UNIVERSITY OF SOUTH BOHEMIA, FACULTY OF SCIENCE

DEPARTMENT OF MOLECULAR BIOLOGY AND BIOCHEMISTRY

Characterization of a putative methyltransferase MT420 in Trypanosoma brucei

MASTER THESIS

AUTHOR: **BC. MICHAELA VESELÍKOVÁ**

SUPERVISORS: **RNDR. ALENA ZÍKOVÁ, PHD.; RNDR. SILVIE TRANTÍRKOVÁ, PHD.**

České Budějovice 2010

Annotation

MASTER THESIS

Veselíková, M., 2010: Characterization of a putative methyltransferase MT420 in *Trypanosoma brucei*. Mgr. Thesis, in English - 49 p., Faculty of Sciences, University of South Bohemia, Czech Republic.

ANNOTATION

Localization and characterization of putative mitochondrial methyltransferase acc. No.: Tb10.6k15.0440 in *Trypanosoma brucei* was performed. Employed molecular methods included immunofluorescence, sub-cellular fractionation and tandem affinity purification. Protein was overexpressed in an *E. coli* expression system, using an in-fusion expression vector pPOPINM with maltose binding protein (MBP) tag.

This work was supported by the grant of GAČR number 540/7036.

Prohlašuji, že jsem tuto magisterskou diplomovou práci vypracovala samostatně, jen s použitím citované literatury.

V Českých Budějovicích dne 25. 4. 2010

. . . . . .
Bc. Michaela Veselíková

Acknowledgement

I originally planned to thank everybody in Czech and then I realized that some people that took a share on this project speak English and the rest of them at least understand. First, I will express my thanks to my former supervisor Silva for her confidence that I can manage, and of course, to my current supervisor, Alena, who made me manage. Then I feel a huge gratitude for support and guidance from my current team: Alena, Brian, and Karolina; and broader team led by prof. Lukes, namely (alphabetically by surname): Som, Esteban, Piya, Gabca, Hassan, Milan, Eva, DHL, Lucka, big Zdenek, Pavel, Petr, Tomas, Jirka, Zuzka, and small Zdenek.

I would like to acknowledge Dr. Obornik for significant help with phylogenetic study and creation of the phylogenetic tree, Peter Konik for LC-MS/MS analysis and Brian Panicucci for the correction of English style. Big thanks belong to my family that supported me in all times and by all means, this part in Czech…

Chci poděkovat svojí rodině za neutuchající morální i finanční podporu v časech dobrých i lepších. Tati, díky za to, že jsi šel se svou kůží na trh a zajistil mi tak pohodové studium. Mami, díky za zásoby kosmetiky, bez tebe bych se ani nechtěla vidět. Babi a dědo, díky, za zásoby buchet a rybiček a za to, že jste mi drželi palce a nepochybovali o mně. Dětičky, Aničko a Pepíku, díky za vaši přízeň, za vyhlížení z okna když se vracím a vyprovázení na vlak, když odjíždím. Milí Slaňáci, díky za všemožnou podporu a nikdy nevynechané přáníčko k narozeninám. Nakonec chci poděkovat Martinovi, Méďovi a Jiříkovi, však oni ví za co.

"Nebát se a nekrást"

T. G. M.

Table of Contents

1 General introduction

1.1 Trypanosomatids

Kinetoplastida is a remarkable group of flagellated protozoans that includes the endoparasitic *Trypanosoma, Trypanoplasma* and *Leishmania* species; the free-living bi-flagellates *Bodonida;* and several ectoparasitic species like *Cryptobia*. Most Kinetoplastids are motile protozoa with a single flagellum that originates near their single large mitochondrion and protrudes from the flagellar pocket in the cell membrane. Kinetoplastids derive their name from kinetoplast, the highly structured mitochondrial DNA (1). Kinetoplast consists of a giant network of thousands of concatenated circular DNA molecules, minicircles (0.5 - 10 kb) and maxicircles (20 – 40 kb) (2).

