant in their telomeres. *Proc Natl Acad Sci U S A* 95: 2366-2371
35. van Leeuwen, F, Wijsman, ER, Kieft, R, van der Marel, GA, van Boom, JH, Borst, P. 1997. Localization of the modified base J in telomeric VSG gene expression sites of *Trypanosoma brucei*. *Genes Dev* 11: 3232-3241
36. Navarro, M, Gull, K. 2001. A pol I transcriptional body associated with VSG mono-allelic expression in *Trypanosoma brucei*. *Nature* 414: 759-763
37. Simmons, JM, Muller, TA, Hausinger, RP. 2008. Fe(II)/alpha-ketoglutarate hydroxylases involved in nucleobase, nucleoside, nucleotide, and chromatin metabolism. *Dalton Trans*: 5132-5142

38. Hausinger, RP. 2004. Fe(II)/ α -ketoglutarate-dependent hydroxylases and related enzymes. *Crit. Rev. Biochem. Mol. Biol.* 39: 21-68
39. Clifton, IJ, McDonough, MA, Ehrismann, D, Kershaw, NJ, Granatino, N, Schofield, CJ. 2006. Structural studies on 2-oxoglutarate oxygenases and related double-stranded beta-helix fold proteins. *J Inorg Biochem* 100: 644-669
40. Hanauske-Abel, HM, Günzler, V. 1982. A stereochemical concept for the catalytic mechanism of prolylhydroxylase. Applicability to classification and design of inhibitors. *J. Theor. Biol.* 94: 421-455
41. Hegg, EL, Que, L, Jr. 1997. The 2-His-1-carboxylate facial triad. *Eur. J. Biochem.* 250: 625-629
42. Ryle, MJ, Padmakumar, R, Hausinger, RP. 1999. Stopped-flow kinetic analysis of *Escherichia coli* taurine/ α -ketoglutarate dioxygenase: interactions with α -ketoglutarate, taurine, and oxygen. *Biochemistry* 38: 15278-15286
43. Price, JC, Barr, EW, Tirupati, B, Bollinger, JM, Jr., Krebs, C. 2003. The first direct characterization of a high-valent iron intermediate in the reaction of an α -ketoglutarate-dependent dioxygenase: a high-spin Fe(IV) complex in taurine/ α -ketoglutarate dioxygenase (TauD) from *Escherichia coli*. *Biochemistry* 42: 7497-7508
44. Proshlyakov, DA, Henshaw, TF, Monterosso, GR, Ryle, MJ, Hausinger, RP. 2004. Direct detection of oxygen intermediates in the non-heme Fe enzyme taurine/ α -ketoglutarate dioxygenase. *J. Am. Chem. Soc.* 126: 1022-1023
45. Riggs-Gelasco, PJ, Price, JC, Guyer, RB, Brehm, JH, Barr, EW, Bollinger, JM, Jr., Krebs, C. 2004. EXAFS spectroscopic evidence for an Fe=O unit in the Fe(IV) intermediate observed during oxygen activation by taurine: α -ketoglutarate dioxygenase. *J. Am. Chem. Soc.* 126: 8108-8109
46. Hoffart, LM, Barr, EW, Guyer, RB, Bollinger, JM, Jr., Krebs, C. 2006. Direct spectroscopic detection of a C-H-cleaving high-spin Fe(IV) complex in a prolyl-4-hydroxylase. *Proc Natl Acad Sci U S A* 103: 14738-14743
47. Fujimori, DG, Barr, EW, Matthews, ML, Koch, GM, Yonce, JR, Walsh, CT, Bollinger, JM, Jr., Krebs, C, Riggs-Gelasco, PJ. 2007. Spectroscopic evidence for a high-spin Br-Fe(IV)-oxo intermediate in the alpha-ketoglutarate-dependent halogenase CytC3 from *Streptomyces*. *J Am Chem Soc* 129: 13408-13409
48. Grzyska, PK, Appelman, EH, Hausinger, RP, Proshlyakov, DA. 2010. Insight into the mechanism of an iron dioxygenase by resolution of steps following the FeIV=HO species. *Proc Natl Acad Sci U S A* 107: 3982-3987

49. Liu, A, Ho, RYN, Que, L, Jr., Ryle, MJ, Phinney, BS, Hausinger, RP. 2001. Alternative reactivity of an α -ketoglutarate-dependent iron(II) oxygenase: enzyme self-hydroxylation. *J. Am. Chem. Soc.* 123: 5126-5127
50. Ryle, MJ, Liu, A, Muthukumaran, RB, Ho, RYN, Koehntop, KD, McCracken, J, Que, L, Jr., Hausinger, RP. 2003. O₂- and α -ketoglutarate-dependent tyrosyl radical formation in TauD, an α -keto acid-dependent non-heme iron dioxygenase. *Biochemistry* 42: 1854-1862
51. Saari, RE, Hausinger, RP. 1998. Ascorbic acid-dependent turnover and reactivation of 2,4-dichlorophenoxyacetic acid/ α -ketoglutarate dioxygenase using thiophenoxyacetic acid. *Biochemistry* 37: 3035-3042
52. Trewick, SC, Henshaw, TF, Hausinger, RP, Lindahl, T, Sedgwick, B. 2002. Oxidative demethylation by *Escherichia coli* AlkB directly reverts DNA base damage. *Nature* 419: 174-178
53. Falnes, PO, Johansen, RF, Seeberg, E. 2002. AlkB-mediated oxidative demethylation reverses DNA damage in *Escherichia coli*. *Nature* 419: 178-182
54. Yu, B, Edstrom, WC, Hamuro, Y, Weber, PC, Gibney, BR, Hunt, JF. 2006. Crystal structures of catalytic complexes of the oxidative DNA/RNA repair enzyme AlkB. *Nature* 439: 879-884
55. Sikora, A, Mielecki, D, Chojnacka, A, Nieminuszczy, J, Wrzesinski, M, Grzesiuk, E. 2009. Lethal and mutagenic properties of MMS-generated DNA lesions in *Escherichia coli* cells deficient in BER and AlkB-directed DNA repair. *Mutagenesis*
56. Wondrack, LM, Hsu, CA, Abbott, MT. 1978. Thymine 7-hydroxylase and pyrimidine deoxyribonucleoside 2'-hydroxylase activities in *Rhodotorula glutinis*. *J Biol Chem* 253: 6511-6515
57. Thornburg, LD, Lai, M-T, Wishnok, JS, Stubbe, J. 1993. A non-heme iron protein with heme tendencies: an investigation of the substrate specificity of thymine hydroxylase. *Biochemistry* 32: 14023-14033
58. Cultrone, A, Scazzocchio, C, Rochet, M, Montero-Moràn, G, Drevet, C, Fernández-Martin, R. 2005. Convergent evolution of hydroxylation mechanisms in the fungal kingdoms: molybdenum cofactor-independent hydroxylation of xanthine via α -ketoglutarate dependent dioxygenase. *Molec. Microbiol.* 57: 276-290
59. Montero-Moran, GM, Li, M, Rendon-Huerta, E, Jourdan, F, Lowe, DJ, Stumpff-Kane, AW, Feig, M, Scazzocchio, C, Hausinger, RP. 2007. Purification and

characterization of the FeII- and alpha-ketoglutarate-dependent xanthine hydroxylase from *Aspergillus nidulans*. *Biochemistry* 46: 5293-5304

60. Shaffer, PM, Cairns, JR, Dorman, DC, Lott, AM. 1984. Separation and characterization of pyrimidine deoxyribonucleoside 2'-hydroxylase and thymine 7-hydroxylase from *Aspergillus nidulans*. *Int. J. Biochem.* 16: 429-434
61. Tsukada, Y-I, Fang, J, Erdjument-Bromage, H, Warren, ME, Borchers, CH, Tempst, P, Zhang, Y. 2006. Histone demethylation by a family of JmjC domain-containing proteins. *Nature* 439: 811-816

CHAPTER 2

***TRYPANOSOMA BRUCEI BRUCEI*: THYMINE 7-HYDROXYLASE-LIKE PROTEINS**

These studies were published in:

Simmons, JM, Koslowsky, DJ, Hausinger, RP. 2009. *Trypanosoma brucei brucei*: Thymine 7-Hydroxylase-Like Proteins. *Experimental Parasitology* **124**: 453-458.

The quantitative reverse-transcriptase PCR experiments were performed by Dr. Donna Koslowsky and the data analysis for the anaerobic chromophore experiment was performed by Dr. Piotr Grzyska. Mass Spectrometry experiments were performed with the aid of Dr. William Kettleman. Additional studies not included in the publication are provided in an appendix.

ABSTRACT

Two genes from *Trypanosoma brucei brucei* are predicted to encode Fe^{II}- and α -ketoglutarate-dependent enzymes related to fungal thymine 7-hydroxylase. Transcript levels of the thymine hydroxylase-like genes are higher in the bloodstream-form of the parasite over the insect form, whereas Western blot analysis indicates more cross-reactive protein in the latter life stage. The genes were cloned, the proteins purified from *Escherichia coli*, and both proteins were shown to bind Fe^{II} and α -ketoglutarate, confirming proper folding. The isolated proteins were incubated with Fe^{II} and α -ketoglutarate plus thymine, thymidine, and other putative substrates, but no activity was detected. Furthermore, no thymine 7-hydroxylase activity was detected in extracts of procyclic or bloodstream form cells. Although the functions of these proteins remain unknown, we conclude they are unlikely to be involved in thymine salvage.

2.1 INTRODUCTION

Trypanosomes are the causative agents of several mammalian diseases, including African Sleeping Sickness and Chagas disease in humans and Nagana in cattle (1). They are transmitted by a suite of insect vectors and infect the bloodstream of their mammalian hosts when the insect takes a blood meal. The genome of *Trypanosoma brucei brucei* was recently elucidated (2), providing important insights into the physiology of this microorganism.

This work describes two *T. b. brucei* genes, located on chromosomes 5 and 7 (Tb927.5.300 and Tb927.7.7500, respectively), which encode full-length orthologues of thymine 7-hydroxylase (T7H) of *Rhodotorula glutinis* (3,4) and are termed TLP5 and TLP7 for thymine hydroxylase-like proteins. T7H is a non-heme iron (Fe^{II})- and α -ketoglutarate (αKG)-dependent hydroxylase that catalyzes sequential oxidations of thymine (Fig. 1.5B) (4,5) as part of a pyrimidine salvage pathway in some fungi (3,6). Hydroxylation of the unactivated C7 methyl group of thymine is driven by the oxidative decarboxylation of αKG , forming succinate and CO_2 . Significantly, trypanosomes have not been reported to contain a pyrimidine salvage pathway and show no mediated uptake of thymine or thymidine (7), nor does the kinetoplastid genome contain an obvious candidate gene encoding uracil-5-carboxylate decarboxylase, the enzyme acting immediately downstream in the salvage pathway. Further, trypanosomes seem to have no direct need for pyrimidine salvage as they are known to be able to synthesize pyrimidine nucleotides (8).

To date, three gene products in trypanosomes are predicted to belong to the $\text{Fe}^{\text{II}}/\alpha\text{KG}$ dependent hydroxylase family. One gene is suggested to encode AlkB, involved in oxidative repair of DNA damaged by alkylation (9). Two others, encoding JBP1 and JBP2 are proposed to catalyze the hydroxylation of selected thymidine bases in trypanosomal DNA as the first step in formation of base J, β -D-glucosyl-hydroxymethyl-deoxyuracil (10,11). Significantly, no direct evidence for hydroxylase activity has been reported for any of these proteins. Here, we examine whether the *TLP5* and *TLP7* genes encode T7H isozymes or have alternative functions.

2.2 MATERIALS AND METHODS

2.2.1. Candidate gene identification

To identify candidate thymine hydroxylase genes, the Basic Local Alignment Search Tool (12) was utilized to search the *T. b. brucei* genome with the gene sequence of *R. glutinis* thymine 7-hydroxylase (4) as the query, resulting in the identification of NCBI accession numbers XM_839611.1 and XM_841277.1. In addition, the available sequences of the kinetoplastid order were searched with the gene encoding uracil-5-carboxylate decarboxylase (3), the next enzyme in the fungal pyrimidine salvage pathway, as a query.

2.2.2. Parasites and quantitative real-time PCR

Transcript levels were compared using *T. b. brucei* strain 427 bloodstream form (BF) paired with strain 427 (MITat-1) procyclic form (PF) and BF 427-2 (221) paired with PF 427, 29-13. Briefly, total RNA was treated with turbo DNase (Ambion) according to the manufacturer's directions and 1 µg of RNA was then reverse transcribed (RT) in the presence of random hexamers (20 mM Tris 8.3, 50 mM KCl, 5 mM MgCl₂, 2 mM mixed dNTPs, 0.25 units RNAsin and 2 units Seikagaku AMV reverse transcriptase). The RT reaction was stopped by heating at 70°C and diluted with 125 µl of 20 mM Tris, 1 mM ethylenediamine tetraacetic acid, pH 8.0. PCR amplification reactions (SYBR Green, Bio-Rad) were set up in triplicate (primers: chromosome 5 forward, 5'-GGT TGG GTA GAG TTG ATG AAC-3'; chromosome 5 reverse, 5'-TGG AGG ATA ATG TAG CAT ACG-3'; chromosome 7 forward, 5'-CAC ACT ATC GCG

ATA TGC GGG AC-3’; chromosome 7 reverse, 5’-GCG GGG TAA TGC ACC ATG CG-3’) using 2 µl of the diluted RT reaction. All reported data were normalized to β -tubulin gene internal control (primers: forward, 5’ GAC GAA GGA GGT TGA TGA GCA GAT 3’; reverse, 5’ TGA AGG TGA CAG CCA TCT TGA GTC 3’) that was included in each run.

2.2.3. Construction of *TLP5*- and *TLP7*- expressing *Escherichia coli* strains

A 963-bp DNA fragment containing *TLP5* was amplified by PCR using genomic *T. b. brucei* strain 427 DNA as a template with “forward” (5’-AGG ATA TAC CAT GGC TCA CGG CTC GAT T-3’) and “reverse” (5’-GAG CAT CCT CGA GCA TCT TTG TTT TGC GAT G-3’) primers which introduce *NcoI* and *XhoI* restriction sites (underlined), and *Taq* polymerase master mix kit (Promega) which leaves a single 3’ adenine nucleotide overhang. The PCR product was treated with a PCR clean up kit (Qiagen, Inc.) and ligated into pGEM-T Easy (Promega). This was transformed into *E. coli* DH5 α (Invitrogen), isolated from several transformants, and sequenced (Davis Sequencing). *TLP5* was excised from the pGEM-T backbone by *NcoI* and *XhoI* restriction and ligated into pET28b (Novagen) which had been previously cut with the same enzymes, putting the coding sequence in frame with a sequence encoding a C-terminal 6-histidine tag. This plasmid was transformed into the expression strain *E. coli* BL21 (DE3).

Similarly, a 963-bp DNA fragment containing *TLP7* was amplified from genomic *T. b. brucei* DNA using “forward” (5’-GGA TAT ACC ATG GTA ATG AAT CAC ACT TCG ATT CCA G-3’) and “reverse” (5’-GAG CAT CCT CGA GCA TAT TTG CCT

TGC G-3') primers, ligated into pGEM-T Easy, and transformed into *E. coli* DH5a. Plasmids isolated from several transformants were sequenced, one containing a minor A186V change was excised from the pGEM-T backbone by using NcoI and XhoI, and this fragment was ligated into pET28b to yield a sequence encoding TLP7 in frame with a C-terminal 6-histidine tag. The recombinant plasmid was transformed into *E. coli* BL21 (DE3) for expression.

2.2.4. Over-expression and purification of TLP5 and TLP7

E. coli BL21 (DE3) cells containing the pET28b derivatives encoding TLP5-His₆ or TLP7-His₆ (hereafter referred to simply as TLP5 and TLP7) were grown at 30 °C in LB medium supplemented with 100 µg/mL kanamycin while shaking at ~160 rpm to an optical density of 0.4 to 0.6 at 600 nm. Cultures were induced to overexpress the desired genes by addition of isopropyl-β-D-thiogalactopyranoside to 0.1 mM, and grown for an additional 4 h, then harvested at 4 °C by centrifugation at ~8,000 g for 8 min. The cell paste was either used immediately for protein purification or stored at -80 °C.

In a typical purification, 2.5 mL of binding buffer (30 mM imidazole, 10 mM Tris, 150 mM NaCl, pH 7.9) was added per gram of cell paste for resuspension. The protease inhibitor phenylmethylsulfonyl fluoride was added to 0.5 mM, cells were lysed by sonication (Branson Sonifier, 3 pulses of 1 min each, 3 W output power, duty cycle 50%, with cooling on ice), and the cell lysates were ultracentrifuged at 100,000 g for 1 h. Soluble cell extracts were loaded onto a Ni-bound nitrilotriacetic acid (NTA) column (Qiagen) pre-equilibrated with binding buffer. The column was washed with binding buffer until the baseline was reestablished, and proteins were released with elution buffer

(150 mM imidazole, 10 mM Tris, 150 mM NaCl, pH 7.9). Fractions containing the purified proteins, as determined by denaturing sodium dodecyl sulfate polyacrylamide gel (12%) electrophoresis (SDS-PAGE) (13) and Coomassie staining, were pooled and exchanged back into binding buffer by using Amicon centrifugal filter units with a 10 kDa molecular weight cutoff. The concentrations of proteins were determined by using their calculated molar absorptivities at 280 nm ($44,390 \text{ M}^{-1} \text{ cm}^{-1}$ for TLP5 and $49,890 \text{ M}^{-1} \text{ cm}^{-1}$ for TLP7 according to the ExPasy protein parameters prediction server (<http://ca.expasy.org/tools/protparam.html>). The proteins were stored at 4 °C and discarded after two weeks or if precipitation developed.

2.2.5. Antibody production and Western blotting

Purified recombinant TLP5 was provided to Lampire Biological Laboratories for polyclonal antibody production in a rabbit. The IgG fraction was isolated from the product antisera by ammonium sulfate precipitation and stored at 4 °C. Trypanosomes (10^7 cells in 20 μL cell wash buffer with 0.5 mM phenylmethylsulfonyl fluoride) were disrupted in a bath sonicator for 20 min. The lysate was centrifuged at $\sim 16,000 g$ for 20 min at 4 °C, and soluble cell extracts were removed to a clean tube and used immediately or stored at -20 °C. Purified recombinant proteins or cell extracts of trypanosomes (both PF and BF) were subjected to 12% SDS-PAGE and Western blots were prepared by following standard protocols using Millipore ImmobilonP membrane. Protein bands were transferred at 4 °C either using 100 V for 1 h or 14 V overnight. Blots were probed with a 1:10,000 dilution of the rabbit anti-TLP5 antibody and a 1:30,000 dilution of goat

anti-rabbit antibody conjugated to alkaline phosphatase from Sigma. Blots were developed using the BCIP/NBT liquid substrate system from Sigma.

2.2.6. Spectroscopy

UV/visible spectra were obtained by using an HP 8453 spectrophotometer (Hewlett Packard) equipped with circulating water bath and magnetic stirrer. TLP5 and TLP7 proteins were concentrated in Amicon centrifugal filter units with 10 kDa cutoff to 250 μ M in binding buffer. Stock solutions of 10 mM Fe^{II} and 5 mM α KG were made by adding the dry reagents to vials, subjecting them to repeated cycles of vacuum and argon, and adding anaerobic water using a syringe. Protein was added to a 1-cm path length, 300 μ L, black-walled cuvette and made anaerobic by the same method followed by scanning with stirring at 9 °C. Data were corrected for baseline shifts arising from cuvette repositioning to have a uniform absorbance at 900 nm by using the software IGOR.

2.2.7. Enzyme assays

Oxygen consumption assays were performed in a YSI Inc. model 5300 Clark-style biological oxygen monitor coupled to a circulating water bath to maintain a temperature of 33 °C. TLP5 or TLP7 was assayed at 1 μ M in 25 mM imidazole, 100 mM KCl, and 5 mM MgCl₂ using variations of the standard conditions: 50 μ M Fe^{II}, 500 μ M α KG, 400 μ M ascorbate, and 500 μ M of the potential substrates.

Succinate production assays utilized 1 μ M enzyme that was incubated with 10 μ M Fe^{II}, 10 μ M ascorbate, and 100 μ M α KG (and sometimes MgCl₂) along with potential

substrates in 25 mM Tris HCl, pH 8.0, at either 28 °C (PF conditions) or 37 °C (BF conditions) for 30 min. Aliquots (300 µL) were taken at various time points, quenched with 5 µL of 6 M sulfuric acid, centrifuged at 13,800 g for 5 min, passed through a 0.45 µm spin filter, and 200 µL was analyzed by Aminex HPX-87H (BioRad) ion-exchange chromatography (Waters, Corp.). Organic acids were eluted with 0.013 M H₂SO₄ as the mobile phase and detected using a refractive index detector (Waters, Corp.) set at 35 °C and a sensitivity level of 16. Standards of αKG and succinate eluted at 19 and 26 min, respectively, and succinate was detectable to a lower limit of 1 nmol.

Purified TLP5 and TLP7 (both at 10 µM), and trypanosome whole cell protein extracts from both the PF and BF life stages (from ~6 x 10⁶ and ~5 x 10⁶ cells, respectively) were incubated with thymine (250 µM) in 1 mM HEPES, pH 7.4, at 30 °C for 1 h with added αKG (500 µM) and ferrous ammonium sulfate (100 µM). Reactions were stopped by addition of hot ethanol, incubated at 60 °C for 5 min, and dried in a speed vac for 2h. Lyophilized pellets were resuspended in 100 µL anhydrous acetonitrile and silylated via treatment with 100 µL *N*-methyl-*N*-(*tert*-butyldimethylsilyl)trifluoroacetamide + 1% *tert*-butyldimethylchlorosilane (Pierce) for 30 min at 135 °C. The solutions were analyzed by gas chromatography-mass spectrometry (GC-MS) using electrospray ionization on a 30 m DB-5MS, I.D. 0.25 cm column for analysis as previously described (4).

2.3. RESULTS AND DISCUSSION

The *T. b. brucei* gene products TLP5 and TLP7 are approximately 30% identical to the fungal T7H sequence and are 85% identical to each other at the amino acid level (Fig. 2.1). All residues suspected to be critical to the active site are conserved, including those likely to coordinate the metal (His 200, Asp 202, and His 257 for both TLP5 and TLP7 numbering) and a residue that could stabilize binding of the α KG cosubstrate (Arg 266). The TLP5 and TLP7 sequences contain several short gaps (of 2, 7, 2, and 8 amino acid residues) compared to the fungal sequence, but the significance of these gaps, if any, cannot be ascertained because the structure of T7H is unknown.

Quantitative real-time PCR methods were used to examine the steady-state expression levels of the two genes in both the BF and PF life stages of *T. b. brucei*. Due to the high sequence identity of *TLP5* and *TLP7* (87% at the DNA level), none of the available programs for primer selection in trypanosomes could be used. Therefore, gene sequences were manually analyzed and primers carefully chosen in order to be unique to each gene. Triplicate runs with mRNA containing *TLP5* and *TLP7* from *T. b. brucei* strain 427 (MiTat 1) (BF and PF), 427 clone 221 (BF), and 427 cell line 29-13 (PF) yielded similar results consistent with both genes being up-regulated in the BF parasite over the insect-form by approximately 4 and 2.5 fold, respectively (Fig. 2.2). This seemingly small effect is considered significant as only about 2% of genes in trypanosomes have been found to be regulated at least 2 fold at the level of mRNA abundance (14).

Figure 2.1. Sequence comparison of TLP5, TLP7, and T7H from *R. glutinis*. Identities are indicated by black shading. Putative metal- or α KG-binding residues are highlighted by asterisks.

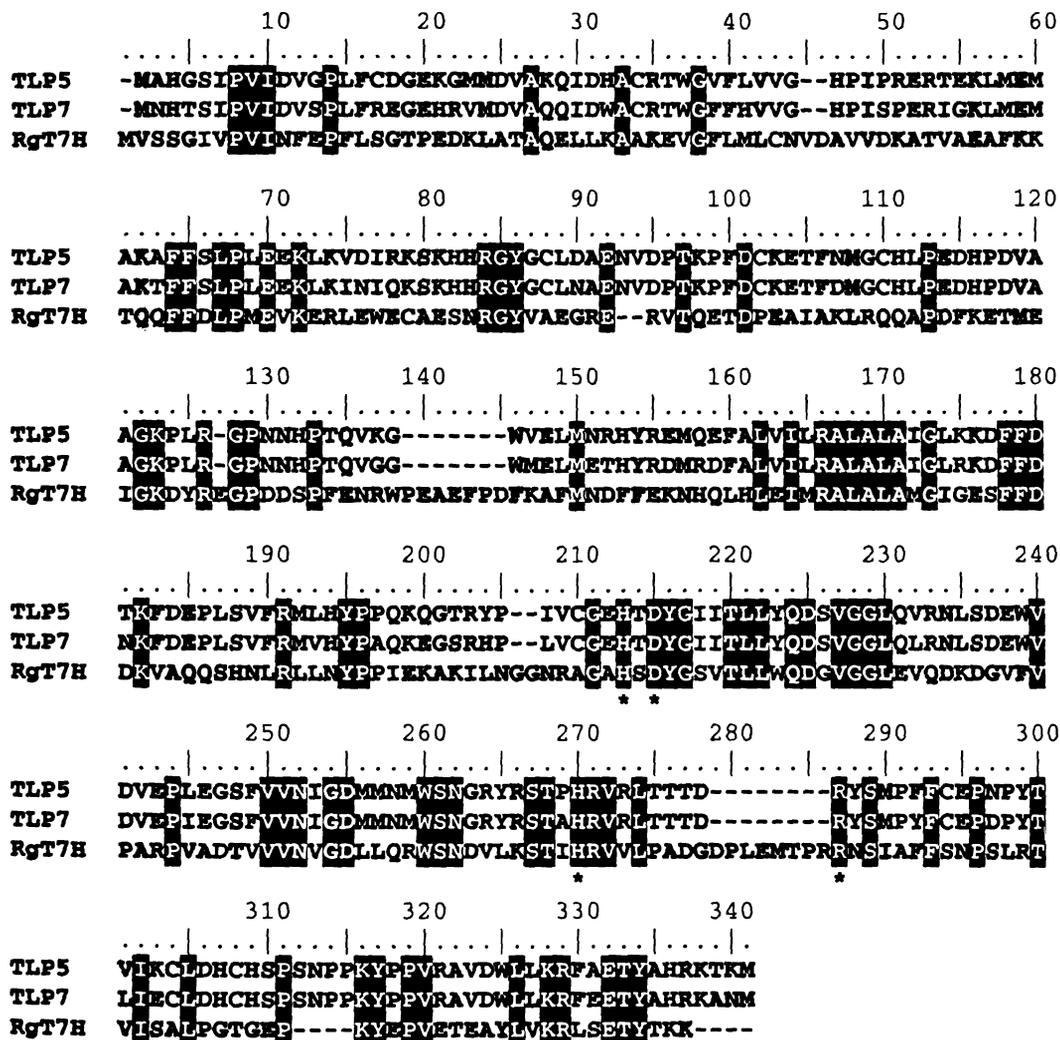
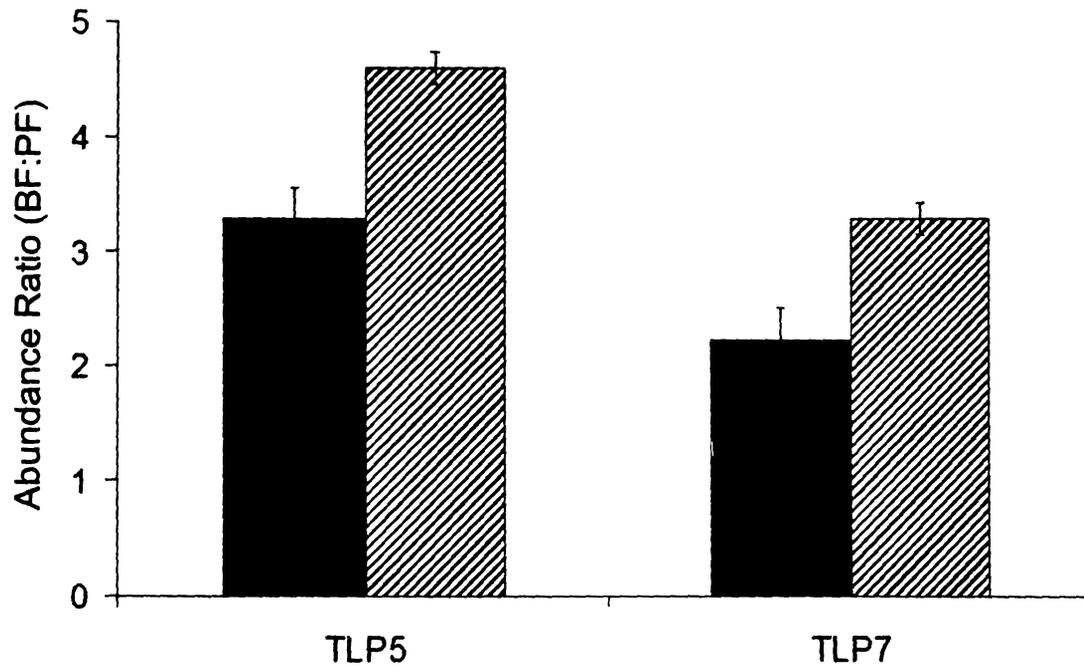


Figure 2.2. Quantitative real-time PCR analysis of *TLP5* and *TLP7* expression in BF- and PF-stages of *T. b. brucei*. The solid bars represent the fold increase of BF over PF in 427 (MITat-1) for the two genes. The hatched bars compare the fold increase of the BF 427 (221) over the PF 29-13. All data are normalized to tubulin internal controls that were included in each run.



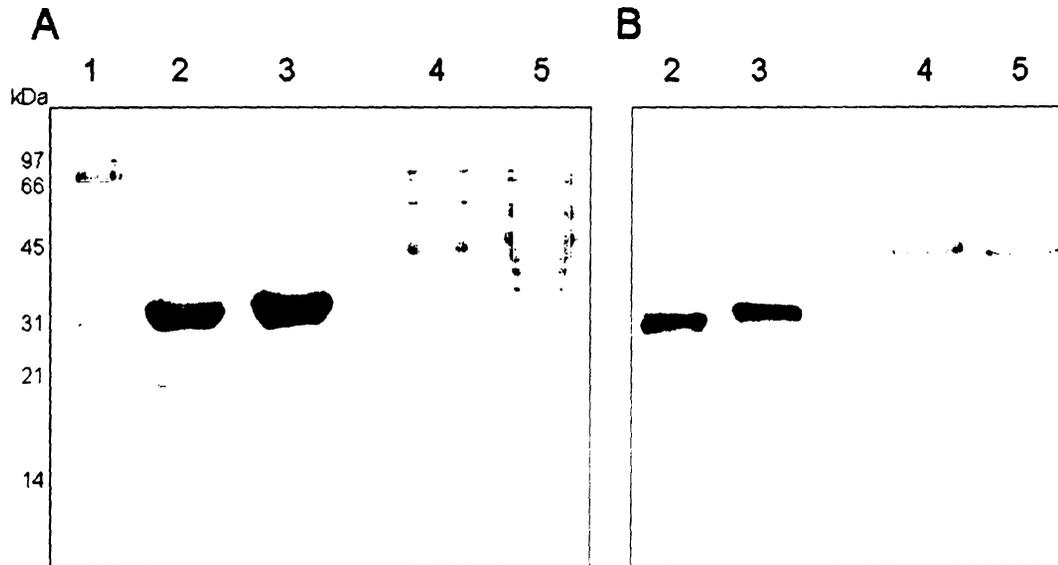
TLP5 and *TLP7* were PCR amplified out of *T. b. brucei* 29-13 genomic DNA and subcloned into pET28b for expression as His-tagged proteins in *E. coli* BL21 (DE3). Sequencing of plasmids isolated from individual colonies revealed a number of differences from the published *T. b. brucei* TREU 927 sequence, most involving amino acid changes in non-catalytic areas of the protein. Of potential interest were a frame shift mutation in *TLP5* and a nonsense mutation in *TLP7* detected in multiple clones from independent PCR amplifications suggesting that one allele may be non-functional in this strain. This phenomenon may be due to an issue inherent to the trypanosomes being studied. The 29-13 cell line used in this study was developed from strain 427 which has been in laboratory culture for over 30 years and is now incapable of differentiating to the BF. Loss of the ability to differentiate back into infective blood forms may reflect the mutation of genes essential for function during this stage without conferring negative selection in culture.

E. coli BL21 (DE3) cells containing the pET28b *TLP5* and *TLP7* derivatives were used as the source of purified *TLP5* and *TLP7*, typically yielding 5 mg of protein per gram of wet cellular weight. The proteins were >95% homogeneous (Fig. 2.3 panel A, lanes 2 and 3) and of the size predicted by the sequences (37.5 kDa when accounting for the His tag), with slight differences in electrophoretic mobility observed. The proteins behaved as monomers when examined by Blue Native gels (Invitrogen) or Superdex 75 gel filtration chromatography (data not shown). Both proteins precipitated when exchanged into a low salt buffer after purification, but exchange back into the column binding buffer provided enhanced stability.

Polyclonal anti-TLP5 antibody was used to perform Western blots of purified recombinant TLP5 and TLP7 along with cell extracts of both PF and BF trypanosomes (Fig. 2.3 panel B). The antibody detected both proteins, as expected from their close similarity. A major cross-reactive species in both BF and PF electrophoresed more slowly than the TLP5/TLP7 doublet; this species may be an unrelated contaminant or it might arise from post-translational modification of these proteins (leading to an apparent ~10-kDa increase in size) in the trypanosome, but lacking in the *E. coli* expression system. Future work could include efforts to identify this band and determine what, if any, modifications are added to the protein *in vivo*. In contrast to the RT-PCR results that indicate higher transcript levels in the BF cells, the immunological analysis suggests that TLP5/7 is more abundant in PF than BF cells; these results are compatible with the protein(s) being turned over more rapidly in the BF life stage or being translated at a greater rate in the PF life stage. Of primary importance, these results indicate that both proteins appear to be present in both life stages.

To confirm proper folding of TLP5 and TLP7, the purified proteins were examined for their ability to form a characteristic chromophore seen in other members of the Fe^{II}/αKG-dependent dioxygenases and associated with binding of the iron cofactor and the αKG cosubstrate (15). The proteins were monitored by UV-visible spectroscopy under anaerobic conditions, using low temperature to maintain stability, while titrating in αKG and Fe^{II}. As shown in Figure 2.4, anaerobic TLP5 generated the diagnostic metal-to-ligand charge-transfer transition at 505 nm when both metal and cofactor were present. The extinction coefficient for this feature was approximately 40 M⁻¹ cm⁻¹, indicating that the chromophore was incompletely formed compared with previously described

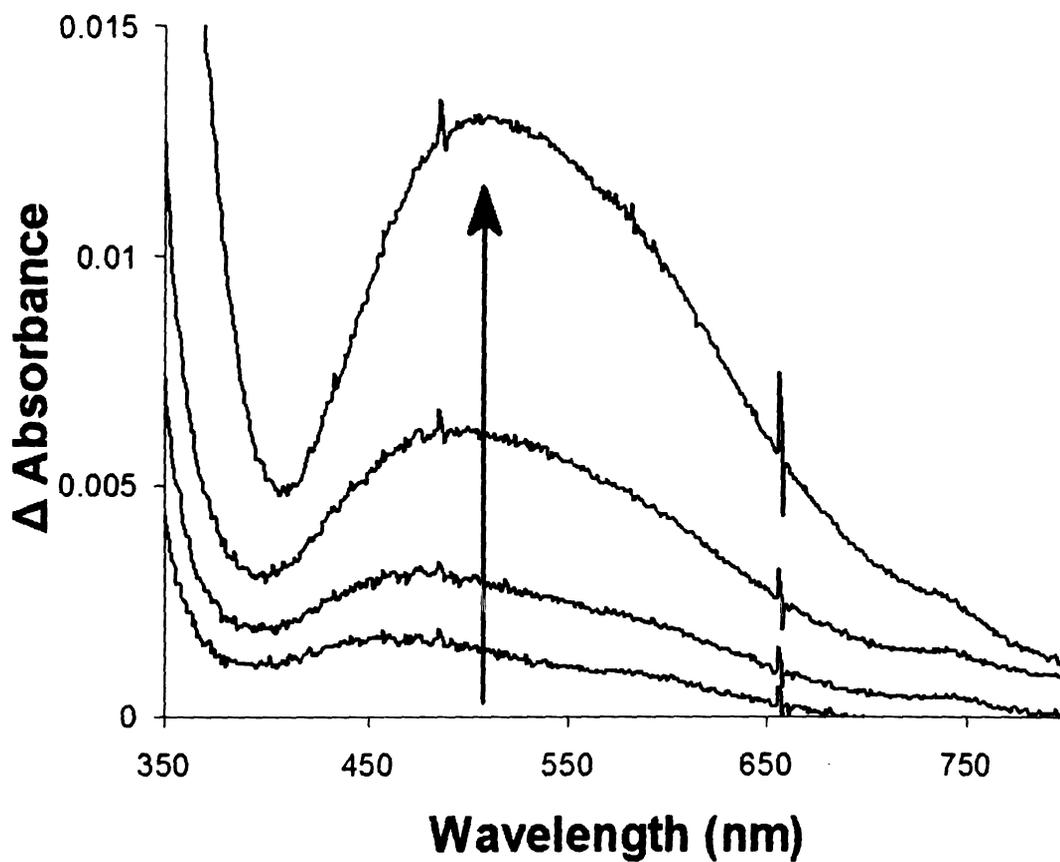
Figure 2.3. Electrophoretic analysis of purified TLP5, TLP7, and extracts of PF and BF cells. Samples were subjected to SDS-PAGE and examined by **A.** Coomassie staining or **B.** Western blot analysis with anti-TLP5 antibodies. Lanes: 1, molecular weight ladder; 2, TLP5; 3, TLP7; 4, extracts of PF cells; 5, extracts of BF cells.



intensities (15), but addition of more metal ions led to protein precipitation. TLP7 also exhibited the diagnostic transition (data not shown), confirming correct folding of the protein, but it was less stable than TLP5 and precipitation during the experiment precluded estimation of the extinction coefficient.

Despite extensive efforts, no *in vitro* thymine hydroxylase activity was detected for purified recombinant TLP5 or TLP7 when assayed by using an oxygen electrode to monitor oxygen consumption, HPLC to assess succinate production, or GC-MS to measure thymine hydroxylation. We cannot exclude the possibility that a post-translational modification, not occurring in *E. coli*, is required for activity. It is also conceivable that the His-tag hinders activity in the recombinant proteins or that, despite formation of the Fe^{II}/αKG chromophore indicating appropriate folding of the active site domains, the full length proteins do not fold properly. On the other hand, these results are compatible with the proteins utilizing alternative substrates. Other members of the Fe^{II}/αKG dioxygenase family recognize a wide array of substrates ranging from small molecules to proteins and DNA (16). Several additional small molecules were tested as putative substrates including other bases, nucleosides, and amino acids, but again no activity was detected. With regard to the possible use of DNA as a substrate, trypanosomes are known to contain the hyper-modified base J in their nuclear DNA (17,18), which requires hydroxylation of an unactivated methyl group of thymidine in DNA to create hydroxymethyl-deoxyuracil (19,20) that is subsequently glucosylated. Thus, we examined whether TLP5 or TLP7 reacted with polynucleotides or thymidine; however no such activity was observed. Furthermore, neither protein bound to linear DNA fragments under any conditions examined (incubation with or without 50 μM Fe^{II},

Figure 2.4. Spectroscopic evidence for binding of Fe^{II} and αKG by TLP5. The anaerobic UV/visible spectrum of TLP5 (250 μM in binding buffer) was examined for the sample as isolated, after adding 1 mM αKG , and while titrating in Fe^{II} with stirring at 9 $^{\circ}\text{C}$. The ($\text{Fe}^{\text{II}}\text{-}\alpha\text{KG}\text{-protein}$ minus $\alpha\text{KG}\text{-protein}$) difference spectra shown correspond to the addition of 0.5, 1, 2, and 3 equivalents of metal ions.



150 μM αKG , and 100 μM MgCl_2 at 37 $^\circ\text{C}$ for 30 min) according to gel band shift studies. Thus, TLP5 and TLP7 are unlikely to possess DNA modifying activity, consistent with the recent proposals that J-binding protein 1 (JBP1), identified on the basis of its ability to bind to J-containing chromatin, and JBP2, related to JBP1, are critically involved in base J synthesis (10,21-24).

Extracts of PF and BF cells were directly assayed for thymine hydroxylase activity by using GC-MS. No significant hydroxylase activity could be detected (i.e., less than 0.5% product formed from 250 μM thymine) after a 1 h incubation.

In conclusion, we have identified two genes encoding T7H-like proteins in *T. b. brucei*, demonstrated that the genes are expressed at greater levels in the BF over the PF stage, shown that both proteins are present in each life stage (possibly with greater abundance in the PF life stage), and expressed the genes in *E. coli*. We purified the two recombinant proteins, demonstrated that they form an $\text{Fe}^{\text{II}}/\alpha\text{KG}$ chromophore (indicating proper folding), and tested for catalytic activity and DNA binding. Our results provide no indication of the recombinant proteins being able to hydroxylate free thymine or thymidine within DNA or to bind DNA. Furthermore, no hydroxylase activity was detected in extracts from trypanosomes of either life stage. The roles of these proteins in *T. b. brucei* remain unknown, but our evidence provides strong arguments against their participation in pyrimidine salvage.

REFERENCES

1. Barrett, MP, Burchmore, RJ, Stich, A, Lazzari, JO, Frasch, AC, Cazzulo, JJ, Krishna, S. 2003. The trypanosomiasis. *Lancet* 362: 1469-1480
2. Berriman, M, Ghedin, E, Hertz-Fowler, C, Blandin, G, Renauld, H, Bartholomeu, DC, Lennard, NJ, Caler, E, Hamlin, NE, Haas, B, Bohme, U, Hannick, L, Aslett, MA, Shallom, J, Marcello, L, Hou, L, Wickstead, B, Alsmark, UCM, Arrowsmith, C, Atkin, RJ, Barron, AJ, Bringaud, F, Brooks, K, Carrington, M, Cherevach, I, Chillingworth, T-J, Churcher, C, Clark, LN, Corton, CH, Cronin, A, Davies, RM, Doggett, J, Djikeng, A, Feldblyum, T, Field, MC, Fraser, A, Goodhead, I, Hance, Z, Harper, D, Harris, BR, Hauser, H, Hostetler, J, Ivens, A, Jagels, K, Johnson, D, Johnson, J, Jones, K, Kerhornou, AX, Koo, H, Larke, N, Landfear, S, Larkin, C, Leech, V, Line, A, Lord, A, MacLeod, A, Mooney, PJ, Moule, S, Martin, DMA, Morgan, GW, Mungall, K, Norbertczak, H, Ormond, D, Pai, G, Peacock, CS, Peterson, J, Quail, MA, Rabinowitsch, E, Rajandream, M-A, Reitter, C, Salzberg, SL, Sanders, M, Schobel, S, Sharp, S, Simmonds, M, Simpson, AJ, Tallon, L, Turner, CMR, Tait, A, Tivey, AR, Van Aken, S, Walker, D, Wanless, D, Wang, S, White, B, White, O, Whitehead, S, Woodward, J, Wortman, J, Adams, MD, Embley, TM, Gull, K, Ullu, E, Barry, JD, Fairlamb, AH, Opperdoes, F, Barrell, BG, Donelson, JE, Hall, N, Fraser, CM, Melville, SE, El-Sayed, NM. 2005. The genome of the African trypanosome *Trypanosoma brucei*. *Science* 309: 416-422
3. Smiley, JA, Kundracik, M, Landfried, DA, Barnes, VR, Sr., Axhemi, AA. 2005. Genes of the thymidine salvage pathway: thymine-7-hydroxylase from a *Rhodotorula glutinis* cDNA library and iso-orotate decarboxylase from *Neurospora crassa*. *Biochim. Biophys. Acta* 1723: 256-264
4. Neidigh, JW, Darwanto, A, Williams, AA, Wall, NR, Sowers, LC. 2009. Cloning and characterization of *Rhodotorula glutinus* thymine hydroxylase. *Chem. Res. Toxicol.* 22: 885-893
5. Holme, E, Lindstedt, G, Lindstedt, S, Tofft, M. 1970. 7-Hydroxylation of thymine in a *Neurospora* strain coupled to oxidative decarboxylation of 2-ketoglutarate. *Biochim. Biophys. Acta* 212: 50-57
6. Shaffer, PM, Arst, HN. 1984. Regulation of pyrimidine salvage in *Aspergillus nidulans*: a role for the major regulatory gene *areA* mediating nitrogen metabolite repression. *Mol. Gen. Genet.* 198: 139-145
7. Gudin, S, Quashie, NB, Candlish, D, Al-Salabi, MI, Jarvis, SM, Ranford-Cartwright, LC, de Koning, HP. 2006. *Trypanosoma brucei*: A survey of pyrimidine transport activities. *Exp. Parasitol.* 114: 118-125

8. Randolph, LB, Krug, CE, Marr, JJ. 1995. Purine and pyrimidine metabolism. In *Biochemistry and Molecular Biology of Parasites*, ed. Marr, JJ, Muller, M, pp. 89-114: Academic Press
9. van den Born, E, Omelchenko, MV, Bekkelund, A, Leihne, V, Koonin, EV, Dolja, VV, Falnes, PO. 2008. Viral AlkB proteins repair RNA damage by oxidative demethylation. *Nuc. Acids Res.* 36: 5451-5456
10. Cliffe, LJ, Kieft, R, Southern, T, Birkeland, SR, Marshall, M, Sweeney, K, Sabatini, R. 2009. JBP1 and JBP2 are two distinct thymidine hydroxylases involved in J biosynthesis in genomic DNA of African trypanosomes. *Nuc. Acids Res.* 37: 1452-1462
11. Iyer, LM, Tahiliani, M, Rao, A, Aravind, L. 2009. Prediction of novel families of enzymes involved in oxidative and other complex modifications of bases in nucleic acids. *Cell Cycle* 8: 1698-1710
12. Altschul, SF, Gish, W, Miller, W, Myers, EW, Lipman, DJ. 1990. Basic local alignment search tool. *J. Mol. Biol.* 215: 403-410
13. Laemmli, UK. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature (London)* 227: 680-685
14. Brems, S, Guilbride, DL, Gundlesdodjir-Planck, D, Busold, C, Luu, VD, Schanne, M, Hoheisel, J, Clayton, C. 2005. The transcriptomes of *Trypanosoma brucei* Lister 427 and TREU927 bloodstream and procyclic trypomastigotes. *Mol. Biochem. Parasitol.* 139: 163-172
15. Ryle, MJ, Padmakumar, R, Hausinger, RP. 1999. Stopped-flow kinetic analysis of *Escherichia coli* taurine/ α -ketoglutarate dioxygenase: interactions with α -ketoglutarate, taurine, and oxygen. *Biochemistry* 38: 15278-15286
16. Simmons, JM, Müller, TA, Hausinger, RP. 2008. Fe^{II}/ α -ketoglutarate hydroxylases involved in nucleobase, nucleoside, nucleotide, and chromatin metabolism. *Dalton Trans.* 38: 5132-5142
17. Gommers-Ampt, JH, Van Leeuwen, F, de Beer, AL, Vliegenthart, JF, Dizdaroglu, M, Kowalak, JA, Crain, PF, Borst, P. 1993. β -D-glucosyl-hydroxymethyluracil: a novel modified base present in the DNA of the parasitic protozoan *T. brucei*. *Cell* 75: 1129-1136
18. Borst, P, Sabatini, R. 2008. Base J: discovery, biosynthesis, and possible functions. *Annu. Rev. Microbiol.* 62: 235-251

19. Ulbert, S, Cross, M, Boorstein, RJ, Teebor, GW, Borst, P. 2002. Expression of the human DNA glycosylase hSMUG1 in *Trypanosoma brucei* causes DNA damage and interferes with J biosynthesis. *Nuc. Acids Res.* 30: 3919-3926
20. Gommers-Ampt, JH, Teixeira, AJ, van de Werken, G, van Dijk, WJ, Borst, P. 1993. The identification of hydroxymethyluracil in DNA of *Trypanosoma brucei*. *Nuc. Acids Res.* 21: 2039-2043
21. Cross, M, Kieft, R, Sabatini, R, Wilm, M, de Kort, M, van der Marel, GA, van Boom, JH, van Leeuwen, F, Borst, P. 1999. The modified base J is the target for a novel DNA-binding protein in kinetoplastid protozoans. *EMBO J.* 18: 6573-6581
22. DiPaolo, C, Kieft, R, Cross, M, Sabatini, R. 2005. Regulation of trypanosome DNA glycosylation by a SWI2/SNF2-like protein. *Mol. Cell* 17: 441-451
23. Yu, Z, Genest, PA, ter Riet, B, Sweeney, K, DiPaolo, C, Kieft, R, Christodoulou, E, Perrakis, A, Simmons, JM, Hausinger, RP, van Luenen, HG, Rigden, DJ, Sabatini, R, Borst, P. 2007. The protein that binds to DNA base J in trypanosomatids has features of a thymidine hydroxylase. *Nuc. Acids Res.* 35: 2107-2115
24. Vainio, S, Genest, PA, ter Riet, B, van Luenen, H, Borst, P. 2009. Evidence that J-binding protein 2 is a thymidine hydroxylase catalyzing the first step in the biosynthesis of DNA base J. *Mol. Biochem. Parasitol.* 164: 157-161

CHAPTER 2

APPENDIX

The docking studies were performed by Mike Howard.