Kinetoplastids are also very interesting as they represent one of the earliest branches of the eukaryotic tree. Since they diverged at such an early point from most higher eukaryotes, they contain several unusual features, including; (i) polycistronic transcription coupled with trans-splicing (3); (ii) unique sub cellular compartmentation of energy metabolism; (iii) unusual mitochondrial DNA architecture; (iv) complex and energy-consuming mitochondrial RNA editing (4); (v) evasion of the host immune response using a variable surface coat (for recent review see (5), (6), (7)). All of these unique attributes of Kinetoplastids are a major focus of basic research as each of these points can be potentially exploited for future disease therapies.

Trypanosomatida is the most studied subgroup of Kinetoplastida because it includes the medically and economically very important parasitic species of genus *Trypanosoma* and *Leishmania.* For example, the species *Trypanosoma brucei* is the causative agent of several serious and neglected tropical diseases in Sub-Saharan Africa. These different diseases are inflicted by various sub-species of *T. brucei*, which are defined by their respective host reservoirs or their geographical location, but they are all transmitted by the tse-tse fly (*Glossina sp.*). *T. brucei brucei* mainly infects the blood of cattle and other domestic and wild animals, causing Nagana, which manifests as the following symptoms: fever, muscular wasting, anemia, swelling of tissues (edema) and eventual paralysis. While *T. brucei brucei* is restricted to non-human mammals, *T. brucei rhodosiense* and *T.brucei gambiense* infect humans in East Africa and West Africa, respectively, causing Human African Trypanosomiases (HAT). Meanwhile, in South America, the parasite *Trypanosoma cruzi*

causes Chagas disease when the infected feces of its insect vector, the reduviid bug, penetrate a sub-sequent bite wound in humans. The acute form of the disease often goes unnoticed; however, the chronic form may develop within the internal organs 10 to 20 years after infection and often leads to heart failure. Furthermore, a sand fly (*Phlebotomus* and *Lutzomya sp*) is responsible for transmission of numerous species of *Leishmania*, which are the causative agents of cutaneous and visceral Leishmaniasis found in many tropical and sub-tropical countries including Americas, Africa, the Middle East and India. These diseases are responsible for the annual deaths and debilitation of hundreds of thousands of people worldwide. Most of the afflicted people live in poverty, which further amplifies the negative effect of these insidious diseases on human lives.

1.2 Human African Trypanosomiasis (HAT)

My project focuses on the basic biology of *Trypanosoma brucei*, the causative agent of HAT. Trypanosomes have developed a complex life cycle (Figure 1) involving an insect vector and a mammalian host. Trypanosomes are transmitted into the mammalian host's tissue by the tse-tse fly as a *metacyclic trypomastigotes.* Then they spread via lymphatic system into the bloodstream and transform into the so-called *bloodstream trypomastigotes.* During this stage, they replicate by binary fission and settle all body fluids like blood, lymph and cerebrospinal fluid. When a tse-tse fly takes a meal of the host's blood, trypanosomes enter the midgut of the fly. There they undergo a transformation into *procyclic trypomastigotes* (PS) before migrating into the salivary glands of the fly where they transform into the dividing *epimastigotes.* Finally, they become the infective *metacyclic trypomastigotes,* during which phase they are able to enter the mammalian host again and thus propagate the disease.