2A.1. INTRODUCTION

Chapter 2 investigated the possible functions of two trypanosomal thymine hydroxylase-like proteins, TLP5 and TLP7 from *Trypanosoma brucei brucei*. These two peptide sequences are 85% identical to each other and about 30% identical to a fungal (*Rhodotorula glutinis*) thymine 7-hydroxylase (T7H), which functions in a pyrimidine salvage pathway (1,2). Trypanosomes are not known to possess a pyrimidine salvage pathway, but they have been shown to employ the activity of a thymidine hydroxylase in the first step of the synthesis of a modified base, β -D-glucosylhydroxymethyldeoxyuracil or base J, in their nuclear DNA (3,4). Thus, I tested whether TLP5 and TLP7 may be involved in this nucleotide modification in trypanosomes.

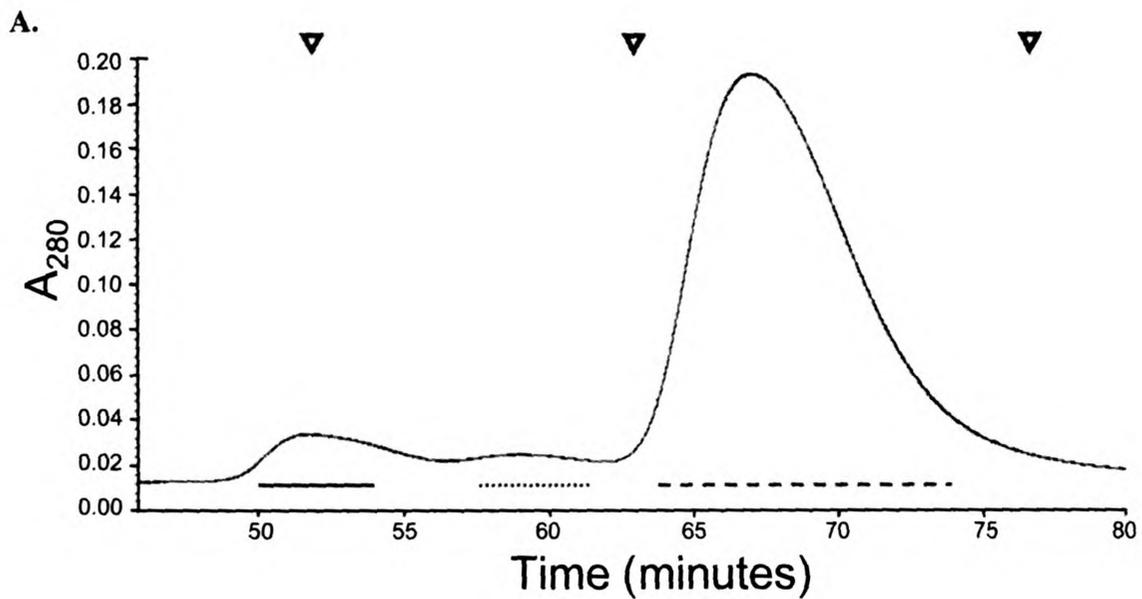
The genes were cloned from the *T. b. brucei* genome and the proteins were overproduced in *E. coli* and purified by column chromatography. The properties of these recombinant proteins were characterized by several biochemical methods and the expression of the genes was compared for the bloodstream and procyclic forms of the protozoan. Much of that work was compiled and published in *Experimental Parasitology* (see main body of Chapter 2); the experiments omitted from that publication are described here. This appendix is organized in a modular fashion. Each module explains the rationale for that method and describes the procedure, results, and any conclusions drawn from the experiments.

2A.2. OLIGOMERIC STATE DETERMINATION

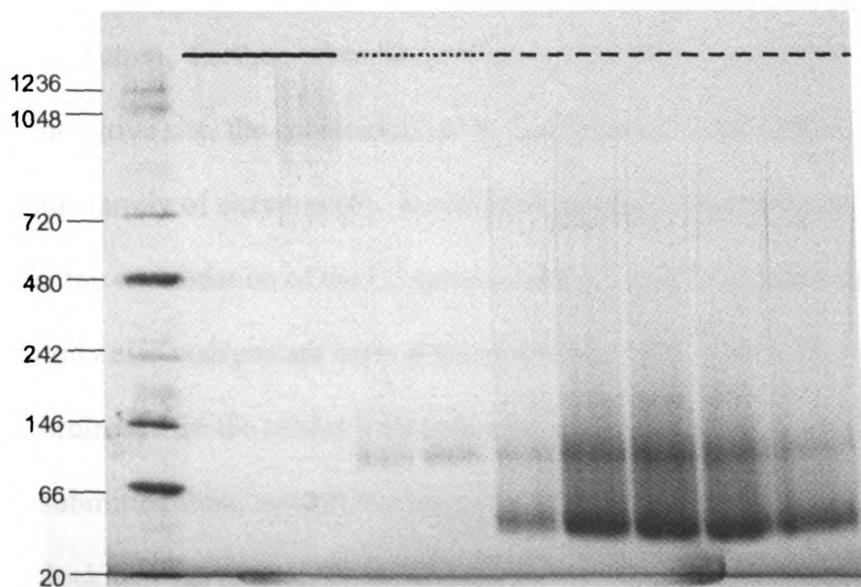
To determine their native sizes and oligomeric states, the purified and concentrated recombinant TLP5 and TLP7 proteins were chromatographed on a Superdex[®]75 size exclusion column and the elution positions were compared to those of standard proteins (BioRad). The collected fractions also were examined by polyacrylamide gel electrophoresis using a 3-12% Bis-Tris Blue native gradient gel from Invitrogen to verify the major oligomeric form of both TLP5 and TLP7. Figure 2A.1 shows that multiple forms of TLP5 were present as purified from *E. coli*. The major peak in the chromatogram (panel A) eluted at a position corresponding to 36.6 ± 4 kDa and contained a monomer as the predominant form, as seen on the native gel (panel B), although significant levels of the dimeric or trimeric species also were present during electrophoresis, consistent with an equilibrium mixture. The minor feature preceding the main peak in the chromatogram corresponded exclusively to a dimer or trimer. Finally, the first small peak in the chromatogram appeared to consist of a high molecular mass contaminant that electrophoresed as a smear on the gel. These results show that the TLP5 protein is predominantly a monomer as purified, though with some dimer or trimer present. Similar results were observed with TLP7 (data not shown).

Figure 2A.1. Native size and oligomeric state determination of TLP5. **A.**

Chromatogram of TLP5 run on a Superdex[®]75 column and compared with BioRad gel filtration standards (retention times of γ -globulin, 158 kDa; ovalbumin, 44 kDa; and myoglobin, 17 kDa, shown as inverted triangles above chromatogram). **B.** Native polyacrylamide gel showing the oligomeric state and native size of TLP5. Dashed lines correspond to those on the chromatogram to indicate the fractions run on the gel.



B.



2A.3. HOMOLOGY MODEL

There is no crystal structure for TLP5 or TLP7, nor has the fungal T7H been structurally characterized. To choose a suitable parent structure for the construction of a homology model, the protein data bank was searched using the Bioinfo Meta Server (<http://meta.bioinfo.pl>) and using the amino acid sequence of TLP5 as a query. Based on sequence identity to proteins of known structure, anthocyanidin synthase from *Arabidopsis thaliana* (PDB 1GP4), with a 33 % identity, was used as a parent structure and the sequence of TLP5 was mapped directly onto the backbone (Ca) atom coordinates from the crystal structure. Side chains were further modeled into the structure using the program SCWRL (5). Finally, the iron atom and the α KG were modeled into the structure based on the coordinates of the anthocyanidin synthase structure containing those elements. The basic structure of the catalytic core, conserved among all iron/ α KG dioxygenases, was observed in TLP5 (Figure 2A.2 panel A). The facial triad residues made up of His 200, Asp 202, and His 257 in TLP5 were correctly positioned to coordinate a metal atom. Further, when the iron atom and the α KG cosubstrate were modeled into the active site, the coordination distances observed were within the normal range seen in this family of enzymes (6). In addition, a highly conserved arginine known to participate in the coordination of the C5 carboxyl group of α KG in other members of this enzyme family also was present here, at position 266.

The coordinates for the model were provided to undergraduate student Mike Howard, who submitted them as a protein target to the program Dock Blaster (<http://blaster.docking.org>) to identify possible classes of ligands that may fit into the

modeled active site (7). A specified active site, using the residues pictured in Figure 2A.3, also was submitted in PDB format. This program used the ZINC database of commercially available small ligands (8) and docked them into the active site based on the lowest energy binding modes. Ligand candidates were then scored and ranked by binding energies. The program identified the three best ligands for the TLP5 homology model active site, which were all ring-based structures resembling nucleic acids.

The C-terminal α -helix of the TLP5 homology model was predicted to rest against the putative opening to the active site and appeared poised to act as a gate, limiting access to the oxygen-activating iron atom (Figure 2A.2 panel B). This model led to two hypotheses that were experimentally tested. First, the polyhistidine tag, added during the cloning process to facilitate protein purification, may have interfered with movement of the C-terminal helix required for activity and would thus explain the lack of activity observed to that point. Therefore, the purification tag was moved to the N-terminus and attempts to characterize the alternatively tagged protein were repeated (subsection 4). Second, one could postulate that another factor in the trypanosome, but not present in my *in vitro* system, could induce movement of the C-terminal helix and allow activity. To test this hypothesis, a truncation mutant was created that eliminated the C-terminal 13 residues making up that helix and the corresponding protein was characterized (subsection 5).

Figure 2A.2. Homology model of TLP5. **A.** Ribbon diagram colored as a spectrum from the N-terminus in blue to the C-terminus in red. Breaks in the model represent areas where there was insufficient homology to model. The iron-coordinating facial triad and α KG-stabilizing arginine are shown as sticks and colored by element (carbon-green, oxygen-red, nitrogen-blue). The α KG is also shown as sticks and colored by element (carbon-cyan, oxygen-red). The iron atom is shown as a magenta sphere. **B.** Surface representation of the same model colored by element (carbon-white, oxygen-red, nitrogen-blue, sulfur-yellow). Figures were created using PyMol.

A.



B.

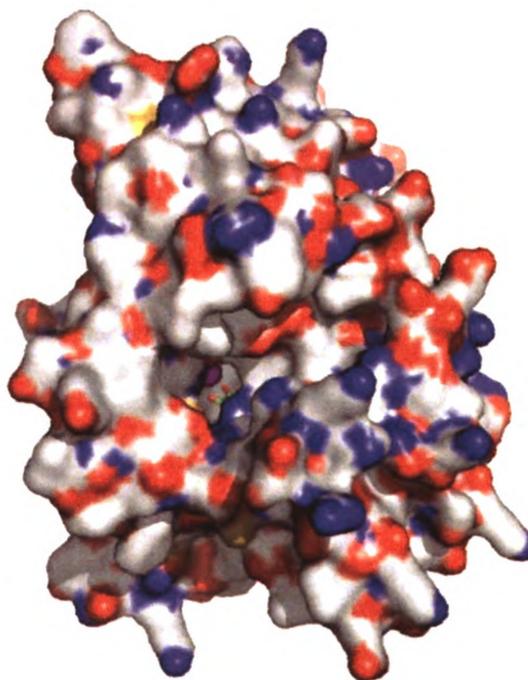
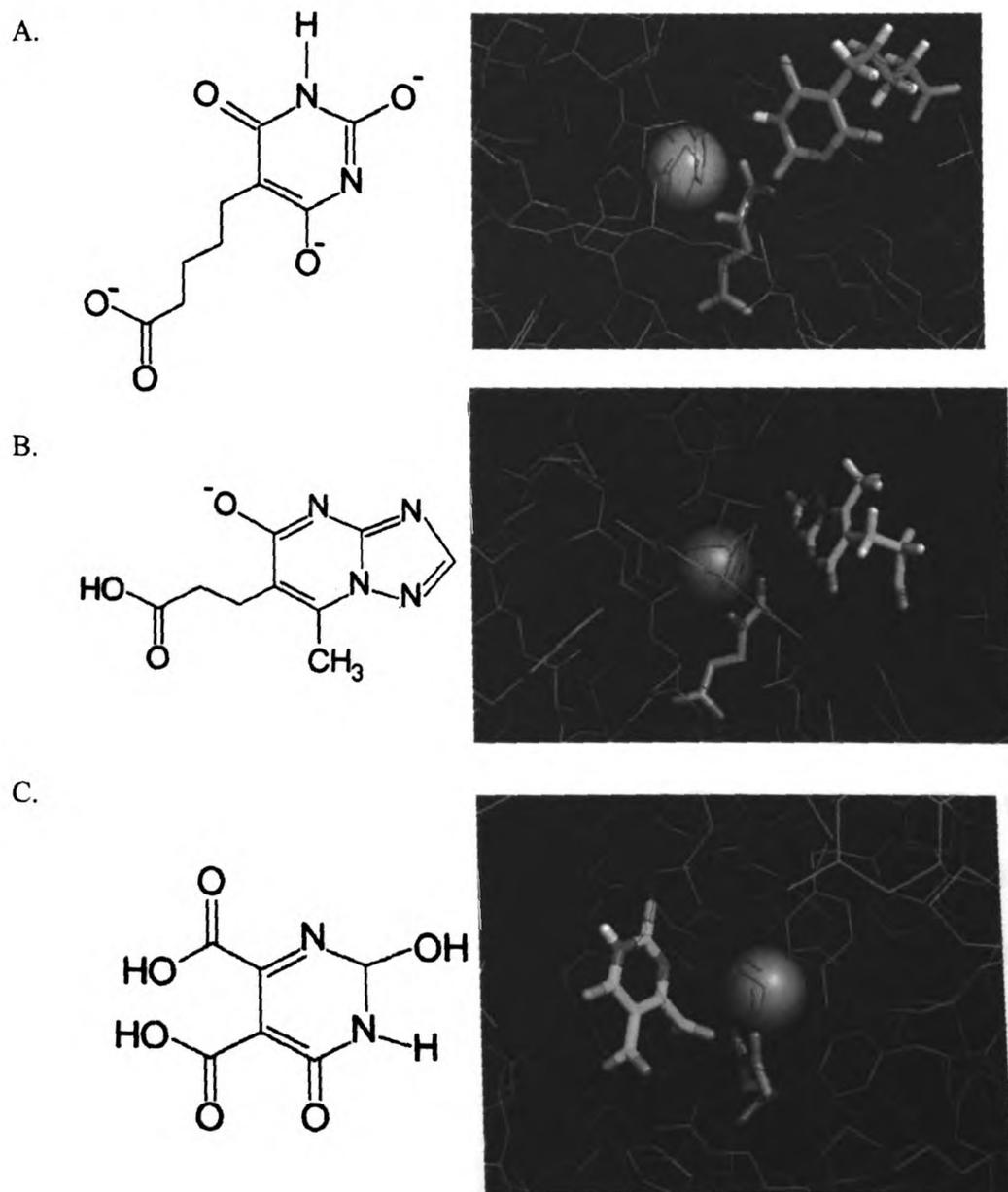


Figure 2A.3. Results of docking a subset of the ZINC small molecule database into the modeled active site of TLP5. Each panel has a chemical structure and a modeled binding depiction of a low-energy, docked ligand molecule. TLP5 is shown as magenta lines, the iron atom as an orange sphere, the α KG as either yellow or purple sticks, and the docked ligand as blue sticks. Panel C is rotated 60° from the angle of panels A and B to better show the possible molecular interactions. Binding figures were created using PyMol.



2A.4. RELOCATION OF PURIFICATION TAG

To investigate whether the C-terminal histidine tag interfered with the activity assays, the purification tag was moved to the N-terminus of the protein. The gene encoding TLP5 was cloned out of its initial expression plasmid, pET28b (Novagen), by using forward (5'-AGGATATAGC TAGCATGGCT CACGGCTCGA T-3') and reverse (5'-GTGGTGGTGC TCGAGTCACA TCTTTGTTTT G-3') primers to reintroduce the stop codon before the C-terminal purification tag and move the start site downstream of a sequence encoding the N-terminal polyhistidine purification tag. This process introduced six extra residues between the purification tag and the initial methionine of the actual protein sequence. The recombinant protein was overproduced in *E. coli* and purified as before (see main body of Chapter 2) though the protein was less stable and precipitated after even brief storage at 4 °C.

The recombinant, N-terminally-tagged TLP5 protein was concentrated to about 180 μ M and assayed for formation of the diagnostic metal-to-ligand charge-transfer chromophore in the presence of Fe^{II} and α KG as seen in the original C-terminally tagged protein preparation. The spectroscopic assay was performed as described in the main body of Chapter 2 and no chromophore was observed. This result indicated that the protein with the extra N-terminal residues did not fold properly and was thus unable to bind the iron and α KG in the correct orientation. Therefore, the N-terminally tagged protein was not used for further experiments.

2A.5. C-TERMINAL TRUNCATION

The C-terminal α -helix of TLP5 appeared to occlude access to the active site in the homology model, fueling the hypothesis that removing those residues may provide access to the active site in the absence of some required *in vivo* post translational modification or protein-induced activation. To test this hypothesis, a 13 residue truncation mutant was created to remove this helix. The gene encoding TLP5 was cloned out of genomic trypanosome DNA using the original forward primer and the new reverse primer (5'-GAGCATCCTC GAGAGTCTTC AGCAACCAAT CC-3') to replace the sequence encoding the C-terminal 13 residues with a stop codon.

Truncated protein, termed TLP5 Δ C13, was produced in *E. coli* and purified as described in the main body of Chapter 2. The shortened protein was concentrated to 65 μ M and assayed anaerobically for the characteristic metal-to-ligand charge-transfer chromophore as before. The feature was present, with an estimated extinction coefficient of 60 $M^{-1} \text{ cm}^{-1}$ indicating that the feature was incompletely formed compared with previously reported intensities (9). TLP5 Δ C13 also was assayed for enzymatic activity by measuring oxygen consumption with an oxygen electrode, but no activity was observed for any of the substrates tested including thymine, thymidine, and oligo polyT. Finally, the truncated protein was assayed for DNA binding (see subsection 6 for method). No binding was observed with any of the substrates tested including supercoiled and linearized plasmids, with or without alkylation, or HindIII lambda DNA fragments. This protein tended to precipitate during experiments; therefore, further experiments were not performed with this mutant.

2A.6. DNA BINDING

The sequence similarity of TLP5 and TLP7 to the fungal T7H led to the hypothesis that the trypanosomal proteins may be nucleotide-modifying enzymes. One possibility is that the substrate is a polynucleotide, therefore the proteins were tested for their abilities to bind to various DNA substrates via an electrophoretic mobility shift assay using agarose gels. Briefly, proteins were incubated with DNA and various supplements for 30 min at 37 °C. Glycerol was added to the samples to 5% final concentration and they were loaded onto a 0.8 % agarose gel and run in 89 mM Tris borate buffer. Additives tested in the binding assays included binding buffer (30 mM imidazole and 10 mM Tris, pH 7.9, containing 150 mM NaCl), Tris buffer (25 mM Tris, pH 8.0, containing 100 mM KCl and with or without 5 mM MgCl₂, 0.5 mM DTT, or 0.2 mM EDTA), or Hepes buffer (50 mM Hepes, pH 8.0, containing 40 μM CoCl₂, 100 μM ascorbate, and 200 μM αKG). Other assay components included in some of the assays included up to 200 μM αKG, 50 μM Fe^{II}, and 100 μM ascorbate. DNA substrates included supercoiled plasmid, linearized plasmid, methylated plasmid, and HindIII λ DNA fragments. No shift in the DNA bands was observed that was greater than the variation in band positions present in lanes without protein. This result, taken together with the lack of detectable enzymatic activity challenged the hypothesis that TLP5 and TLP7 were polynucleotide-modifying enzymes. Therefore, other methods were employed to look for possible substrates and functions of these proteins.

2A.7. *IN VIVO* EPITOPE TAGGING OF THE *TLP5* GENE

Data acquired by quantitative reverse transcriptase PCR amplification of *TLP5* and *TLP7* indicated that the steady state mRNA levels were higher in the bloodstream form trypanosomes than the procyclic forms. While interesting, the picture remained incomplete as most of the genetic regulation in trypanosomes takes place at the level of protein expression and turnover (10). Therefore, a construct was designed to add sequence coding for a nine residue hemagglutinin (HA) epitope tag to the chromosomal copy of *TLP5*. This would allow the analysis of *in vivo* protein levels under various stress conditions as well as determination of the subcellular localization for the protein, and it would provide a method to observe changes in the protein production levels in concert with other procedures such as RNAi.

The initial HA tagging construct was acquired from Professor Donna Koslowsky. It consisted of the sequence coding for the HA tag, an α/β tubulin spacer region, and a blasticidin resistance gene. About 100 nucleotides corresponding to the 3' end of the *TLP5* gene were added to the 5' end of the construct by PCR amplification using the primer (5'- CCATTCGCCA AGCAATCCCC CCAAATATCC ACCAGTCCGT GCTGTGGATT GGTTGCTGAA GCGTTTCGCG GAAACATATG CCCATCGCAA AACAAAGATG TACCCATACG ACGTCCCAGA CTACGCTTAA- 3') where the stop codon is underlined. In addition, about 100 nucleotides corresponding to the beginning of the 3' untranslated region was added to the 3' end of the construct by PCR amplification using the primer (5'- GAAAAGAAAA AAAAAGAGTT TGAAGACACA TTCATTCATT CATTGCCACC GCATCCGTAC CGACGGGGGC CTAACGAAAA GATTACGTTC

AGTTATTGAT TTAGCCCTCC CACACATAAC CAGAG -3') where the reverse complement of the stop codon for the blasticidin resistance gene is underlined. This first round of PCR amplification created the unique tagging construct needed for my tagging procedure. Next, this unique construct was amplified by a second round of PCR using the forward primer (5'- CCATTCGCCA AGCAATCCCC CAAA -3') and the reverse primer (5'- GAAAAGAAAA AAAAAGAGTT TGAAGACACA TTCATTCATT -3') (Figure 2A.4). The PCR products were separated from unused reagents by using the Qiagen PCR clean-up kit and used directly for trypanosome transfections.

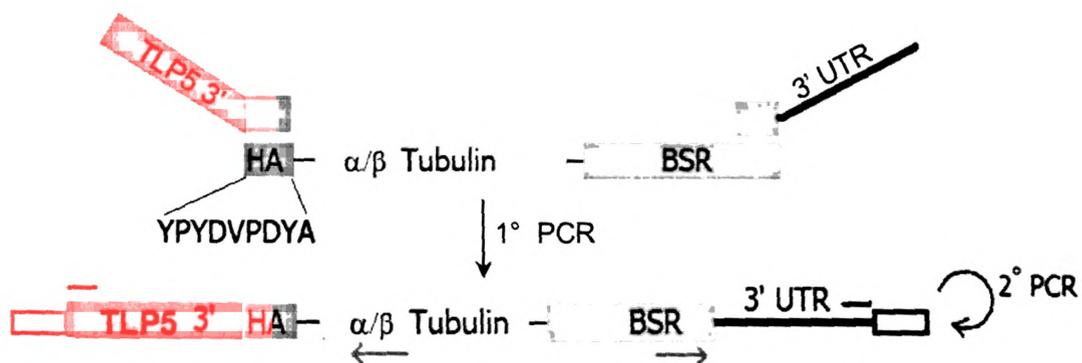
In replicate trials, 2.5 or 5 μg of the DNA-tagging construct were added to 10^8 procyclic trypanosome cells in 0.5 mL phosphate-buffered sucrose in a BioRad electroporation cuvette. Cells were treated in a BioRad electroporator at 1.5 kV, 25 μF , and infinite resistance for two pulses spaced by 10 sec. Cells were transferred immediately to a small culture flask containing 4.5 mL of SDM-79 medium with 10 % fetal bovine serum and 7.5 $\mu\text{g}/\text{mL}$ hemin (as an iron source) along with penicillin and streptomycin to select against bacteria. Cultures were allowed to grow at 27 $^{\circ}\text{C}$ overnight with gentle shaking before adding the eukaryotic selection drugs: 15 $\mu\text{g}/\text{mL}$ neomycin (to maintain the tet repressor), 100 $\mu\text{g}/\text{mL}$ hygromycin (to maintain the T7 polymerase), and 10 $\mu\text{g}/\text{mL}$ blasticidin (to maintain our tagging construct). Cells were grown for an additional two days and then serially diluted in the same selective media to create a more clonal population. One of the dilution cultures was grown for approximately a week before some of the cells were harvested and genomic DNA extracted to test the incorporation of the tagging construct.

The entire *TLP5* gene segment of chromosome 5 was amplified by PCR using the forward primer (5'- GCCATTTTT CTGTGAGCCT AATCC -3') and the reverse primer (5'- TCAGG ACAGATACGA GCAAAGAAG-3'). With the incorporated tag, the lengths of the products would be a mixture of approximately 1 kb for the tagged allele and 300 bases for the untagged allele, since trypanosomes are diploid. In addition, the segment was amplified by using primers internal to the tagging sequence that would further test its placement when paired with the primers described above; these hybridize to regions just upstream and downstream of the sequences included in the original unique-tagging construct. The internal primers corresponded to a forward segment near the end of the blasticidin resistance gene (5'-GGGATTCGTGAATTGCTGCCCTCT-3') and a reverse segment (5'-TGTACCACGCTGCAACAGTGTGAAAATTCGA-3') near the beginning of the α/β tubulin spacer (Figure 2A.4).

One culture out of six separate tagging experiments, each with six transfections, appeared to have the correct tag incorporation in at least one allele as seen by the PCR verification. Therefore, a whole cell lysate was made from a sample of that culture and was Western blotted using the anti-HA antibody conjugated to alkaline phosphatase as a probe. Briefly, 10^7 cells were harvested by centrifugation and resuspended in 1 mL cell wash buffer (20 mM Tris, pH 7.5, containing 100 mM NaCl and 3 mM $MgCl_2$) and placed in a bath sonicator with ice for about 20 min. Debris was removed by centrifugation at 16,000 g for about 10 min. The supernatant was electrophoresed on a denaturing 12% polyacrylamide gel, proteins were transferred to an ImmobilonP PVDF membrane (Millipore), and the membrane was probed with rabbit anti-HA antibody from Sigma. Unfortunately, no protein band of the appropriate size was detected. Several high

molecular weight bands were detected that are commonly seen in blots of trypanosome extracts probed with the anti-HA antibody (Donna Koslowsky personal communication).

Figure 2A.4. Hemagglutinin epitope tagging of *TLP5*. This figure is a construction schematic where the nine residue hemagglutinin (HA) epitope is shown in green, the α/β tubulin spacer in yellow, and the blasticidin resistance gene in blue. The 3' end of the *TLP5* gene is shown in red and the 3' untranslated region in brown as parts of the primers used to make the unique tagging construct by PCR. This construct was amplified by a 2^o PCR reaction using primers whose positions are indicated by the red and brown lines above the scheme. Finally, primers whose positions are indicated by the red and brown open boxes and the grey arrows were used to test the incorporation of the tag into chromosome 5 of the trypanosome.



2A.8. PULL-DOWN ASSAYS WITH TLP5 AND TRYPANOSOME CELL EXTRACT

All attempts to detect activity with various nucleic acid substrates failed, therefore pull down assays with trypanosome cell extracts were used in an effort to find potential protein interaction partners of TLP5. *E. coli* cells over-expressing TLP5 were lysed, sonicated, and ultracentrifuged to generate soluble cell extracts. This sample was added to Ni-NTA cellulose beads and the histidine tagged TLP5 allowed to bind to the resin. The beads were washed with binding buffer twice and then trypanosome soluble cell extracts (prepared as before) were added and incubated at room temperature for 20 min to allow binding. Again, the resin was washed with binding buffer and eluted in elution buffer (150 mM imidazole and 10 mM Tris, pH 7.9, containing 150 mM NaCl). Extract, wash, and elution samples were electrophoresed on a denaturing 12% polyacrylamide gel and silver stained to detect proteins of low abundance.

The gel lanes were compared as described above for *E. coli* extracts, *T. b. brucei* extracts, and mixtures of the extracts. With three separate experiments, including analyses using free resin in an eppendorf centrifuge tube and resin packed into a small column, no new bands were identified in the combined extract runs that were not also present in the control runs. Thus, the pull-down experiments failed to identify potential interaction partners of TLP5.

2A.9. CONCLUSIONS

The thymine hydroxylase-like proteins TLP5 and TLP7 from *T. b. brucei* have been characterized and determined to be members of the non-heme iron and α KG dependent dioxygenase superfamily of enzymes. Extensive efforts were undertaken to identify the activity of these proteins, though none of the tested hypotheses were supported. Under the conditions used, no DNA binding was observed and no activity was demonstrated using various substrates, even when examining modified proteins to eliminate possible active site access restrictions. Pull-down assays failed to identify protein interaction partners and docking studies did not yield obvious ligand choices to pursue. While the true purpose of these proteins in the cell remains enigmatic, they probably do not participate in polynucleotide modification but are very likely to catalyze some other type of hydroxylation reaction in the cells.

REFERENCES

1. Smiley, JA, Kundracik, M, Landfried, DA, Barnes, VR, Sr., Axhemi, AA. 2005. Genes of the thymidine salvage pathway: thymine-7-hydroxylase from a *Rhodotorula glutinis* cDNA library and iso-orotate decarboxylase from *Neurospora crassa*. *Biochim Biophys Acta* 1723: 256-264
2. Neidigh, JW, Darwanto, A, Williams, AA, Wall, NR, Sowers, LC. 2009. Cloning and characterization of *Rhodotorula glutinis* thymine hydroxylase. *Chem Res Toxicol* 22: 885-893
3. Gommers-Ampt, JH, Van Leeuwen, F, de Beer, AL, Vliegthart, JF, Dizdaroglu, M, Kowalak, JA, Crain, PF, Borst, P. 1993. beta-D-glucosyl-hydroxymethyluracil: a novel modified base present in the DNA of the parasitic protozoan *T. brucei*. *Cell* 75: 1129-1136
4. Gommers-Ampt, JH, Teixeira, AJ, van de Werken, G, van Dijk, WJ, Borst, P. 1993. The identification of hydroxymethyluracil in DNA of *Trypanosoma brucei*. *Nucleic Acids Res* 21: 2039-2043
5. Canutescu, AA, Shelenkov, AA, Dunbrack, RL, Jr. 2003. A graph-theory algorithm for rapid protein side-chain prediction. *Protein Sci* 12: 2001-2014
6. Hausinger, RP. 2004. Fe(II)/ α -ketoglutarate-dependent hydroxylases and related enzymes. *Crit. Rev. Biochem. Mol. Biol.* 39: 21-68
7. Irwin, JJ, Shoichet, BK, Mysinger, MM, Huang, N, Colizzi, F, Wassam, P, Cao, Y. 2009. Automated docking screens: a feasibility study. *J Med Chem* 52: 5712-5720
8. Irwin, JJ, Shoichet, BK. 2005. ZINC--a free database of commercially available compounds for virtual screening. *J Chem Inf Model* 45: 177-182
9. Ryle, MJ, Padmakumar, R, Hausinger, RP. 1999. Stopped-flow kinetic analysis of *Escherichia coli* taurine/ α -ketoglutarate dioxygenase: interactions with α -ketoglutarate, taurine, and oxygen. *Biochemistry* 38: 15278-15286
10. Clayton, C, Shapira, M. 2007. Post-transcriptional regulation of gene expression in trypanosomes and leishmanias. *Mol Biochem Parasitol* 156: 93-101

CHAPTER 3

SEARCH FOR HYDROXYLASE ACTIVITY IN THE TRYPANOSOMAL PROTEIN, JBP1

Portions of these studies were published along with experiments conducted by others in:

Yu, Z, Genest, PA, ter Riet, B, Sweeney, K, DiPaolo, C, Kieft, R, Christodoulou, E, Perrakis, A, Simmons, JM, Hausinger, RP, van Luenen, HG, Rigden, DJ, SabatiniR, Borst, P. 2007. The protein that binds to DNA base J in trypanosomatids has features of a thymidine hydroxylase. *Nuc. Acids Res.* **35**: 2107-2115

The introduction for this chapter mentions some of the other studies, citing this reference, as well as prior background and more recent results by others.

3.1 INTRODUCTION

Trypanosomes are eukaryotic kinetoplastid flagellates and the causative agents of several neurological diseases affecting mammals. Chagas Disease, caused by *Trypanosoma cruzi* and transmitted by reduvid bugs, leishmaniasis, associated with various strains of *Leishmania* and transmitted by sandflies, and African Sleeping Sickness, caused by *Trypanosoma brucei* and carried by the tsetse fly are examples in humans (1,2). About one percent of the thymine in the telomeric repeats of nuclear DNA in the mammalian bloodstream form (BF) of *T. brucei* is replaced with β -D-glucosyl-hydroxymethyldeoxyuracil (base J) (3,4). Even smaller amounts of the modification are found in subtelomeric regions and very little in chromosome internal locations. This modified base has several possible functions. First, it is proposed to induce chromatin compaction and thereby stabilize the chromosome from rearrangements within these repetitive sequences (5,6). Second, base J has a proposed role in the process of antigenic variation characteristic of *T. brucei* and other related parasites. Base J is found throughout the silent variable surface glycoprotein (VSG) expression sites and the flanking repetitive sequences, but no J is present in the one active expression site, suggesting a specific role for J in gene silencing (7). Finally, base J has been identified in the chromosomal internal polymerase II polycistronic transcription unit strand switch regions of several kinetoplastids, including *T. brucei* and *L. major*, as well as the spacer regions between individual open reading frames, suggesting a role for base J in the wider family of kinetoplastids that do not all undergo antigenic variation (8).

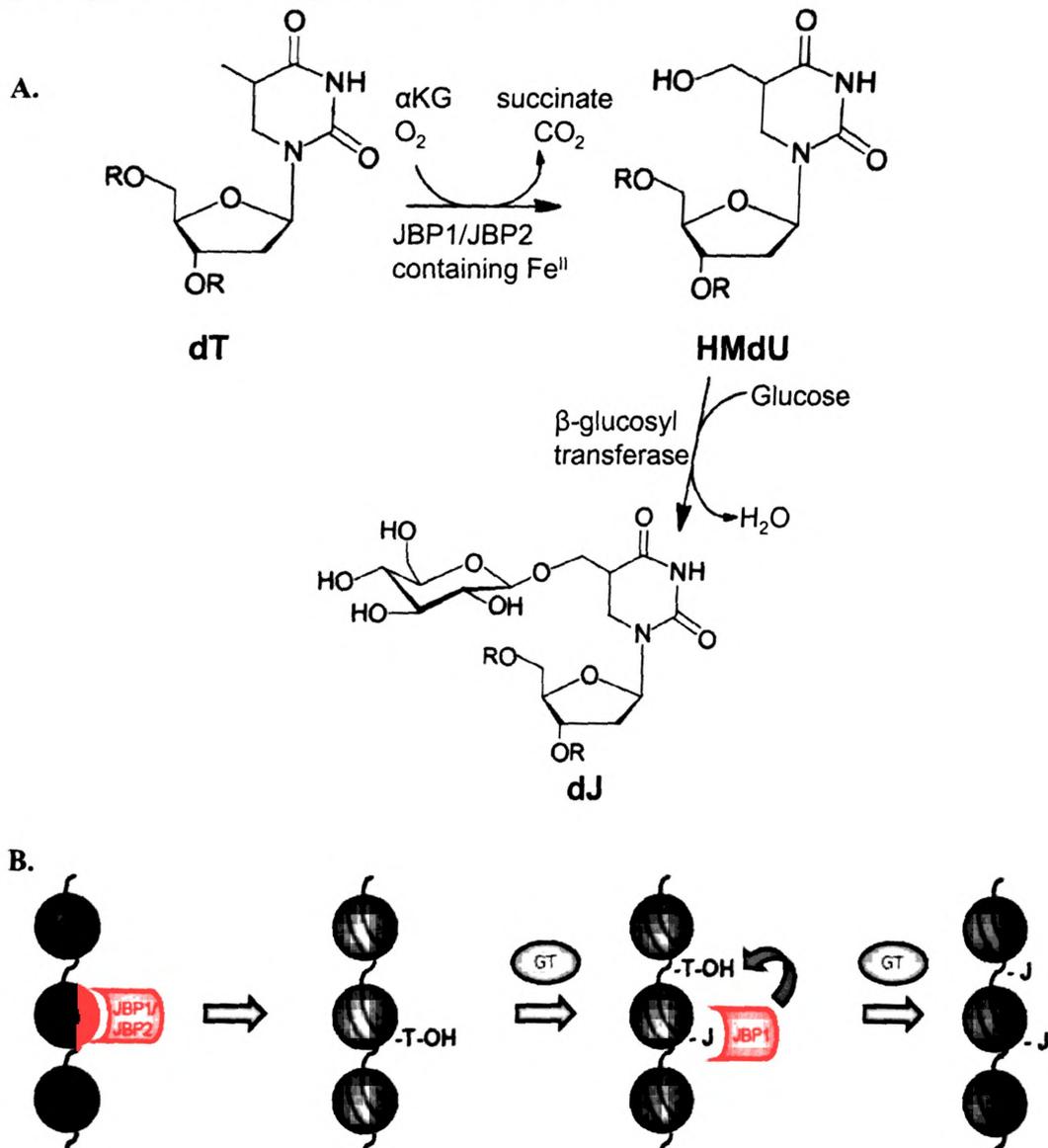
Base J is synthesized in two steps. First, the methyl group on the C5 position of a thymidine already incorporated into DNA must be hydroxylated to create the

intermediate hydroxymethyldeoxyuridine (HOMedU) (Figure 3.1 panel A), as evidenced by the presence of low levels of HOMedU in BF trypanosome DNA (9). Next, the hydroxymethyl moiety must be glucosylated. This reaction is not specific to the sequence or life cycle-stage as evidenced by the random placement of base J in the bloodstream and procyclic forms of cells grown on HOMedU (5).

Recently, a protein was identified in *T. brucei* that specifically binds to base J in DNA (10). This J binding protein 1 (JBP1) is ~90 kDa protein whose N terminal domain shows some sequence similarity to other non-heme iron and α -ketoglutarate (α KG) dependent hydroxylases by manual alignment (11). Secondary structure prediction further places this protein in the double-stranded beta helix or “jelly-roll” core fold family of proteins (11). This homology and the obvious functional similarity to the methylated DNA hydroxylase AlkB (a DNA repair enzyme) and the fungal thymine-7-hydroxylase, which is involved in the pyrimidine salvage pathway and acts only on the free base, fueled the hypothesis that JBP1 may be the thymine hydroxylase in trypanosomes.

JBP2, identified *in silico* based on its sequence identity to JBP1, does not bind to J-containing DNA, but does possess a SWI-2/SNF-2 chromatin remodeling domain at its C-terminal end and is capable of inducing *de novo* base J synthesis when expressed in procyclic *T. brucei* cells (12,13). These cells normally do not express JBP1 or JBP2 and lack base J (3). JBP1 is suggested to maintain J levels by stimulating additional J synthesis in nearby sequences while cells containing modified DNA, but lacking JBP1, lose base J by dilution during replication (14). Further work to establish the roles of these proteins has shown that a JBP1 knockout in *Crithidia fasciculata* (a non-pathogenic

Figure 3.1. Synthesis of base J. **A.** Scheme showing the two step molecular synthesis of base J. The R group represents the continuing strands of DNA. **B.** Cartoon depicting the proposed chromatin specific introduction and maintenance of base J. JBP2 is responsible for the initial introduction of telomeric and subtelomeric base J, while JBP1 is responsible for the initial introduction of chromosome internal base J (8). JBP1 then binds to these regions and stimulates amplification of the modified base (15). GT-glucosyl transferase. Figure modified from (11).



strain) reduces J levels to 5% of the wild-type cells (14), a JBP2 knockout reduces base J levels to 20% of the wild-type cells in studies involving *T. brucei* and *L. tarentolae* (12,16), and a double knockout creates a J-null cell line in *T. brucei* (8,15). Surprisingly, both JBP1 and JBP2 proteins are dispensable in *T. brucei* with no phenotypic or growth effects (8,15), while JBP1 is essential in *L. tarentolae* (17) and JBP2 is not (16). Finally, it has been shown that mutation of the key conserved residues making up the “facial triad” of the active site or the conserved arginine believed to stabilize the α KG cosubstrate in these proteins abolishes their ability to stimulate base J synthesis (11,15,16).

Understanding the enzymes involved in J biosynthesis may lead to the development of drugs to treat trypanosomiasis, as this modified base is found only in the kinetoplastid flagellates. Yet, no direct *in vitro* hydroxylase activity has been detected for either JBP1 or JBP2. Therefore, this study sought to demonstrate enzymatic or uncoupled turnover of JBP1. To test JBP1 for α KG dependent hydroxylase activity, I used recombinant protein with *in vitro* assays to look for product formation and substrate consumption. Other members of this enzyme family have been shown to decarboxylate α KG and consume oxygen in the absence of their primary substrate, termed uncoupled turnover. These studies primarily focused on the uncoupled reactions of JBP1 due to the lack of sufficient amounts of primary substrate, genomic trypanosome DNA containing base J. Some putative alternate substrates were tested in attempts to measure the coupled turnover or stimulate uncoupled turnover.

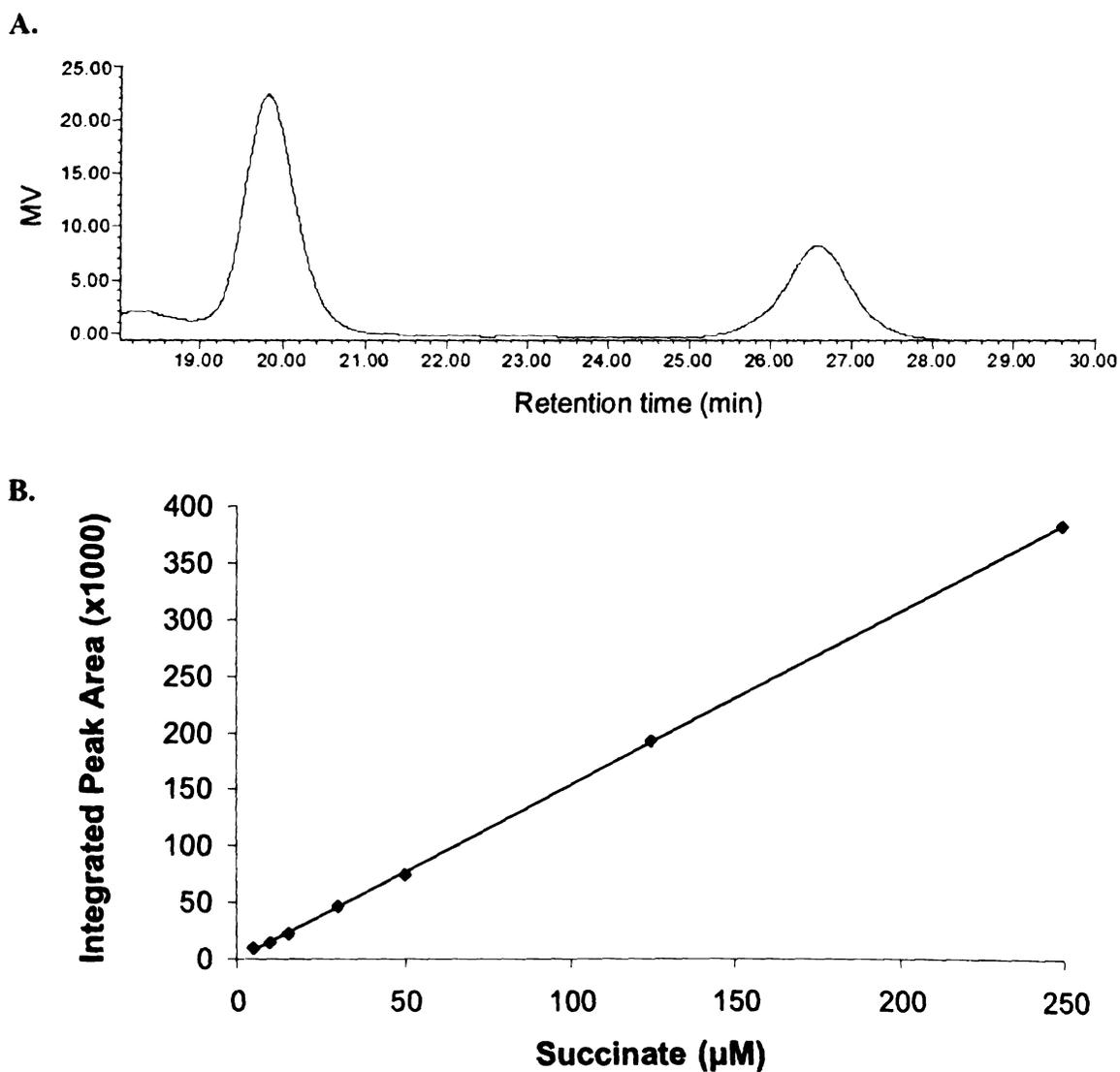
3.2 MATERIALS AND METHODS

3.2.1 Determination of succinate production as a measure of uncoupled enzymatic turnover.

Purified recombinant JBP1 was supplied by Henri van Luenen of Piet Borst's lab at The Netherlands Cancer Institute in Amsterdam. JBP1 was from *C. fasciculata* (not pathogenic to mammals) or *Leishmania tarentolae* in wild type or mutated forms. One mL reactions in 13 x 100 glass tubes, to provide for efficient aeration, were incubated at either 30 or 37 °C in either 25 mM Tris or 50 mM Hepes buffer, with the pH ranging from 7.0 to 8.5. Conditions tested included assays containing ferrous iron (0.05-1 mM), ascorbate (0.05-5 mM), α KG (0.5-20 mM), NaCl (20 mM), MgCl₂ (0.5-10 mM), KCl (10-100 mM), EDTA (0 or 20 μ M), glycerol (0 or 5%), or Triton X-100 (0 or 0.02%), and as potential substrates, thymidine (20 mM), DNA oligo containing base J (20 μ M), and genomic trypanosome DNA (40 μ M).

Aliquots of 300 μ L were taken at various time points and quenched with 5 μ L of 6 M sulfuric acid. Samples were centrifuged at 16,000 g for 5 min to remove any protein. 250 μ L was loaded onto a 0.45 μ m cutoff spin filter and centrifuged. 200 μ L of the flow through fraction was loaded onto an Aminex HPX-87H ion exchange organic acids column (BioRad) (300 mm x 7.8 mm) and analyzed by high performance liquid chromatography (HPLC). Organic acids were eluted with 0.013 M H₂SO₄ as the mobile phase and detected by using a Waters refractive index detector set at 35 °C and a sensitivity level of 16. The standards of α KG and succinate eluted at 19.8 and 26.6 min,

Figure 3.2. α KG and succinate detection by HPLC. **A.** A representative chromatogram showing the retention times for α KG (19.8 min) and succinate (26.6 min). **B.** Standard curve showing the detection of succinate down to a lower limit of 5 μ M in the injected volume. The R^2 value was 0.9999 and the fitted line was used to determine the amount of succinate produced in an assay based on the integrated area of the detected peak.



respectively, and succinate was detectable to a lower limit of 1 nmol or 5 μ M in the injected volume (Figure 3.2).

3.2.2 Oxygen consumption as a measure of uncoupled enzymatic turnover.

Assays were performed by using a micro oxygen electrode in an Instech chamber coupled to a circulating water bath to maintain a temperature of 33° C and a Diamond Electro-Tech chemical microsensor. Percent oxygen saturation was continuously monitored throughout the assay. The upper and lower saturation limits were set by flushing the system with buffer (25 mM imidazole, pH 7.0) that was bubbled with air or degassed buffer bubbled with nitrogen, respectively.

JBP1, as received from the Netherlands, was in 25 mM Hepes, pH 7.0, with 150 mM NaCl and 1 mM EDTA. A portion of the protein was dialyzed against 25 mM imidazole, pH 7.0, and 150 mM NaCl to remove the EDTA. JBP1 with or without EDTA was assayed at 1 μ M using variations of the standard conditions: 50 μ M ferrous iron, 1 mM α KG, 400 μ M ascorbate, 6.5 mM NaCl, and 1 mM of either thymine, thymidine, or oligo polyT. Assays were carried out in 25 mM imidazole, pH 7.0, in the micro-oxygen electrode with an assay volume of 600 μ L.

3.2.3 Spectroscopic detection of a Metal-to-Ligand Charge-Transfer (MLCT) Feature.

UV/visible spectra were obtained by using an HP 8453 spectrophotometer (Hewlett Packard) equipped with a circulating water bath. Stock solutions of 10 mM Fe^{II} and 5 mM α KG were made by adding the dry reagents to vials, subjecting them to repeated cycles of vacuum and argon, and adding anaerobic water by using a syringe.

The JBP1 protein was provided in a high salt buffer with a reducing agent for shipping, therefore to prevent possible interference in these experiments it was first exchanged (at 4 °C by repeating concentration/dilution by using an Amicon double-membrane centrifugal concentrator with a 10 kDa cutoff) into 25 mM Tris, pH 8.0, containing 50 mM KCl. 50 μM protein was added to a 1-cm path length, 300 μL, black-walled cuvette and made anaerobic by the same method, followed by scanning at 9 °C. Four equivalents of αKG were added followed by a titration of one equivalent of Fe^{II} into the cuvette in 0.2 equivalent increments with scans after each addition. Data were analyzed by using the software IGOR6. First, data were corrected for baseline shifts arising from cuvette repositioning to have a uniform absorbance at 850 nm. Then, data points over a 5 nm range of wavelengths on each side of every point were averaged to smooth the data and the wavelength maxima of the observed features were determined by taking the second derivative.

3.3 RESULTS AND DISCUSSION

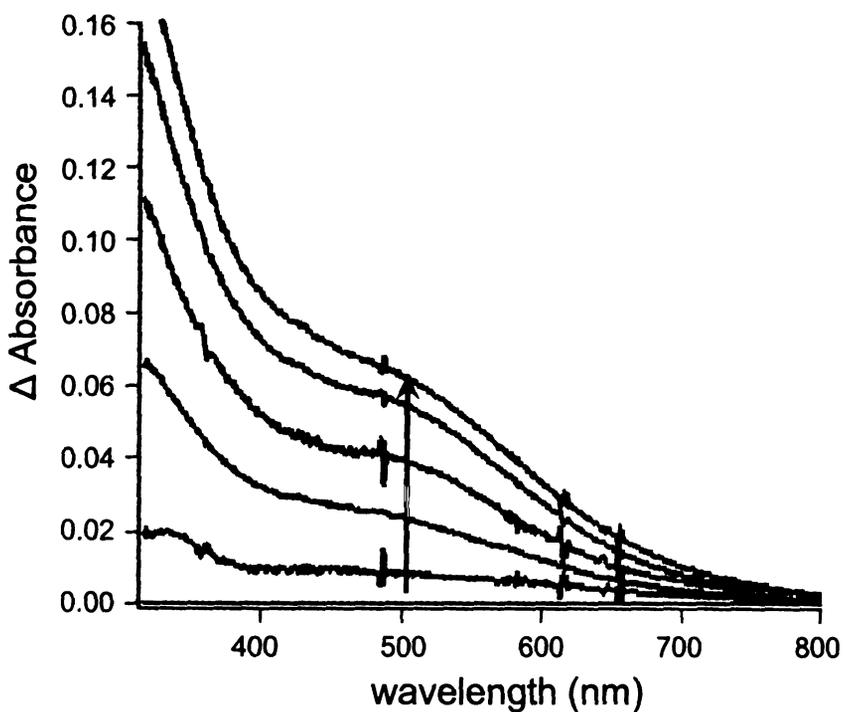
The trypanosome protein JBP1 is known to bind to J-containing DNA and has been implicated in the process of antigenic variation and chromatin maintenance. It was hypothesized that the C-terminal domain functions in J base recognition and binding while the N-terminal domain serves as a non-heme iron and α KG-dependent hydroxylase responsible for the initial step in the synthesis of newly modified bases in adjacent DNA regions. If this is the case, then oxygen and α KG would be consumed in the reaction and succinate would be produced.

Assays for oxygen consumption and α KG decomposition were performed by using an oxygen electrode and HPLC, respectively, with purified recombinant JBP1 that was known to bind DNA oligos containing base J (as evidenced by gel band shifts). Methods included assays for both coupled and uncoupled turnover. The assay was verified using taurine hydroxylase. Various nucleic acid substrates were tested, though the amounts available were severely limiting. None of the assays analyzed by oxygen electrode exhibited oxygen consumption greater than the background levels measured in the absence of protein. The background oxygen consumption is thought to arise from the oxidation of ascorbate in the presence of iron (18). In a similar manner, none of the assays analyzed by HPLC showed appreciable production of succinate over a basal level that was measurable when no protein was present. This background amount of succinate is thought to arise from the spontaneous breakdown of α KG in the presence of iron, oxygen, and ascorbate.

To confirm proper folding of JBP1, the purified recombinant protein was examined for its ability to form a characteristic MLCT chromophore seen in other members of the Fe^{II}/αKG-dependent dioxygenases and associated with binding of the iron cofactor and the αKG cosubstrate (19). The protein was monitored by UV-visible spectroscopy under anaerobic conditions, using low temperature to maintain stability, while titrating in αKG and Fe^{II}. As shown in Figure 3.3, anaerobic JBP1 generated the diagnostic MLCT transition at 504 nm when both metal and cofactor were present. Due to the light scattering underneath the feature, an extinction coefficient was not estimated from this data.

While JBP1 and JBP2 have been shown by mutagenesis studies to be responsible for the specific introduction of base J into kinetoplastid DNA, *in vitro* activity has not been directly detected for either protein. JBP1 purified from recombinant cells was shown to be properly folded and capable of binding iron and αKG in a manner that formed a characteristic MLCT spectrum. This result verifies the protein's membership in the non-heme iron and αKG dependent dioxygenase superfamily of enzymes. Since the activity assays performed here were unsuccessful, it is possible that the substrate for the JBP1 enzyme may be a specific form of chromatin that is recognized by its structure rather than sequence. Therefore, even the genomic trypanosome DNA containing base J, would not elicit enzymatic turnover under the conditions we tested.

Figure 3.3. Spectroscopic evidence for binding of Fe^{II} and α KG by JBP1. The anaerobic UV/visible spectrum of JBP1 (50 μ M in 25 mM Tris, pH 8.0 containing 50 mM KCl) was examined for the sample as prepared, after adding 200 μ M α KG, and while titrating in Fe^{II} with stirring at 9 °C. The (Fe^{II}- α KG-protein minus α KG-protein) difference spectra shown correspond to the addition of 0 (baseline), 0.2, 0.4, 0.6, 0.8, and 1 equivalent of metal ions.



REFERENCES

1. Barrett, MP, Burchmore, RJ, Stich, A, Lazzari, JO, Frasch, AC, Cazzulo, JJ, Krishna, S. 2003. The trypanosomiasis. *Lancet* 362: 1469-1480
2. Herwaldt, BL. 1999. Leishmaniasis. *Lancet* 354: 1191-1199
3. Gommers-Ampt, JH, Van Leeuwen, F, de Beer, AL, Vliegenthart, JF, Dizdaroglu, M, Kowalak, JA, Crain, PF, Borst, P. 1993. beta-D-glucosyl-hydroxymethyluracil: a novel modified base present in the DNA of the parasitic protozoan *T. brucei*. *Cell* 75: 1129-1136
4. van Leeuwen, F, Dirks-Mulder, A, Dirks, RW, Borst, P, Gibson, W. 1998. The modified DNA base beta-D-glucosyl-hydroxymethyluracil is not found in the tsetse fly stages of *Trypanosoma brucei*. *Mol Biochem Parasitol* 94: 127-130
5. van Leeuwen, F, Kieft, R, Cross, M, Borst, P. 1998. Biosynthesis and function of the modified DNA base beta-D-glucosyl-hydroxymethyluracil in *Trypanosoma brucei*. *Mol Cell Biol* 18: 5643-5651
6. van Leeuwen, F, Kieft, R, Cross, M, Borst, P. 2000. Tandemly repeated DNA is a target for the partial replacement of thymine by beta-D-glucosyl-hydroxymethyluracil in *Trypanosoma brucei*. *Mol Biochem Parasitol* 109: 133-145
7. van Leeuwen, F, Wijsman, ER, Kieft, R, van der Marel, GA, van Boom, JH, Borst, P. 1997. Localization of the modified base J in telomeric VSG gene expression sites of *Trypanosoma brucei*. *Genes Dev* 11: 3232-3241
8. Cliffe, LJ, Siegel, TN, Marshall, M, Cross, GA, Sabatini, R. 2010. Two thymidine hydroxylases differentially regulate the formation of glucosylated DNA at regions flanking polymerase II polycistronic transcription units throughout the genome of *Trypanosoma brucei*. *Nucleic Acids Res* 38: 3923-3935
9. Gommers-Ampt, JH, Teixeira, AJ, van de Werken, G, van Dijk, WJ, Borst, P. 1993. The identification of hydroxymethyluracil in DNA of *Trypanosoma brucei*. *Nucleic Acids Res* 21: 2039-2043
10. Cross, M, Kieft, R, Sabatini, R, Wilm, M, de Kort, M, van der Marel, GA, van Boom, JH, van Leeuwen, F, Borst, P. 1999. The modified base J is the target for a novel DNA-binding protein in kinetoplastid protozoans. *EMBO J* 18: 6573-6581
11. Yu, Z, Genest, PA, ter Riet, B, Sweeney, K, DiPaolo, C, Kieft, R, Christodoulou, E, Perrakis, A, Simmons, JM, Hausinger, RP, van Luenen, HG, Rigden, DJ,

- Sabatini, R, Borst, P. 2007. The protein that binds to DNA base J in trypanosomatids has features of a thymidine hydroxylase. *Nucleic Acids Res* 35: 2107-2115
12. DiPaolo, C, Kieft, R, Cross, M, Sabatini, R. 2005. Regulation of trypanosome DNA glycosylation by a SWI2/SNF2-like protein. *Mol Cell* 17: 441-451
 13. Kieft, R, Brand, V, Ekanayake, DK, Sweeney, K, DiPaolo, C, Reznikoff, WS, Sabatini, R. 2007. JBP2, a SWI2/SNF2-like protein, regulates de novo telomeric DNA glycosylation in bloodstream form *Trypanosoma brucei*. *Mol Biochem Parasitol* 156: 24-31
 14. Cross, M, Kieft, R, Sabatini, R, Dirks-Mulder, A, Chaves, I, Borst, P. 2002. J-binding protein increases the level and retention of the unusual base J in trypanosome DNA. *Mol Microbiol* 46: 37-47
 15. Cliffe, LJ, Kieft, R, Southern, T, Birkeland, SR, Marshall, M, Sweeney, K, Sabatini, R. 2009. JBP1 and JBP2 are two distinct thymidine hydroxylases involved in J biosynthesis in genomic DNA of African trypanosomes. *Nucleic Acids Res* 37: 1452-1462
 16. Vainio, S, Genest, PA, ter Riet, B, van Luenen, H, Borst, P. 2009. Evidence that J-binding protein 2 is a thymidine hydroxylase catalyzing the first step in the biosynthesis of DNA base J. *Mol Biochem Parasitol* 164: 157-161
 17. Genest, PA, ter Riet, B, Dumas, C, Papadopoulou, B, van Luenen, HG, Borst, P. 2005. Formation of linear inverted repeat amplicons following targeting of an essential gene in *Leishmania*. *Nucleic Acids Res* 33: 1699-1709
 18. Klein, SM, Cohen, G, Cederbaum, AI. 1981. Production of formaldehyde during metabolism of dimethyl sulfoxide by hydroxyl radical generating systems. *Biochemistry* 20: 6006-6012
 19. Ryle, MJ, Padmakumar, R, Hausinger, RP. 1999. Stopped-flow kinetic analysis of *Escherichia coli* taurine/ α -ketoglutarate dioxygenase: interactions with α -ketoglutarate, taurine, and oxygen. *Biochemistry* 38: 15278-15286

CHAPTER 4

CHARACTERIZATION OF A *TRYPANOSOMA BRUCEI* ALKB HOMOLOG CAPABLE OF REPAIRING ALKYLATED DNA

Portions of the introduction to this chapter were published in:

Simmons, JM, Müller, TA, Hausinger, RP. 2008. Fe(II)/alpha-ketoglutarate hydroxylases involved in nucleobase, nucleoside, nucleotide, and chromatin metabolism. *Dalton Trans.* 5132-5142.

ABSTRACT

Trypanosoma brucei encodes a protein homologous to AlkB of *Escherichia coli*, raising the possibility that trypanosomes catalyze oxidative repair of alkylation-damaged DNA. The gene encoding the *T. brucei* AlkB homolog (TbABH) was cloned and expressed in *E. coli*, the recombinant protein was purified and characterized, and the capacity of the protozoal gene to complement *alkB* bacteria cells was examined. When anaerobic TbABH was incubated with Fe^{II} and α -ketoglutarate (α KG), the sample produced a characteristic metal-to-ligand charge-transfer chromophore, confirming its membership in the Fe^{II}/ α KG dioxygenase superfamily. TbABH was shown to bind to DNA, exhibiting a clear preference for binding to alkylated DNA according to results derived by electrophoretic mobility shift assays. Further, TbABH partially complemented an *alkB* knockout in *E. coli* cells stressed with methylmethanesulfonate, confirming the assignment of TbABH as a functional AlkB protein in *T. brucei*.

4.1 INTRODUCTION

Trypanosomes are eukaryotic extracellular parasites that cause diseases in various mammalian hosts (1). They are evolutionarily divergent organisms that branched early from the phylogenetic tree, before fungi, plants, or animals (2). *Trypanosoma brucei*, the causative agent of African Sleeping Sickness in humans, has a dense exterior protein coat that limits access to the parasite's cellular membrane (3,4). In bloodstream form cells, the coat protein is expressed from a single allele and changes periodically in a population to evade detection by the host immune system; a process called antigenic variation (5). The genes encoding the repertoire of coat proteins are located primarily in the subtelomeric regions among large patches of repetitive sequence and are relocated by homologous recombination into a promoter-containing expression site as needed (6,7).

The many repetitive regions in the *T. brucei* genome and the high frequency of recombination events lead to the need for efficient DNA repair and maintenance mechanisms. Components for most of the common DNA repair pathways have been identified in trypanosomes and some have been characterized. Proteins homologous to certain elements of the mammalian mismatch repair pathway (8), base excision repair (9), nucleotide excision repair (10) and homologous recombination (7) are present. One method of DNA repair not yet described in trypanosomes is direct repair which can be catalyzed by several DNA glycosylases, alkyl transferases, and hydroxylating enzymes in other systems. One such hydroxylase is AlkB, which has been best characterized in *Escherichia coli*.

E. coli alkB has been studied for its role in the adaptive response to alkylation damage since the 1980s (11,12). This gene was long known to confer resistance to certain methylating agents, and in 2002 the encoded protein was discovered to be a member of the Fe^{II}/α-ketoglutarate (αKG) dioxygenases (13,14). The enzyme catalyzes N-dealkylation of 1-methyladenine (1meA) and 3-methylcytosine (3meC) in DNA by the mechanism shown in Figure 4.1, creating an unstable intermediate that spontaneously releases an aldehyde to regenerate the native base. AlkB repairs the analogous lesions in RNA (15), including mRNA and tRNA (16). Furthermore, the protein dealkylates 1-methylguanine, 3-methylthymine and several etheno adducts of DNA (17-20).

The crystal structure of *E. coli* AlkB in complex with Fe^{II}, αKG, and the tri-deoxynucleotide substrate T-1meA-T (Fig. 4.2) reveals details of the metallocenter active site and helps to clarify how the enzyme can accommodate its various substrates that include single-strand (ss)-DNA, double-strand (ds)-DNA, and various forms of RNA (21). Residues His 131, Asp 133, and His 187 bind the metal, αKG coordinates Fe^{II} in a bidentate manner (via the C1 carboxylate and C2 keto groups), and a salt bridge from Arg 204 stabilizes the αKG C5 carboxylate. The open coordination site for oxygen binding does not point toward the substrate (a situation termed “off-line” geometry) (22); thus requiring a shift in position of the C1 carboxylate prior to oxygen binding or a “ferryl flip” to occur after generating the Fe^{IV}-oxo intermediate (23). The trimer substrate is bent to extend the damaged base deep into the active-site pocket, making ten identified hydrogen bonds and many more van der Waals contacts with the protein.

Figure 4.1. Reactions of AlkB with 1meA and 3meC in DNA or RNA.

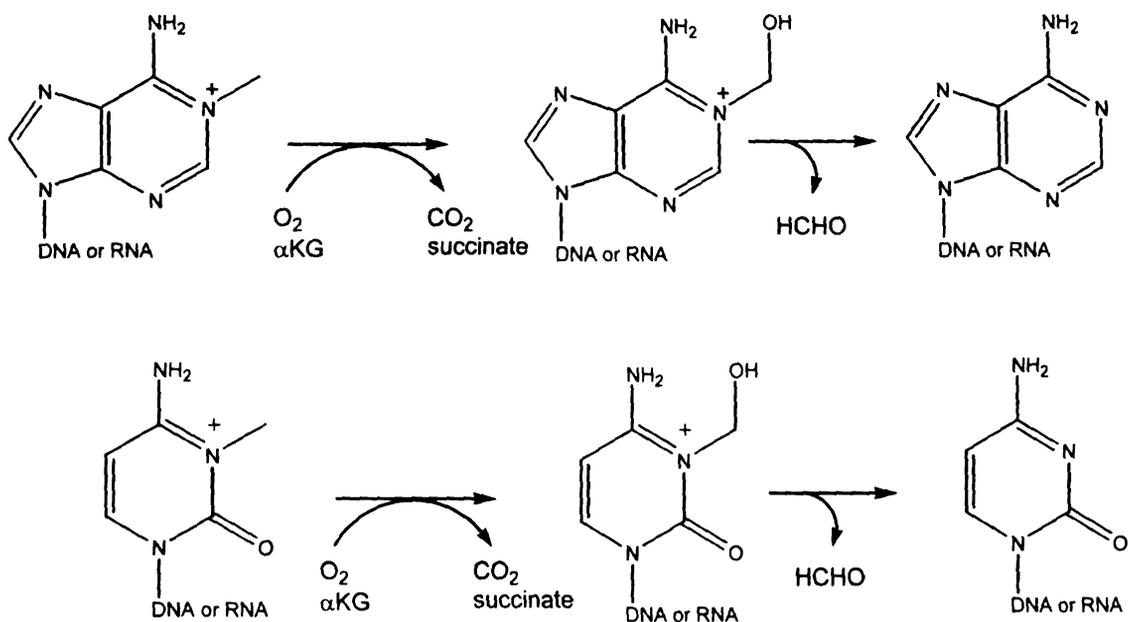
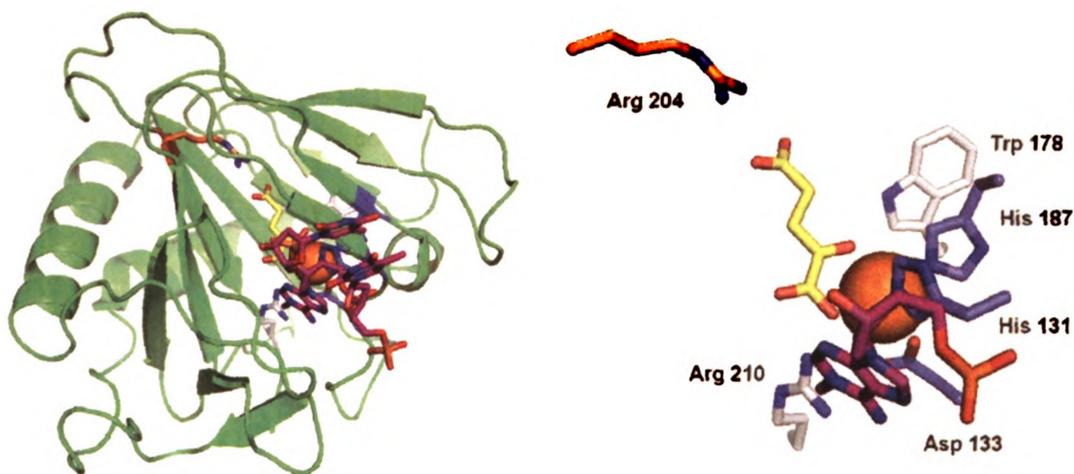


Figure 4.2. Structure of AlkB (PDB 2fd8). The left panel depicts the overall fold of the enzyme, while the right panel highlights the active site residues. Fe^{II} is an orange sphere, αKG has yellow carbon atoms, metal ligands have blue carbon atoms, a cofactor-stabilizing residue is shown with orange carbon atoms, and other postulated active site residues are indicated with white carbon atoms. The bound substrate is shown with magenta carbon atoms (abbreviated in the right panel). The nucleotide-binding lid (residues 45-90) is comprised of three β -strands (the left-most strand in the pair of β -sheets and two short β -strands at the bottom) along with several loops (bottom left) lacking well-defined secondary structures.



The AlkB N-terminus is responsible for about half of the identified contacts with the substrate via a highly flexible nucleotide recognition domain or lid made up of 3 β -strands (residues 45-90). This protein region is flexible based upon the superimposed structures of the bound and nucleotide-free states, thus allowing movement and reorientation without losing critical contacts. Substrate specificity at the base level is provided deep in the pocket; a hydrogen bond between Arg 210 and the adenine moiety could also be made with cytosine, but not with the other two bases, explaining why AlkB prefers 1meA and 3meC as substrates. Finally, AlkB's modest preference for ss-DNA/RNA substrates (15,24) is explained by the nucleotide polymer being oriented to extend into a hydrophobic pocket with stabilization provided by base stacking with nearby aromatic residues. By contrast, ds-DNA would need to unwind partially in order to allow the section of DNA containing the lesion to extend into the active site.

Kinetic studies have examined AlkB activity using a wide range of substrates. A rate of 12 min^{-1} was reported for randomly methylated poly(dA) (25), or about 4 min^{-1} for two specifically mono-methylated oligonucleotides (26). Catalytic rates with the tri-deoxynucleotide T-1meA-T used in the crystal structure studies were determined separately by two groups and found to be 7.4 and 4.5 min^{-1} , respectively (21,25). A phosphate 5' of the methylated base is critical for proper recognition based upon a 6-fold increase in the apparent K_m when it is missing (25), and this result is compatible with the crystal structure which shows both a hydrogen bond and a salt bridge involving this 5' phosphate (21).

In addition to the dealkylation reactions observed with the above substrates, AlkB is known to exhibit additional types of reactivity. The minimal substrate that could be

demethylated was identified as 1me-dAMP (25), whereas the free nucleosides 1meA, 1me-2'-dA, 3meC, and 3me-2'-dC were shown to stimulate α KG decomposition without undergoing demethylation (27). In the absence of nucleotide, but the presence of α KG and O₂, AlkB hydroxylates its own Trp 178 (located near the active site, see Fig. 4.2) to generate a hydroxy-tryptophan side chain. A ligand-to-metal charge-transfer transition between this modified residue and the oxidized metal site results in a blue protein that absorbs maximally at 595 nm (28).

This chapter focuses on an AlkB homolog of *T. brucei* and describes the DNA repair activity observed for this member of the non-heme Fe^{II}- and α KG-dependent hydroxylase family of enzymes in kinetoplastids.

4.2 MATERIALS AND METHODS

4.2.1 Gene identification and multiple sequence alignment

The Basic Local Alignment Search Tool (29) was utilized to search the protein-encoding sequences of the *Trypanosoma brucei brucei* genome with the protein sequence of *E. coli* AlkB as the query, resulting in the identification of a sequence with the NCBI accession number XP_844196. The sequence of the identified trypanosomal AlkB homolog (TbABH) was aligned to representative group 1A (30) AlkB sequences of *E. coli* (NP_416716), *Brucella abortus* (ZP_05894130.1), *Pseudomonas putida* (AAN69003.1), *Pseudomonas syringae* (NP_792910.1), and the first three human AlkB homologs (ABH1, AAH25787.1; ABH2, Q6NS38.1; ABH3, Q96Q83.1) by using Clustal W (31) for analysis. Further, the TbABH sequence was analyzed to predict the subcellular location by several online servers including LOCTree (32), PSORTII (<http://psort.hgc.jp/form2.html>), SubLoc (33), and ESLPred (34).

4.2.2 Cloning

A 991-bp DNA fragment containing *TbABH* was amplified by PCR using genomic *T. b. brucei* strain 427 DNA as a template with “forward” (5'-AGGATATACC ATGGAAGACC CCGTGC-3') and “reverse” (5'-GA GCATCCTCGAG TTCGTTAAG GAACTCAC-3') primers which introduce NcoI and XhoI restriction sites (underlined), and *Taq* polymerase master mix kit (Promega) which leaves a single 3' adenine nucleotide overhang. The PCR product was treated with a PCR clean up kit (Qiagen, Inc.) and ligated into pGEM-T Easy (Promega) to create pGEM-TbABH. The pGEM-

TbABH plasmid was transformed into *E. coli* DH5 α (Invitrogen), isolated from several transformants, and sequenced (Davis Sequencing). *TbABH* was excised from the pGEM-T backbone by NcoI and XhoI restriction and ligated into pET28b (Novagen) which had been cut previously with the same enzymes, creating pET-TbABH and putting the coding sequence in frame with a sequence encoding a C-terminal 6-histidine tag. This plasmid was transformed into the expression strain *E. coli* BL21 (DE3).

4.2.3 Protein production and purification

E. coli BL21 (DE3) cells containing pET-TbABH encoding TbABH-His₆ (hereafter referred to simply as TbABH) were grown at 30 °C in lysogeny broth (LB) medium supplemented with 100 μ g/mL kanamycin while shaking at ~160 rpm to an optical density of 0.4 to 0.6 at 600 nm. Cultures were induced to overexpress the desired gene by addition of isopropyl- β -D-thiogalactopyranoside (IPTG) to 0.1 mM, and grown for an additional 4 h, then harvested at 4 °C by centrifugation at ~8,000 g for 8 min. The cell paste was either used immediately for protein purification or stored at -80 °C.

In a typical purification, 3 mL of binding buffer (30 mM imidazole, 10 mM Tris, 150 mM NaCl, pH 7.9) was added per g of cell paste for resuspension. The protease inhibitor phenylmethylsulfonyl fluoride was added to 0.5 mM, cells were lysed by sonication (Branson Sonifier, 3 pulses of 1 min each, 3 W output power, duty cycle 50%, with cooling on ice), and the cell lysates were ultracentrifuged at 100,000 g for 1 h. Soluble cell extracts were loaded onto a Ni-bound nitrilotriacetic acid column (Qiagen) pre-equilibrated with binding buffer. The column was washed with binding buffer until the baseline was reestablished, and proteins were released with elution buffer (150 mM

imidazole, 10 mM Tris, 150 mM NaCl, pH 7.9). Fractions containing the purified proteins, as determined by denaturing sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE, 12% acrylamide) (35) and Coomassie staining, were pooled and dialyzed back into binding buffer by using 12-14 kDa molecular weight cutoff dialysis tubing (Fisher) at 4 °C overnight with stirring. The concentration of protein was determined by using its calculated molar absorptivity at 280 nm ($45,350 \text{ M}^{-1} \text{ cm}^{-1}$) according to the ExPasy protein parameters prediction server (<http://ca.expasy.org/tools/protparam.html>) (36). The protein was either used immediately for assays or was stored at 4 °C and discarded after two weeks or if precipitation developed. A typical purification yielded approximately 3 mg protein per g of cell paste.

4.2.4 Gel filtration chromatography

To determine the native size and oligomeric state of the protein, purified TbABH was concentrated to 200 μM (7.4 mg/mL) of protomer in an Amicon centrifugal filter unit with a 10 kDa molecular weight cutoff and was chromatographed on a Superdex®75 size exclusion column preequilibrated with binding buffer. Retention times were compared to those of gel filtration standards (BioRad). The collected fractions also were examined by PAGE using a 3-12% Bis-Tris Blue native gradient gel (Invitrogen) to verify the major oligomeric forms of TbABH.

4.2.5 Spectroscopy

UV/visible spectra were obtained by using an HP 8453 spectrophotometer (Hewlett Packard) equipped with a circulating water bath and magnetic stirrer. TbABH was concentrated in an Amicon centrifugal filter unit with a 10 kDa cutoff to 265 μM of protomer in binding buffer. Stock solutions of 10 mM Fe^{II} and 50 mM αKG were prepared by adding the dry reagents to vials, subjecting them to repeated cycles of vacuum and argon, and adding anaerobic water by using a syringe. Protein was added to a 1-cm path length, 1.5 mL, quartz cuvette, the sample was made anaerobic by gentle vacuum/argon cycling, and it was scanned while stirring at 9 $^{\circ}\text{C}$. Four equivalents of αKG (to 1 mM final concentration) were added and the sample was re-scanned, followed by titration with iron in 8 increments up to 500 μM with scanning after each addition. Data were corrected for dilution in Excel and plotted as difference curves with the absorbance from the protein alone set as the baseline. The difference absorbance of the feature observed at 530 nm (ΔA) was plotted versus the concentration of total added iron (L_t) to calculate a K_d for the binding of iron to the αKG -bound active site using equation 1 (37). In this equation, ΔA_{max} is the predicted maximal absorbance change, $[E_t]$ is the total enzyme concentration, and n is the number of iron binding sites per protein molecule.

$$\text{Eq. 1 } \Delta A = \Delta A_{\text{max}} * \frac{((K_d + L_t + n[E_t]) - ((K_d + L_t + n[E_t])^2 - (4L_t * n[E_t]))^{0.5})}{2 n[E_t]}$$

4.2.6 Electrophoretic mobility shift assays (EMSA)

Gel band shift assays were conducted after incubating 50 ng of a plasmid ds-DNA substrate (7.5 nM in assay) with increasing amounts of TbABH (up to 20 μM) for 1-2 h at

30 °C. This maximum amount of TbABH equates to a final ratio of protein to DNA base pairs of 1:2. These assays were analyzed by electrophoresis using a 0.8% agarose gel and visualized via staining with ethidium bromide. DNA binding was evidenced by a shift in the electrophoretic mobility of the DNA band in the gel upon increasing protein concentration.

Other gel band shift assays were conducted by incubating ³²P-labeled oligonucleotides with increasing amounts of TbABH and monitoring the effects on electrophoretic mobility when using a 4% polyacrylamide gel via autoradiography. Briefly, oligonucleotides (forward 5'-CGA TCC AGA CTC GAC TAC GCC ATC CGA TCC-3' and reverse 5'-GGA TCG GAT GGC CTA CTC CAC TCT GGA TCG-3') were treated with methyl methanesulfonate (MMS) to 1.5% in 20 mM cacodylate buffer, pH 7.5, for 5 h at 31 °C then ethanol precipitated to recover the DNA. 50 pmol of the forward oligo (either untreated or methylated) was treated with 50 µCi of γ-³²P-ATP and 10 units of T4 polynucleotide kinase (Gibco) in the provided buffer for 50 min at 37 °C. DNA was denatured and purified on a 12% agarose gel containing 7 M urea. Gel slices containing the labeled DNA oligonucleotides were excised and the DNA extracted, ethanol precipitated to purify, and quantified by scintillation counting. 50 nM stocks were made for each of the DNA substrates used in the binding assays, each at 10 kcpm/µL. For ds-substrate, forward and reverse oligos were combined and incubated at 70 °C for 1 min, and then slow cooled to 27 °C to allow hybridization.

For binding assays, 10 nM oligo was combined with increasing TbABH concentration from 0-30 µM in 10 µL (up to a 100-fold excess of protein subunit per base pair) containing 20% glycerol as a molecular crowding agent for 1 h at room temperature.

Samples were electrophoresed on a 4% polyacrylamide gel containing a 37.5:1 ratio of acrylamide/bisacrylamide in Tris-glycine in a BRL V16-2 system at 20 mA per gel until the dye front was about half way through the gel. Gels were fixed in a solution containing 10% each of methanol and acetic acid and dried by vacuum. Dried gels were exposed to a phosphorescence screen for at least 2 h on a Typhoon 9200 variable mode imager. The images were analyzed by using Image Quant, with the count values of the shifted bands normalized to the total counts in each lane, converted to nM shifted DNA, and plotted against the total TbABH concentration in the sample. The resulting data were fit in Sigma Plot to equation 2 for hyperbolic single-site saturation, where B_{\max} is the theoretical maximum DNA bound, and used to estimate the K_d of DNA binding. The results shown are the average of duplicate experiments.

Eq. 2
$$y = \frac{B_{\max} * [E]}{K_d + [E]}$$

To test binding of TbABH to another form of damaged DNA, oligonucleotides containing abasic sites were created and used in analogous EMSA studies. Briefly, oligonucleotides (forward 5'-AGT AGA CAG CTA CCA TGC CTG CAC GAA GUT AGC AAT TCG TAA TCA TGG TCA TAG CTA GTA -3' and reverse 5'- TAC TAG CTA TGA CCA TGA TTA CGA ATT GCT AGU TTC GTG CAG GCA TGG TAG CTG TCT ACT-3') each containing a single deoxyuridine (in bold) were hybridized by heating to 95 °C for 1 min then allowed to slow cool to room temperature. The ds-DNA was treated with uracil-DNA glycosylase (New England Biolabs) following the manufacturer's instructions to induce formation of abasic sites. For the assays, 1 μ M of

treated or untreated oligonucleotides was incubated with increasing amounts of TbABH (0-30 μM) in 25 mM Tris, pH 8, containing 50 mM KCl for 20 min at 37 °C. Samples were loaded onto a 6% polyacrylamide gel (19:1 acrylamide:bisacrylamide ratio) and electrophoresed at 90 V in TBE buffer (89 mM each of Tris and borate, 1 mM EDTA, pH 8) and visualized by staining with ethidium bromide.

4.2.7 *In vitro* enzyme assays

Production of formaldehyde from methylated DNA was monitored by using a formaldehyde dehydrogenase-coupled assay which converts the product into formate under conditions that are essentially irreversible (42). The reaction was monitored spectrophotometrically at 340 nm to follow the coupled reduction of NAD^+ to NADH. Assays containing 1 μM TbABH, 100 μM αKG , 5 μM Fe^{II} , 13 μM methylated 25-mer poly-dA oligonucleotide (~30 μM methylated bases assuming a typical ~10% methylation efficiency), 1 mM NAD^+ , and 0.05 units of formaldehyde dehydrogenase in 50 mM HEPES, pH 8, were mixed in a 300 μL quartz, black walled cuvette and continuously monitored at 340 nm in a Shimadzu spectrophotometer. A series of control experiments were carried out, omitting individual components of the TbABH assay. A standard curve demonstrated that formaldehyde was detectable in a linear fashion from 150 μM down to 5 μM in the assay volume.

Consumption of αKG was monitored by using *ortho*-phenylenediamine (OPDA) which is known to react with several α -keto acids (43). Assays containing 1 μM TbABH, 13 μM methylated poly-dA oligonucleotide, 5 μM Fe^{II} , and 100 μM αKG in 50 mM HEPES, pH 8, were incubated at 37 °C. Samples (225 μL) were taken at selected time

points and quenched by addition of 75 μ L OPDA solution (10 mg/mL OPDA in 1 M phosphoric acid adjusted to pH 2 with .25% β -mercaptoethanol). Quenched samples were heated in a 100 $^{\circ}$ C block for 3 min, cooled to room temperature, and monitored at 340 nm in a quartz, black-walled, microcuvette by using a Shimadzu spectrophotometer. The standard curve for the OPDA reaction with α KG was linear from 5 μ M to 100 μ M.

4.2.8 Complementation assays

TbABH was cloned out of pET-TbABH by using PCR amplification primers (forward 5'-CGTGGAATTC ATGGAAGACC CCGTGC-3' and reverse 5'-GCT CCAAGCTTTC ATTCGTTAAG GAACTCAC-3') that introduced EcoRI and HindIII restriction sites (underlined). The DNA containing the gene was excised using those enzymes and ligated into pUC18 to create pUC-TbABH. This construct was transformed into the BW25113 Δ *alkB* cell line (38) and grown to exponential phase before plating 100 μ L on LB agar containing 200 μ g/mL ampicillin. After allowing the plates to dry, sterile paper discs soaked in 1% or 5% MMS were placed on top of the agar, the plates were incubated overnight at 37 $^{\circ}$ C, and the resulting zones of inhibition were analyzed to monitor cell stress according to the Kirby-Bauer disc diffusion method (39).

In an attempt to increase TbABH protein production during the complementation studies, *TbABH* was cloned into the IPTG-inducible vector pEXT20 (40) by using pET-TbABH as a template and PCR amplification primers (forward 5'-GAT GAATTCTGAA GGAGATATAC ATGGAAGACC CCGTGCG-3' and the same reverse primer as above) that introduced an EcoRI restriction site (underlined) while preserving the

ribosomal binding site (boxed). The construct was transformed into BW25113 Δ *alkB*, and the culture was grown, plated, and stressed with MMS as above.

Another vector used to enhance TbABH protein production during complementation was pMMB67, which was created by and obtained from Dr. Michael Bagdasarian. This is a low copy number plasmid with an IPTG-inducible *Taq* promoter, ampicillin resistance gene, RSF1010 replicon, and a multiple cloning site derived from pUC18. *TbABH* was subcloned from pUC-TbABH by using the restriction enzymes EcoRI and HindIII and ligated into similarly digested pMMB67, and the xenobiotic stress assays were repeated with BW25113 Δ *alkB* cultures containing this recombinant plasmid.

TbABH was found to be out of frame in pUC-TbABH, so this plasmid was mutagenized by using PCR amplification primers (forward 5'-GATTACGAAT TCGATGGAAG ACCCCG-3' and reverse 5'-CGGGGTCTTC CATCGAATTC GTAATC-3') that inserted a base downstream of the EcoRI site (underlined) and before the *TbABH* start site (boxed), placing the gene in-frame with the *lacZa* start site which added 7 extra residues to the N-terminal end of the produced protein. The new construct, denoted pUC-TbABH2, was verified by sequencing (Davis Sequencing) and transformed into BW25113 Δ *alkB* cells. Cultures were grown to exponential phase before being induced with 50 μ M IPTG. After one doubling time of additional growth to allow protein production, 100 μ L of this culture (and the necessary control cultures) were plated on LB agar containing 200 μ g/mL ampicillin and 50 μ M IPTG. After reaching dryness, the Kirby-Bauer disc diffusion assays were repeated using 1% and 5% MMS.

Finally, *TbABH* was cloned into pWKS30 (41) by using pET-TbABH as a template and PCR amplification primers (forward 5'-CGTGGGTACC GAAGGAGATA

TACCATGG-3' and the same reverse primer used to clone the gene into pUC18 from pET-TbABH) that introduced a KpnI restriction site, underlined, while preserving the ribosomal binding site (boxed) and the start site of *TbABH* (bold). The gene was ligated into pGEM-T Easy (Promega), verified by sequencing (Davis Sequencing), subcloned into pWKS30 by using the KpnI and HindIII restriction sites, and the latter plasmid was transformed into the knockout cell line. Kirby-Bauer assays were repeated with this construct as before.

In order to verify the sensitivity differences between the wild type and *alkB* knockout BW25113 cell lines to MMS, a viable cell count assay was performed. Briefly, cultures were grown to exponential phase in LB medium, stressed with 0.5 % MMS, and samples were taken at various time points, diluted, and plated on LB agar. Plates were incubated overnight at 37 °C and the colony forming units (CFUs) were counted. CFUs were corrected for dilution and plotted as total viable cell count versus time under the stress condition.

4.3 RESULTS AND DISCUSSION

4.3.1 Analysis of the TbABH Sequence

The genome of *T. brucei brucei* encodes a full length ortholog of the *E. coli* AlkB protein, that we have termed *T. brucei* AlkB homolog (TbABH). TbABH was predicted to be a nuclear protein by the ESLPred, SubLoc, and PSORTII online prediction servers, and further identified as a DNA-binding protein by LocTree. TbABH is 23% identical to *E. coli* AlkB with the most conserved regions predicted to be located in the double-stranded β -helix and including all the putative metal-coordinating facial triad residues (His 213, Asp 215, and His 269) as well as the probable α KG-stabilizing Arg 305 (Fig. 4.3). These residues, along with several others, were completely conserved among the bacterial and eukaryotic AlkB-related sequences aligned. This is not surprising since the core fold is one of the only defining features of this enzyme family. In fact, most of the conserved residues mapped to the β -strands making up the catalytic core of the *E. coli* AlkB crystal structure (21). In addition, several more residues were completely conserved between TbABH and the bacterial sequences including a few components of the nucleotide recognition lid in *E. coli* (residues 45-90) known to contact the nucleotide substrate (Trp 151, Tyr 158, and Ser 161 in TbABH).

Interestingly, there was much greater similarity between the TbABH sequence and that of the bacterial AlkB sequences than between TbABH and the other eukaryotic AlkB homologs represented in the multiple sequence alignment. This result is consistent with this protozoan gene being acquired by horizontal gene transfer. Such is the case for several other genes in the trypanosomatids including elements of the heme

biosynthetic pathway (44), phospholipase A1 (45) and isopropyl alcohol dehydrogenase (46). Despite TbABH's greater similarity to the bacterial AlkB-like proteins than to homologs in eukaryotes, it is distinct in possessing a much longer amino terminus and inserts of 37, 1, and 19 residues compared to the *E. coli* protein. The largest insert extends from within the nucleotide-binding lid of AlkB, the extra amino acid at residue 193 corresponds to an unstructured region on the back face of the *E. coli* protein in the view of Figure 4.2, and the 19 amino acid insertion (residues 281-299) corresponds to an unstructured region at the very top of AlkB in the view shown.

4.3.2 General biochemical properties of TbABH

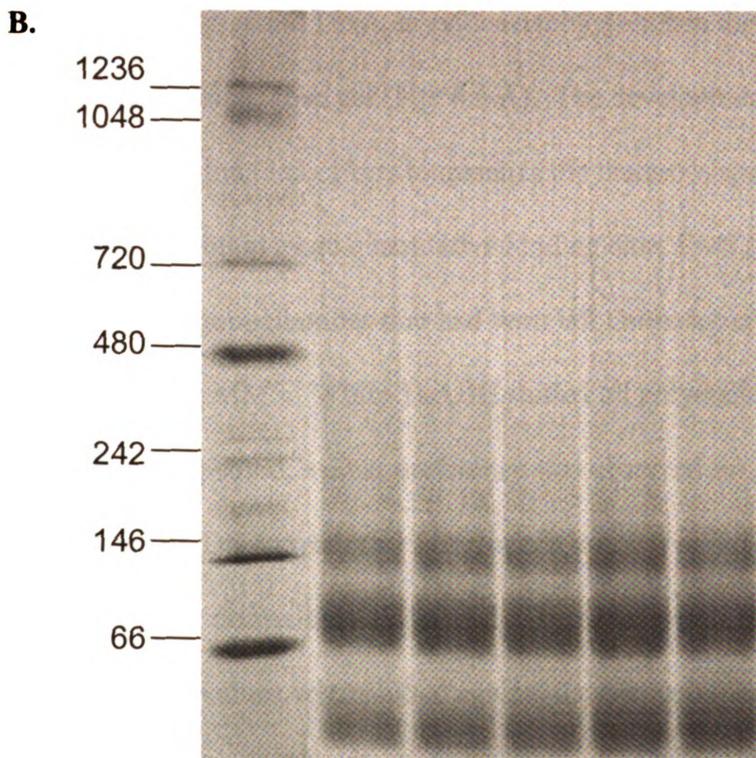
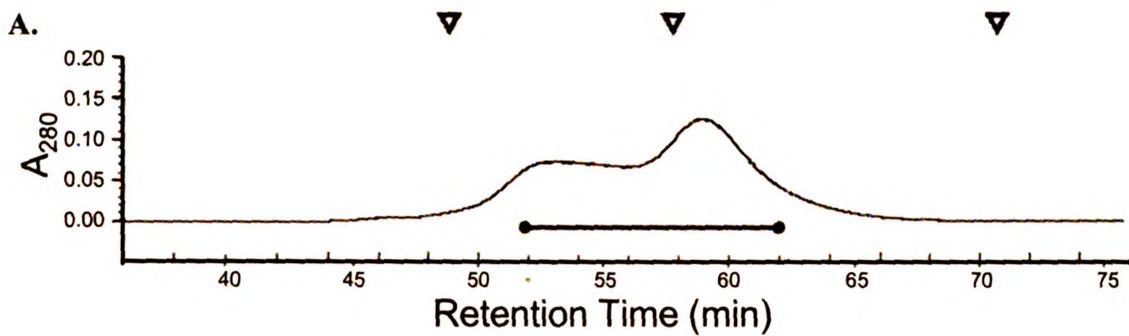
To characterize this protein, *TbABH* was cloned from *T. b. brucei* genomic DNA, expressed recombinantly in *E. coli*, and the protein purified by affinity chromatography to about 90% purity as visualized via SDS-PAGE and Coomassie staining (data not shown). Figure 4.4 shows that multiple forms of the protein were present as purified from *E. coli*. The predominant peak in the chromatogram (panel A) chromatographed as a monomer while the less prominent peak eluted as a dimer; however, native gel electrophoresis of the fractions was most consistent with both protein species forming an equilibrium mixture of 37 kDa monomers, 74 kDa dimers, and 148 kDa tetramers.

To confirm proper folding of TbABH, the purified protein was examined for its ability to form a characteristic chromophore seen in other members of the Fe^{II}/αKG-dependent dioxygenases and associated with binding of the iron cofactor and the αKG cosubstrate (47). The protein was monitored by UV-visible spectroscopy under anaerobic conditions, using low temperature to maintain stability, while titrating in αKG

Figure 4.4. Native size and oligomeric state determination of TbABH. **A.**

Chromatogram of TbABH analyzed by using a Superdex[®]75 column and compared with BioRad gel filtration standards (retention times of γ -globulin, 158 kDa; ovalbumin, 44 kDa; and myoglobin, 17 kDa, are shown as inverted triangles above the chromatogram).

B. Native polyacrylamide gradient gel showing the oligomeric states and native sizes of TbABH. The bar on the chromatogram indicates the fractions shown on the gel.



and Fe^{II}. As shown in Figure 4.5, anaerobic TbABH generated the diagnostic metal-to-ligand charge-transfer transition at 530 nm when both metal and cofactor were present. The extinction coefficient for this feature was approximately 190 M⁻¹ cm⁻¹, consistent with values previously described for members of this enzyme family (37,48). The relative absorbance change at 530 nm was plotted versus the total concentration of iron titrated into the protein sample and fit using equation 1. This allowed for calculation of an Fe^{II} K_d of ~4 μ M with 1.3 \pm 0.07 iron atoms binding per active site.

4.3.3 DNA binding by TbABH

To test the hypothesis that TbABH is a DNA repair protein, the ability of this protein to bind to potential DNA substrates was tested. TbABH was able to bind to supercoiled ds-plasmid DNA as evidenced by a shift in the electrophoretic mobility of the DNA band in an agarose gel (Fig. 4.6 A). The development of multiple TbABH-DNA complexes confounded efforts to quantify the thermodynamics of this interaction. Therefore, to obtain more quantitative binding data, EMSA assays were performed using ³²P-labeled oligonucleotides that had been left untreated or had been subjected to methylation by MMS. While TbABH shifted all polynucleotides to some extent when tested by this method, a clear preference was observed for ds-DNA and especially for methylated ds-DNA over other compounds tested (Fig 4.6 B). These band intensities were analyzed, normalized to the total ³²P counts in each lane, and converted to nM DNA bound. These values were plotted against the concentration of added TbABH and the resulting curve was fit to a hyperbolic single site binding equation (equation 2). This allowed the estimation of a K_d of 7.1 \pm 1.9 μ M (Fig 4.6 C). Although these assays

Figure 4.5. Spectroscopic evidence for binding of Fe^{II} and α KG by TbABH. **A.** The anaerobic UV/visible spectrum of TbABH (266 μ M protomer in binding buffer) was examined for the sample as isolated (baseline), after adding 1 mM α KG, and while titrating in Fe^{II} with stirring at 9 °C. The (Fe^{II}- α KG-protein minus protein) difference spectra shown correspond to the addition of 0, 63 μ M, 125 μ M, 188 μ M, 250 μ M, 313 μ M, 375 μ M, 438 μ M, and 500 μ M metal ions. **B.** The intensity of the absorbance difference at 530 nm was examined as a function of added Fe^{II}. The data were fit to equation 1.

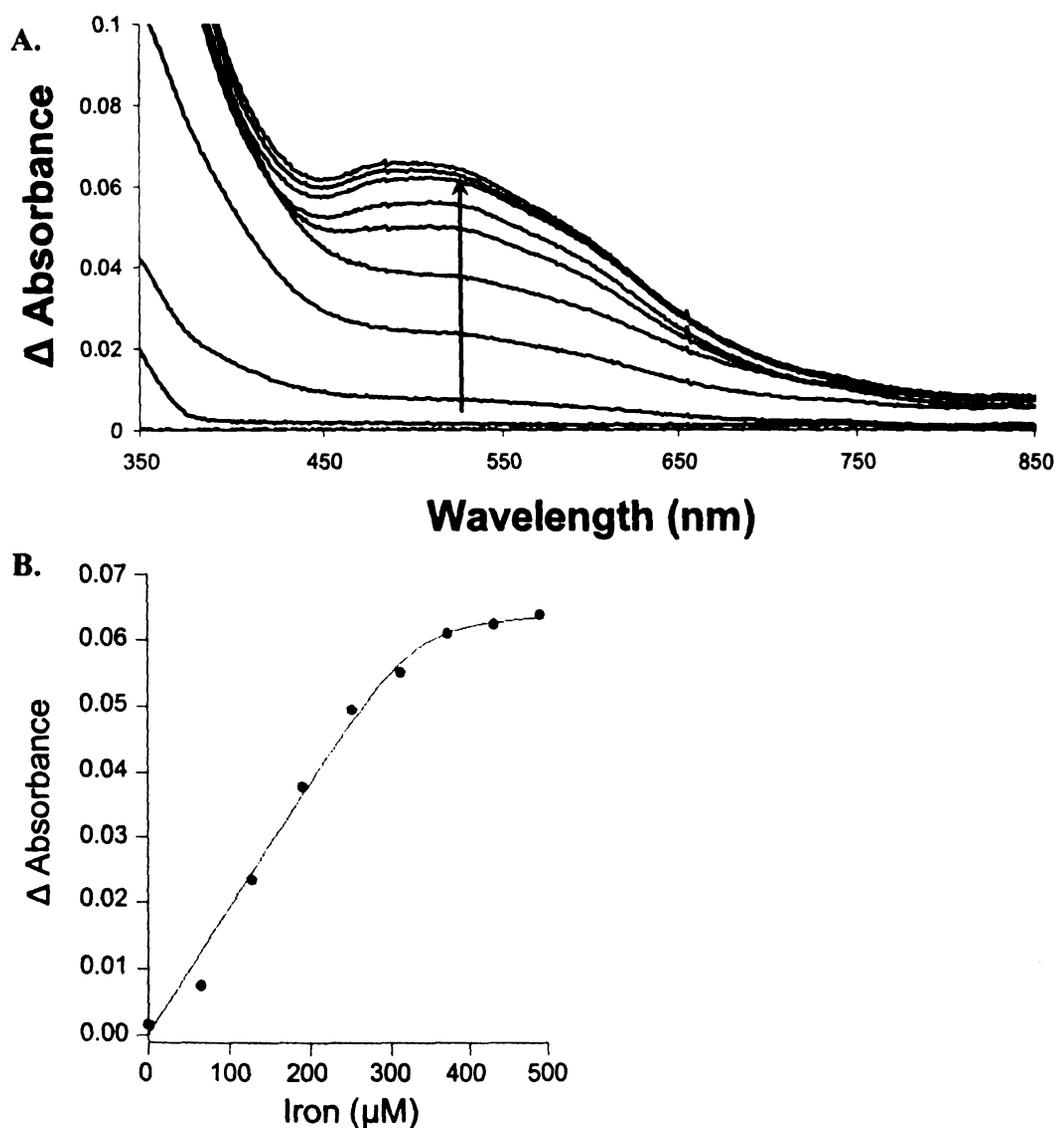
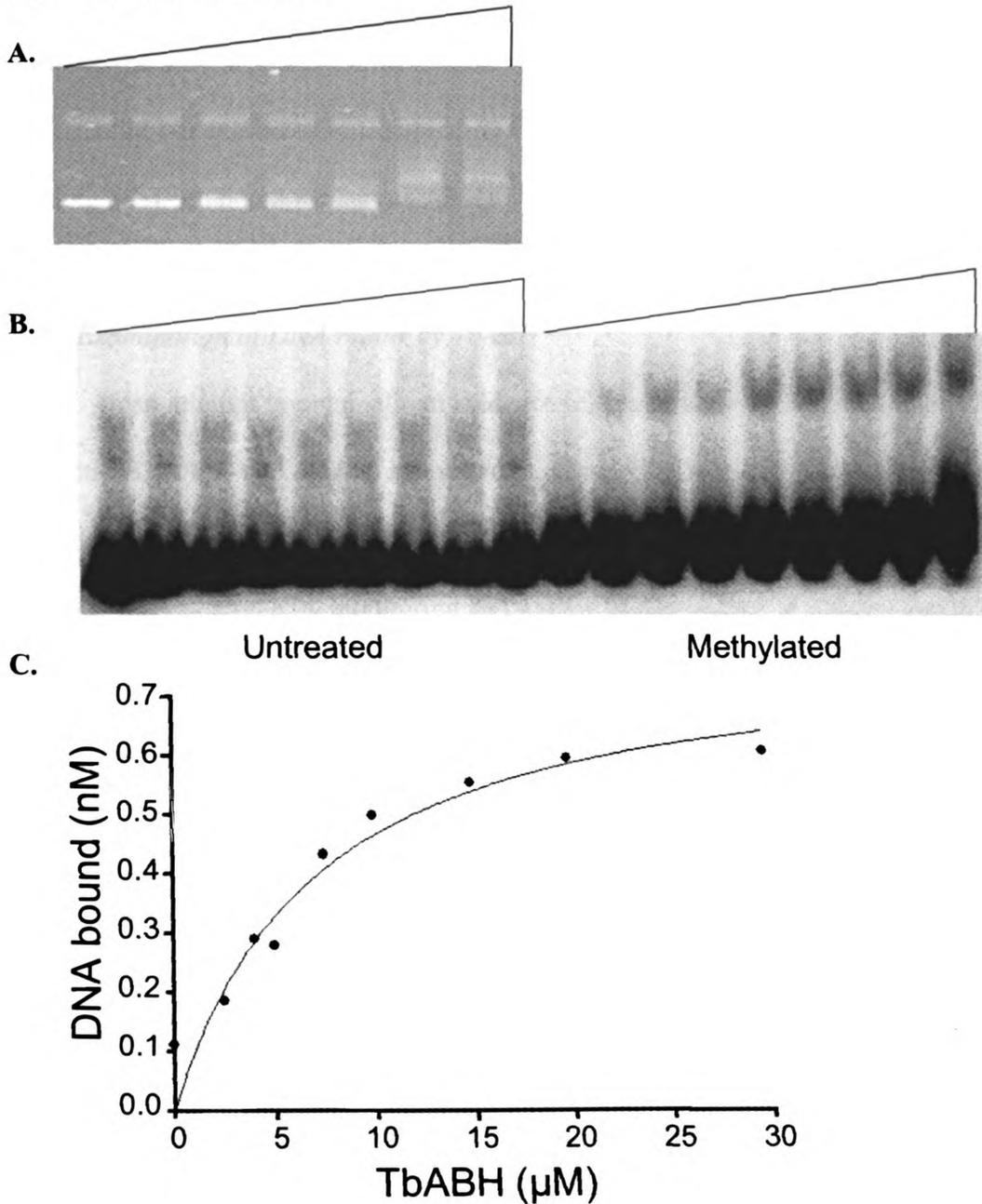


Figure 4.6 EMSA studies of TbABH and various DNA substrates. **A.** Agarose gel depicting the shift in mobility of supercoiled pGEX5 plasmid (7.5 nM) with increasing TbABH concentration (shown by the wedge, 0 to 20 μM). **B.** Native polyacrylamide gel showing the shift in mobility of ^{32}P -labeled ds-oligonucleotide (7.5 nM; untreated on left and methylated on right) with increasing TbABH concentration (0 to 30 μM). **C.** Fit of data from B to equation 2.



contained 10 nM DNA, the process used for methylating the substrate typically exhibits poor efficiency (~10%); thus, the observation that saturation (B_{max}) corresponded to 0.79 ± 0.08 nM DNA indicates that nearly stoichiometric binding (close to 80%) of the methylated substrate was bound. In contrast to the above results showing TbABH's preference for binding to methylated ds-DNA, other oligonucleotides were bound with lower affinity. Very little binding was observed for ss-DNA with no differences noted between the untreated and the methylated substances (data not shown). Furthermore, TbABH did not exhibit enhanced binding to DNA containing an abasic site; where as this binding has been noted for human ABH1 (49).

4.3.4 Examination of DNA repair by TbABH

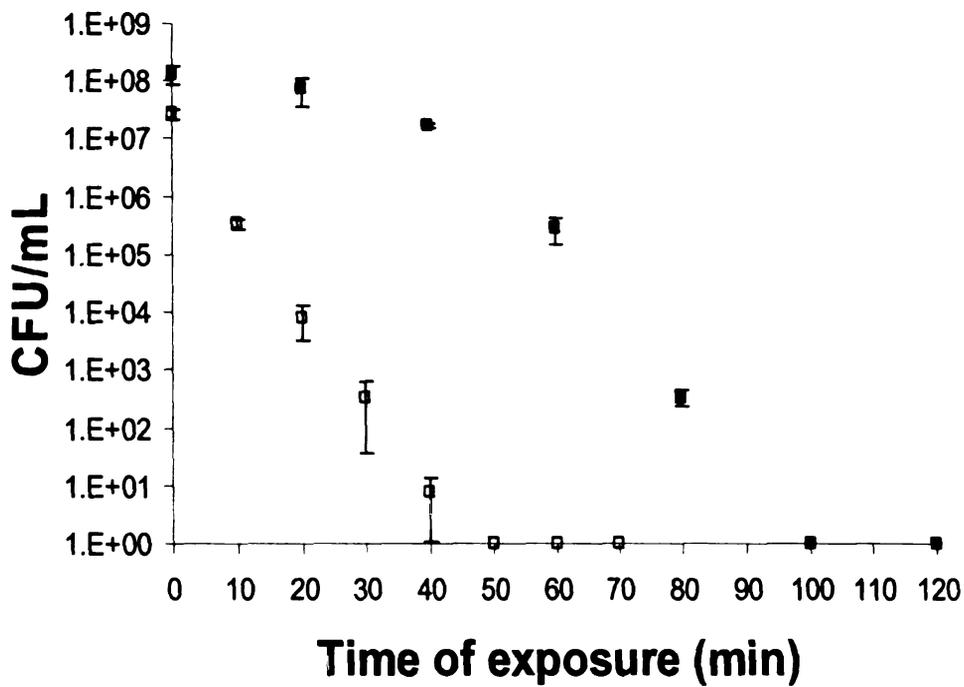
In an attempt to obtain direct evidence for the demethylation reaction expected of a functional AlkB, two *in vitro* assays were examined. The purified recombinant protein was used in a formaldehyde dehydrogenase-coupled reaction that allows for detection of released formaldehyde by monitoring an increase in absorbance at 340 nm due to the conversion of NAD^+ to NADH as part of the coupled reaction. While there was an indication of turnover in one experiment using this assay, only background rates were found in follow-up studies where the protein-free control also yielded a slightly positive reaction. Efforts to demonstrate α KG consumption by an OPDA assay also were inconclusive due to the low sensitivity of measuring substrate loss.

As an alternative approach, TbABH was tested for the ability to complement an *alkB* knockout in an *E. coli* cell line. The *alkB* gene of BW25113 Δ *alkB* cells is replaced with a kanamycin resistance cassette resulting in 3.5-fold greater susceptibility to

methylation damage by MMS as evidenced by the decreased viable cell counts over time when exposed to 0.5% of this methylating agent (Fig. 4.7). The *alkB* knockout cell line was transformed with pUC-TbABH2 (which placed the trypanosomal gene in the correct reading frame while adding 7 extra residues to the N-terminus of the protein) and tested for MMS sensitivity by the Kirby-Bauer method. At 1% MMS, wild type cells showed no inhibition of growth while the *alkB* knockout exhibited an average zone of inhibition (the distance between the edge of the paper disc and the edge of non-growth) of 4.0 ± 0.7 mm. The complemented strain containing pUC-TbABH2 partially complemented the knockout exhibiting a zone of inhibition of 1.5 ± 0.3 mm (Fig. 4.8).

To extend the *alkB* complementation approach, the trypanosomal gene was cloned into several other vectors and the disc diffusion assays were repeated. The BW25113 cell line does not contain the DE3 lysogen and thus cannot support T7 dependent expression plasmids, so efforts focused on three alternative vectors which would use the natural methionine start of *TbABH*. However derivatives of pEXT20, pMMB67, and pWKS30 containing *TbABH* failed to produce soluble protein in this cell line so the complementation studies could not be performed.

Figure 4.7 Time course viability test with wild-type and $\Delta alkB$ BW25113 cell lines stressed with 0.5% MMS. BW25113 wild type (filled symbols) or BW25113 $\Delta alkB$ (open symbols) cultures were grown to exponential phase, stressed with 0.5% MMS, sampled in a time course, diluted and plated to determine viable cell count. The time of exposure to MMS required to reduce viability by half was 3.5 fold greater for the wild type than the knockout cultures.



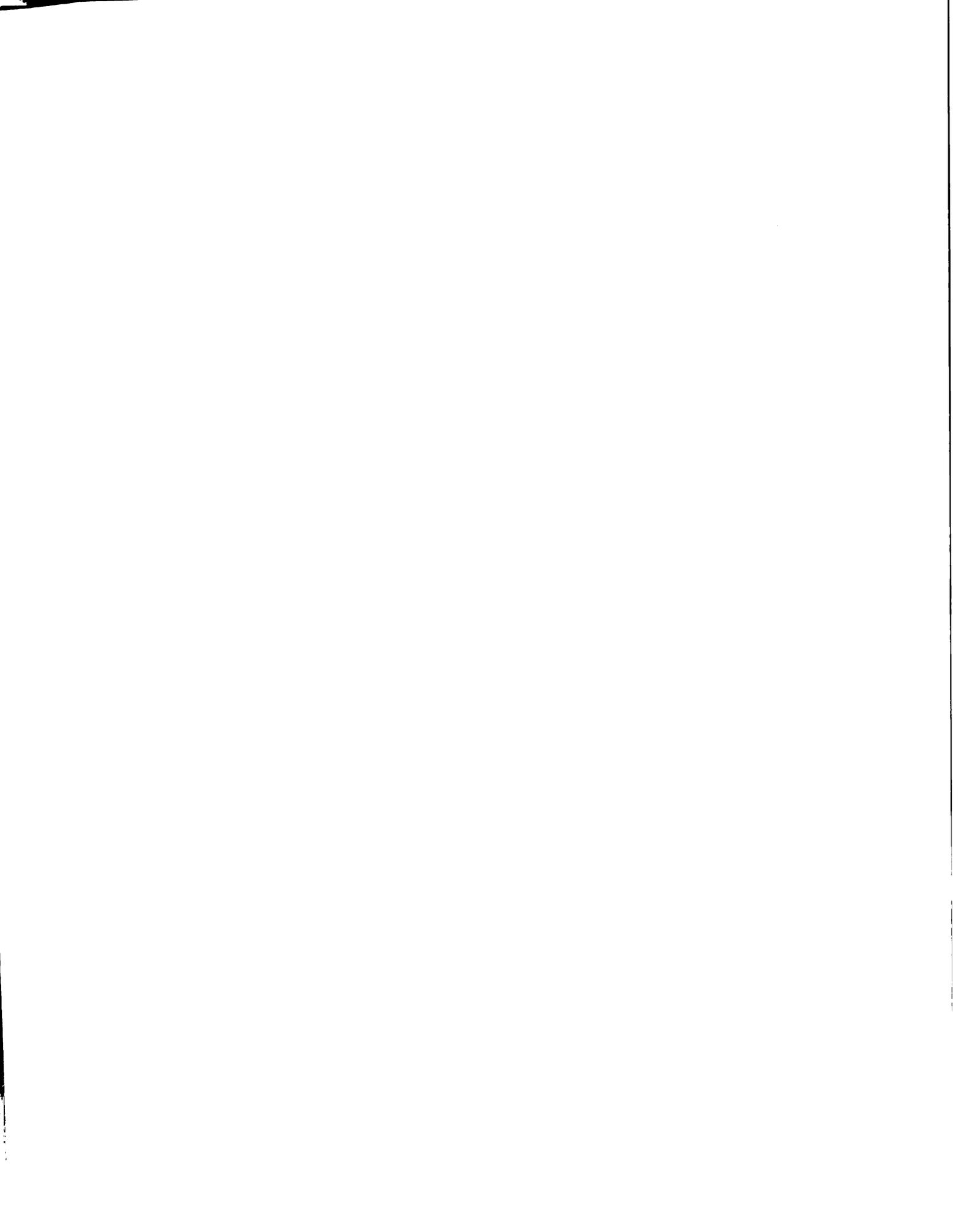
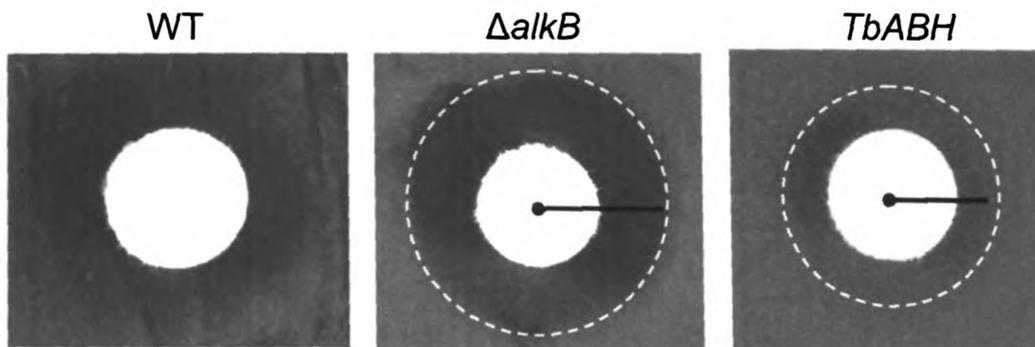


Figure 4.8 Complementation of an *E. coli alkB* mutant with *TbABH* under alkylation stress. BW25113 cells containing either pUC18 or the derivative pUC-TbABH2 were stressed with 1% MMS according to the Kirby-Bauer method (39). Plates were incubated overnight at 37 °C and the resulting zones of inhibition were measured. Below is a representative experiment showing the wild type cells containing pUC18 (labeled WT), $\Delta alkB$ cells containing pUC18 ($\Delta alkB$), and $\Delta alkB$ cells containing pUC-TbABH2 (*TbABH*). The zones of inhibition are shown by the black lines and white dashed circles.



4.3.5 Conclusions

Trypanosomes contain a full length ortholog of the *E. coli* AlkB which exhibits characteristics representative of the Fe^{II} and α KG dependent dioxygenase superfamily of enzymes including the formation of a diagnostic metal-to-ligand charge-transfer chromophore when incubated with the metal and cofactor under anaerobic conditions. Further, while *E. coli* AlkB binds DNA only very weakly (50), TbABH forms a tight complex and exhibits a clear preference for alkylated ds-DNA. Although direct evidence for DNA repair was not obtained *in vitro*, an *E. coli* cell line deficient in *alkB* was complemented by the expression of *TbABH* upon alkylation stress conditions confirming its assignment as an AlkB protein.

REFERENCES

1. Barrett, MP, Burchmore, RJ, Stich, A, Lazzari, JO, Frasch, AC, Cazzulo, JJ, Krishna, S. 2003. The trypanosomiases. *Lancet* 362: 1469-1480
2. Pace, NR. 2009. Mapping the tree of life: progress and prospects. *Microbiol Mol Biol Rev* 73: 565-576
3. Roditi, I, Furger, A, Ruepp, S, Schürch, N, Bütikofer, P. 1998. Unravelling the procyclin coat of *Trypanosoma brucei*. *Mol Biochem Parasitol* 91: 117-130
4. Field, MC, Carrington, M. 2009. The trypanosome flagellar pocket. *Nat Rev Microbiol* 7: 775-786
5. Cross, GA, Wirtz, LE, Navarro, M. 1998. Regulation of vsg expression site transcription and switching in *Trypanosoma brucei*. *Mol Biochem Parasitol* 91: 77-91
6. Horn, D, Barry, JD. 2005. The central roles of telomeres and subtelomeres in antigenic variation in African trypanosomes. *Chromosome Res* 13: 525-533
7. Glover, L, McCulloch, R, Horn, D. 2008. Sequence homology and microhomology dominate chromosomal double-strand break repair in African trypanosomes. *Nucleic Acids Res* 36: 2608-2618
8. Bell, JS, Harvey, TI, Sims, AM, McCulloch, R. 2004. Characterization of components of the mismatch repair machinery in *Trypanosoma brucei*. *Mol Microbiol* 51: 159-173
9. Castillo-Acosta, VM, Estevez, AM, Vidal, AE, Ruiz-Perez, LM, Gonzalez-Pacanowska, D. 2008. Depletion of dimeric all-alpha dUTPase induces DNA strand breaks and impairs cell cycle progression in *Trypanosoma brucei*. *Int J Biochem Cell Biol* 40: 2901-2913
10. Lee, JH, Nguyen, TN, Schimanski, B, Gunzl, A. 2007. Spliced leader RNA gene transcription in *Trypanosoma brucei* requires transcription factor TFIIF. *Eukaryot Cell* 6: 641-649
11. Kataoka, H, Yamamoto, Y, Sekiguchi, M. 1983. A new gene (*alkB*) of *Escherichia coli* that controls sensitivity to methyl methane sulfonate. *J. Bacteriol.* 153: 1301-1307

12. Kondo, H, Nakabeppu, Y, Kataoka, H, Kuhara, S, Kawabata, S, Sekiguchi, M. 1986. Structure and expression of the *alkB* gene of *Escherichia coli* related to the repair of alkylated DNA. *J. Biol. Chem.* 261: 15772-15777
13. Trewick, SC, Henshaw, TF, Hausinger, RP, Lindahl, T, Sedgwick, B. 2002. Oxidative demethylation by *Escherichia coli* AlkB directly reverts DNA base damage. *Nature* 419: 174-178
14. Falnes, PO, Johansen, RF, Seeberg, E. 2002. AlkB-mediated oxidative demethylation reverses DNA damage in *Escherichia coli*. *Nature* 419: 178-182
15. Aas, PA, Otterlei, M, Falnes, PO, Vagbo, CB, Skorpen, F, Akbari, M, Sundheim, O, Bjoras, M, Slupphaug, G, Seeberg, E, Krokan, HE. 2003. Human and bacterial oxidative demethylases repair alkylation damage in both RNA and DNA. *Nature* 421: 859-863
16. Ougland, R, Zhang, CM, Liiv, A, Johansen, RF, Seeberg, E, Hou, YM, Remme, J, Falnes, PO. 2004. AlkB restores the biological function of mRNA and tRNA inactivated by chemical methylation. *Molec. Cell* 16: 107-116
17. Koivisto, P, Robins, P, Lindahl, T, Sedgwick, B. 2004. Demethylation of 3-methylthymine in DNA by bacterial and human DNA dioxygenases. *J. Biol. Chem.* 279: 40470-40474
18. Delaney, JC, Essigmann, JM. 2004. Mutagenesis, genotoxicity, and repair of 1-methyladenine, 3-alkylcytosines, 1-methylguanine, and 3-methylthymine in *alkB* *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* 101
19. Mishina, Y, Yang, C-G, He, C. 2005. Direct repair of the exocyclic DNA adduct 1,N6-ethenoadenine by the DNA repair AlkB proteins. *J. Am. Chem. Soc.* 127: 14594-14595
20. Delaney, JC, Smeester, L, Wong, C, Frick, LE, Taghizadeh, K, Wishnok, JS, Drennan, CL, Samson, LD, Essigmann, JM. 2005. AlkB reverses etheno DNA lesions caused by lipid oxidation in vitro and in vivo. *Nature Struct. Molec. Biol.* 12: 855-860
21. Yu, B, Edstrom, WC, Hamuro, Y, Weber, PC, Gibney, BR, Hunt, JF. 2006. Crystal structures of catalytic complexes of the oxidative DNA/RNA repair enzyme AlkB. *Nature* 439: 879-884
22. Hausinger, RP. 2004. Fe(II)/ α -ketoglutarate-dependent hydroxylases and related enzymes. *Crit. Rev. Biochem. Mol. Biol.* 39: 21-68

23. Clifton, IJ, McDonough, MA, Ehrismann, D, Kershaw, NJ, Granatino, A, Schofield, CJ. 2006. Structural studies on 2-oxoglutarate oxygenases and related double-stranded β -helix fold protein. *J Inorg Biochem* 100: 844-868
24. Falnes, PO, Bjoras, M, Aas, PA, Sundheim, O, Seeberg, E. 2004. Substrate specificities of bacterial and human AlkB proteins. *Nucleic Acids Res* 32: 3450-3461
25. Koivisto, P, Duncan, T, Lindahl, T, Sedgwick, B. 2003. Minimal methylated substrate and extended substrate range of *Escherichia coli* AlkB protein, a 1-methyladenine-DNA dioxygenase. *J Biol Chem* 278: 44348-44354
26. Roy, TW, Bhagwat, AS. 2007. Kinetic studies of *Escherichia coli* AlkB using a new fluorescence-based assay for DNA demethylation. *Nucleic Acids Res* 35: e147
27. Welford, RWD, Schlemminger, I, McNeill, LA, Hewitson, KS, Schofield, CJ. 2003. The selectivity and inhibition of AlkB. *J Biol Chem* 278: 10157-10161
28. Henshaw, TF, Feig, M, Hausinger, RP. 2004. Aberrant activity of the DNA repair enzyme AlkB. *J Inorg Biochem* 98: 856-861
29. Altschul, SF, Madden, TL, Schaffer, AA, Zhang, JH, Zhang, Z, Miller, W, Lipman, DJ. 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res* 25: 3389-3402
30. van den Born, E, Bekkelund, A, Moen, MN, Omelchenko, MV, Klungland, A, Falnes, PO. 2009. Bioinformatics and functional analysis define four distinct groups of AlkB DNA-dioxygenases in bacteria. *Nucleic Acids Res* 37: 7124-7136
31. Thompson, JD, Higgins, DG, Gibson, TJ. 1993. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucl. Acids Res* 22: 4673-4680
32. Nair, R, Rost, B. 2005. Mimicking Cellular Sorting Improves Prediction of Subcellular Localization. *J. Mol. Biol.* 348: 85-100
33. Hua, S, Sun, Z. 2010. Support vector machine approach for protein subcellular localization prediction. *Bioinformatics*: in press
34. Bhasin, M, Raghava, GP. 2004. ESLpred: SVM-based method for subcellular localization of eukaryotic proteins using dipeptide composition and PSI-BLAST. *Nucleic Acids Res* 32: W414-419

35. Laemmli, UK. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature (London)* 227: 680-685
36. Hulo, N, Bairoch, A, Bulliard, V, Cerutti, L, De Castro, E, Langendijk-Genevaux, PS, Pagni, M, Sigrist, CJ. 2006. The PROSITE database. *Nucleic Acids Res* 34: D227-230
37. Ryle, MJ, Padmakumar, R, Hausinger, RP. 1999. Stopped-flow kinetic analysis of *Escherichia coli* taurine/ α -ketoglutarate dioxygenase: interactions with α -ketoglutarate, taurine, and oxygen. *Biochemistry* 38: 15278-15286
38. Datsenko, KA, Wanner, BL. 2000. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci U S A* 97: 6640-6645
39. Bauer, AW, Kirby, WM, Sherris, JC, Turck, M. 1966. Antibiotic susceptibility testing by a standardized single disk method. *Am J Clin Pathol* 45: 493-496
40. Dykxhoorn, DM, St. Pierre, R, Linn, T. 1996. A set of compatible *tac* promoter expression vectors. *Gene* 177: 133-136
41. Wang, RF, Kushner, SR. 1991. Construction of versatile low-copy-number vectors for cloning, sequencing and gene expression in *Escherichia coli*. *Gene* 100: 195-199
42. Lizcano, JM, Unzeta, M, Tipton, KF. 2000. A spectrophotometric method for determining the oxidative deamination of methylamine by the amine oxidases. *Anal Biochem* 286: 75-79
43. Hayashi, T, Tsuchiya, H, Todoriki, H, Naruse, H. 1982. High-performance liquid chromatographic determination of alpha-keto acids in human urine and plasma. *Anal. Biochem.* 122: 173-179
44. Koreny, L, Lukes, J, Obornik, M. 2010. Evolution of the haem synthetic pathway in kinetoplastid flagellates: an essential pathway that is not essential after all? *Int J Parasitol* 40: 149-156
45. Richmond, GS, Smith, TK. 2007. A novel phospholipase from *Trypanosoma brucei*. *Mol Microbiol* 63: 1078-1095
46. Molinas, SM, Altabe, SG, Opperdoes, FR, Rider, MH, Michels, PA, Uttaro, AD. 2003. The multifunctional isopropyl alcohol dehydrogenase of *Phytomonas* sp. could be the result of a horizontal gene transfer from a bacterium to the trypanosomatid lineage. *J Biol Chem* 278: 36169-36175

47. Ryle, MJ, Padmakumar, R, Hausinger, RP. 1999. Stopped-flow kinetic analysis of *Escherichia coli* taurine/ α -ketoglutarate dioxygenase: interactions with α -ketoglutarate, taurine, and oxygen. *Biochemistry* 38: 15278-15286
48. Hegg, EL, Whiting, AK, Saari, RE, McCracken, J, Hausinger, RP, Que, L, Jr. 1999. Herbicide-degrading α -keto acid-dependent enzyme TfdA: metal coordination environment and mechanistic insights. *Biochemistry* 38: 16714-16726
49. Muller, TA, Meek, K, Hausinger, RP. 2010. Human AlkB homologue 1 (ABH1) exhibits DNA lyase activity at abasic sites. *DNA Repair (Amst)* 9: 58-65
50. Dinglay, S, Trewick, SC, Lindahl, T, Sedgwick, B. 2000. Defective processing of methylated single-stranded DNA by *E. coli alkB* mutants. *Genes Develop.* 14: 2097-2105

CHAPTER 5

PERSPECTIVES

5.1 INTRODUCTION

Enzymes of the non-heme iron and α -ketoglutarate (α KG) dependent dioxygenase superfamily are ubiquitous in nature and versatile in catalytic activity. Members have been identified in nearly every branch of life and catalyze oxidative reactions ranging from hydroxylations to desaturations, ring formation, or ring expansion to name a few (1). The defining features of this family are the double-stranded β -helix core fold and the requirement for a ferrous ion at the active site. Most members oxidatively decarboxylate α KG to form CO_2 and succinate while creating a highly reactive ferryl-oxo species that then carries out the primary oxidative reaction (1). One branch of phylogeny for which no direct activity for this enzyme family has been shown is the kinetoplastid flagellates including trypanosomes.

Trypanosomes are the source of medical and economic challenges in their endemic areas, causing disease in humans and livestock (2). These parasites have unique anatomy and physiology (see Chapter 1) and complex relationships with their mammalian and insect hosts. They undergo drastic physiological changes during their life cycle and persist in mammalian infection through host immune system evasion involving a long studied, yet poorly understood, process called antigenic variation (3). For all of these reasons, trypanosomes are interesting topics of study. The experiments described in this thesis sought to identify and characterize several putative Fe^{II} and α KG-dependent hydroxylases from trypanosomes in order to increase understanding of these parasites and expand knowledge of this remarkable enzyme family.

5.2 SUMMARY OF RESULTS

5.2.1 *TLP5 and TLP7 are unlikely to participate in pyrimidine salvage*

Two *Trypanosoma brucei* genes were identified based on sequence similarity to thymine 7-hydroxylase from *Rhodotorula glutinis* and termed *TLP5* and *TLP7*. The genes were cloned into *Escherichia coli* and the recombinant proteins expressed, purified by column chromatography, and characterized. Both proteins were soluble as expressed and existed predominantly as monomers in solution. In addition, proper folding and binding of the Fe^{II} and α KG cosubstrates were verified by observation of a characteristic metal-to-ligand charge-transfer (MLCT) feature by UV-visible spectroscopy. However, neither protein bound to the tested nucleic acids or exhibited enzymatic activity as measured by oxygen consumption or succinate production.

Steady state mRNA levels of *TLP5* and *TLP7* were tested by quantitative reverse transcriptase PCR and were found to be several fold higher in the bloodstream form (BF) of the parasite over the procyclic form (PF) for both genes, suggesting that the RNAs are stabilized in this life stage. Epitope tagging of the chromosomal copy of *TLP5* was attempted in order to be able to test the subcellular localization and expression levels of the protein, but this effort failed on six separate attempts.

Protein production was verified in trypanosome cell extracts by western blotting with a polyclonal antibody raised against *TLP5*. The antibodies were shown to be cross reactive against *TLP7*, and both proteins were identified in extracts of both BF and PF parasites. Each cell form also contained a cross reactive band of higher molecular weight suggesting possible posttranslational modification of these proteins in the trypanosome

cell. Pull-down assays were performed to identify possible protein partners of TLP5 and TLP7 in the cell by passing trypanosome cell extracts over column-bound recombinant TLP5 or TLP7 and examining the eluant on a polyacrylamide gel, but no interacting proteins were observed.

In efforts to detect enzymatic activity from these proteins in their natural states, cell extracts were incubated with possible substrates along with cosubstrates and analyzed for product formation by mass spectrometry, but no product formation was observed. In addition, a homology model of TLP5 was created and used in docking studies in an effort to generate hypotheses about the enzyme specificity; although several heterocyclic ring structures that resembled nucleotides were identified as the best ligands, no convincing potential substrates were identified.

5.2.2 *JBP1 is inactive in its recombinant form*

Trypanosomes contain a unique base in their nuclear DNA, β -D-glucosylhydroxymethyldeoxyuracil or base J, which replaces up to 1% of thymine bases (4). Extensive evidence has been obtained that JBP1, a protein shown to bind specifically to J-containing DNA, is responsible along with JBP2 for the introduction of base J into chromatin *in vivo*, though no *in vitro* activity has been reported (5,6). Recombinant JBP1, purified from *E. coli*, was shown to fold properly and bind Fe^{II} and αKG via observation of the MLCT feature; however, no activity was observed when it was tested for *in vitro* activity by monitoring succinate production via high performance liquid chromatography or oxygen consumption via oxygen electrode. While it has been established that JBP1 and JBP2 are critical to the production of base J in trypanosome

DNA, the *in vitro* systems tested were insufficient to observe any activity of these enzymes.

5.2.3 *TbABH* is a DNA-binding protein homologous to *E. coli AlkB*

TbABH is an ortholog of *E. coli alkB*, and was predicted to be a nuclear, DNA-binding protein. Its peptide sequence of the catalytic core aligned well with bacterial AlkBs of group 1A (7) including all metal-coordinating residues and several other highly conserved amino acids known to make contact with the nucleic acid substrate in *E. coli* AlkB (8). *TbABH* was cloned into *E. coli* and the protein expressed and purified. *TbABH* was found to be soluble and in equilibrium between monomer, dimer, and tetramer forms. Proper folding was verified by the formation of the MLCT feature when anaerobic sample was incubated with Fe^{II} and α KG, and titration studies provided a ferrous ion K_d of approximately 4 μ M.

TbABH bound to supercoiled plasmid DNA, as evidenced by electrophoretic mobility shift (EMSA) assays in agarose gels, and to linear alkylated oligonucleotides, as evidenced by EMSA assays in polyacrylamide gels. The latter EMSAs, performed with varying amounts of *TbABH*, allowed determination of a methylated DNA K_d of 7 μ M with near stoichiometric binding to the alkylated portion of the substrate, given the typical low alkylation efficiency. The specific shift observed with methylated DNA was not observed with untreated DNA, and was not due to the creation of abasic sites indicating that *TbABH* binds preferentially to methylation-damaged DNA.

TbABH was tested for *in vitro* activity by monitoring oxygen consumption via oxygen electrode, α KG consumption via the OPDA assay, and formaldehyde production

via the FDH-coupled assay. Unfortunately, none of the approaches provided conclusive results. *TbABH* was also tested for the ability to complement an *alkB* knockout in an *E. coli* cell line stressed with methylmethane sulfonate. The complemented cell line containing pUC-TbABH2 was approximately two fold more resistant to MMS growth inhibition than the knockout cell line containing underivatized pUC18. This complementation along with the DNA binding confirms the assignment of TbABH as a functional AlkB-like protein in *T. brucei* expanding the knowledge of DNA repair mechanisms in this organism.

5.3 CONCLUSIONS AND FUTURE DIRECTIONS

Trypanosomes contain several representatives of the Fe^{II} and α KG-dependent dioxygenase superfamily, and while all the proteins investigated in this study were expressed well in *E. coli* and yielded soluble protein that was properly folded and capable of binding iron and α KG, none of the *in vitro* assays detected enzymatic activity. It may be that trypanosomes contain some stabilizing or activating factor that is missing in our assay conditions. Also, it is possible that these proteins require some type of posttranslational modification for activity, consistent with the results obtained in western blots of trypanosome cell extracts with anti-TLP5 antibodies where prominent cross-reactive bands were observed. On the other hand, I was unable to detect thymine 7-hydroxylase activity using a mass spectrometry assay after incubating thymine with trypanosome cell extracts, perhaps because the protein levels were too low to allow detectable product formation or the substrates tested may have been incorrect.

In order to address these problems, one could attempt to purify the proteins directly from trypanosomes, bypassing the recombinant system entirely, so as to maintain the native form of the protein as translated and modified in the cell. This would pose a new problem of yield, however, since trypanosomes do not grow to a high density and the proteins of interest would not be over expressed. Many culture flasks would be needed to obtain the number of cells required to achieve a comparable amount of protein purified from 1 L of an *E. coli* culture.

Another approach would be to use trypanosome cell extracts as an additive in *in vitro* assays with recombinant protein. This could possibly provide a needed stabilizing

or activating factor, however then the reactions become less defined and any observed activity cannot be directly attributed to the recombinant protein. It would provide a starting point though, and the *in vitro* assay could be refined and optimized once it was known that something in the cell extract could render the protein active.

Finally, one could look into other potential enzyme family members in these amazing parasites. The genome of *Trypanosoma brucei* contains many orthologs of Fe^{II} and α KG-dependent dioxygenases and studying several of these could lead to further understanding of trypanosome biology and new therapeutic targets. As one example, locus Tb09.211.3730 (XP_827514) encodes an ortholog of the oxygen-sensing asparaginyl hydroxylase (FIH) of humans (25% identity over 343 amino acids). Whereas the human enzyme is 349 amino acids, the trypanosomal protein is 1145 amino acids in length and consistent with a multi-domain architecture. The role of FIH in humans is oxygen-sensing via hydroxylation of the hypoxia inducible factor (HIF) (9), so it is reasonable to suspect a similar function of the protozoal protein. In this regard, the bloodstream form of the parasite strictly uses glycolysis for its energy needs while the intra-insect or procyclic form uses oxidative phosphorylation; thus, oxygen sensing plays a critical role in protozoan cellular regulation. It would be interesting to investigate this trypanosomal FIH homolog and determine if it has asparaginyl hydroxylase activity by testing several Asn-containing possible substrates by the methods used previously or a more sensitive approach utilizing a radiolabeled cosubstrate such as 1-[¹⁴C]- α KG.

Another example of potentially interesting proteins to study is a group of highly related sequences arranged in a cluster. Trypanosomes contain large gene arrays within polycistronic units spanning megabase regions of their chromosomes (10). Some genes

whose protein products are needed in abundance, such as tubulin, are known to appear in groups of tandem copies in order to obtain more of the primary transcript. Therefore, the presence of five tandemly repeated genes in *T. brucei* (Tb927.2.6180, 6210, 2130, 6270, and 6310) that encode nearly identical proteins of 319 amino acid residues and are likely to be Fe^{II}/αKG dioxygenases based on sequence comparisons is interesting because it implies that the protein product of these genes is important for the cell and needed in abundance. While there is no close homolog to these sequences to suggest a function for them, it would be interesting to pursue a screening approach to identify possible substrates. This could be accomplished using ¹⁴C-labeled αKG that would release ¹⁴CO₂ upon its oxidative decomposition. Any positive results from those assays would have to be followed up with substrate specific assays to confirm its assignment as a substrate.

Trypanosomes are of great interest because they present many health threats to society and their unique biology offers exceptional challenges to scientists. They will continue to be studied even after effective and affordable cures of their diseases are available because they offer an interesting window into phylogenetic relationships represented in the tree of life. Having branched so early from the eukaryotic lineage, they are very divergent from the more widely studied model organisms for eukaryotic systems, such as yeast (11). They harbor several genes that appear to have been obtained by gene transfer from bacteria and others that seem to be distantly related to plant sequences based on alignments. The Fe^{II}/αKG dioxygenases represent an important enzyme family responsible for many reactions in the cell including several related oxygen sensing functions. Sensing oxygen levels is critical to any organism and especially for trypanosomes as they undergo changes associated with their life cycle which includes

forms exhibiting very different oxygen utilization. I am confident that in time, members of this enzyme family will be found to be involved in the regulation of these processes, expanding their roles across nature.

REFERENCES

1. Hausinger, RP. 2004. Fe(II)/ α -ketoglutarate-dependent hydroxylases and related enzymes. *Crit. Rev. Biochem. Mol. Biol.* 39: 21-68
2. Barrett, MP, Burchmore, RJ, Stich, A, Lazzari, JO, Frasch, AC, Cazzulo, JJ, Krishna, S. 2003. The trypanosomiasis. *Lancet* 362: 1469-1480
3. Morrison, LJ, Marcello, L, McCulloch, R. 2009. Antigenic variation in the African trypanosome: molecular mechanisms and phenotypic complexity. *Cell Microbiol* 11: 1724-1734
4. Borst, P, Sabatini, R. 2008. Base J: discovery, biosynthesis, and possible functions. *Annu Rev Microbiol* 62: 235-251
5. Cliffe, LJ, Kieft, R, Southern, T, Birkeland, SR, Marshall, M, Sweeney, K, Sabatini, R. 2009. JBP1 and JBP2 are two distinct thymidine hydroxylases involved in J biosynthesis in genomic DNA of African trypanosomes. *Nucleic Acids Res*
6. Yu, Z, Genest, PA, ter Riet, B, Sweeney, K, DiPaolo, C, Kieft, R, Christodoulou, E, Perrakis, A, Simmons, JM, Hausinger, RP, van Luenen, HG, Rigden, DJ, Sabatini, R, Borst, P. 2007. The protein that binds to DNA base J in trypanosomatids has features of a thymidine hydroxylase. *Nucleic Acids Res* 35: 2107-2115
7. van den Born, E, Bekkelund, A, Moen, MN, Omelchenko, MV, Klungland, A, Falnes, PO. 2009. Bioinformatics and functional analysis define four distinct groups of AlkB DNA-dioxygenases in bacteria. *Nucleic Acids Res* 37: 7124-7136
8. Yu, B, Edstrom, WC, Hamuro, Y, Weber, PC, Gibney, BR, Hunt, JF. 2006. Crystal structures of catalytic complexes of the oxidative DNA/RNA repair enzyme AlkB. *Nature* 439: 879-884
9. Webb, JD, Coleman, ML, Pugh, CW. 2009. Hypoxia, hypoxia-inducible factors (HIF), HIF hydroxylases and oxygen sensing. *Cell Mol Life Sci* 66: 3539-3554
10. Myler, PJ, Audleman, L, deVos, T, Hixson, G, Kiser, P, Lemley, C, Magness, C, Rickel, E, Sisk, E, Sunkin, S, Swartzell, S, Westlake, T, Bastien, P, Fu, G, Ivens, A, Stuart, K. 1999. *Leishmania major* Friedlin chromosome 1 has an unusual distribution of protein-coding genes. *Proc Natl Acad Sci U S A* 96: 2902-2906
11. Pace, NR. 2009. Mapping the tree of life: progress and prospects. *Microbiol Mol Biol Rev* 73: 565-576

MICHIGAN STATE UNIVERSITY LIBRARIES



3 1293 03063 8773