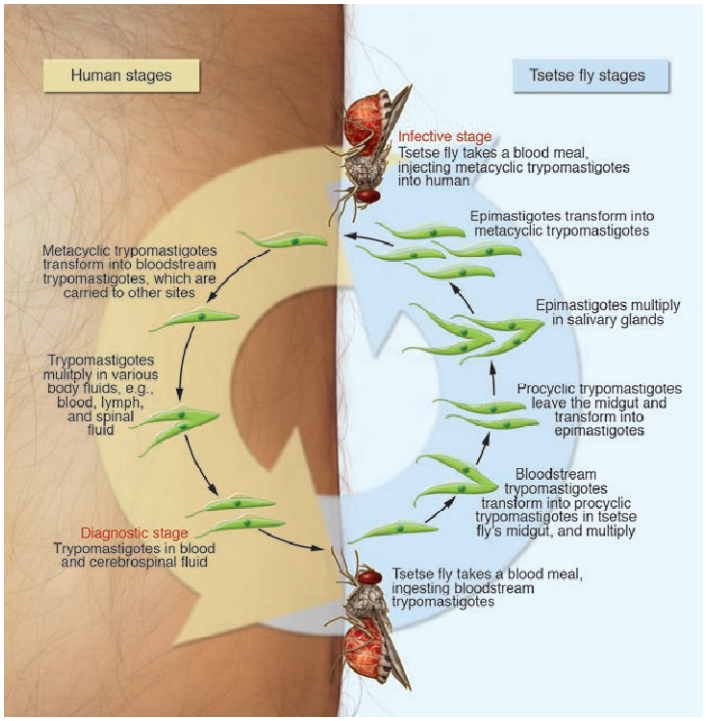

Figure 1, The life cycle of *Trypanosoma brucei.* Adapted from (8)

The symptoms of this disease are often debilitating and without treatment, HAT is invariably fatal. The symptom progression of HAT can be divided into two stages that correlate with the transition of the parasite from the bloodstream through the blood-brain barrier. The first hemolymphatic stage is accompanied by fever, headache and joint pain – symptoms similar to influenza, which often leads to the misdiagnosis of the disease. Rapid parasite growth is at first countered by the host immune response, but the parasite's ability of coat antigenic variations enables it to evade the host's immune system. This results in waves of parasitemia that significantly weakens the host and after a variable period, allows the parasite to cross the blood-brain barrier to invade the central nervous system and produce an array of neurological features. The symptoms of this second phase give the disease its common name (African sleeping sickness). The sleep cycle is disturbed and bouts of fatigue are punctuated with manic periods that progress to daytime slumber and night time insomnia. This progressive mental deterioration eventually leads to coma and death if untreated. The specific pathogenesis of HAT, especially the second stage of the disease, is not very well characterized because there is significant lack of knowledge from a biochemical and molecular point of view.

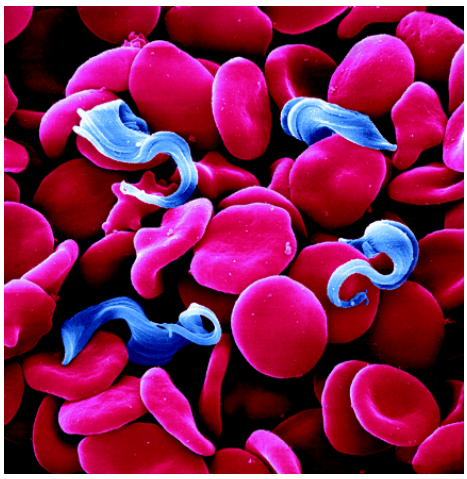

Figure 2, Parasite, *Trypanosoma brucei* surrounded by red blood cells in a smear of infected blood. Adapted from **(53)**.

Current diagnostic tools are not completely satisfactory, as the majority of them usually are not able to identify which subspecies of *T. brucei* has infected the patient. Even after 20 years of searching for molecular markers, the only distinguishing factor is the presence of the SRA (human serum resistance –associated) gene found only in *T. brucei rhodesiense* genome. Therefore, reliable methods of discrimination are not usually applicable in the field because either they require extensive diagnostic tools or they are expensive. Furthermore, detection of late stage HAT is even more demanding because it requires the analysis of a patient's cerebrospinal fluid, which is very difficult to obtain out in field conditions. Poor diagnostics are in case of this disease combined with limited drug effectivity.

HAT is a significant threat to developing countries as a half a billion people, primarily in subtropical and tropical areas, live at risk of the infection by this disease. It is estimated that

more than 20 million individuals are infected with the pathogen, resulting in extensive suffering and more than 100 000 deaths per year (8). Major flare-ups of the disease occurred in 1990s in Angola, The Democratic Republic of the Congo, Uganda and Sudan; however, there were still 17600 cases reported in 2004. Common factors leading to the disease outbreak are still prevalent in these countries, particularly ineffective or unstable governments, poverty and poor infrastructure.

While new treatments of HAT are imperative for all of these at-risk populations, the development of a HAT vaccine is highly unlikely because these parasites are able to change their surface antigens and thus evade the host's immune system. The ability to switch the gene expression of thousands of the VSGs (variable surface glycoprotein), encoded throughout the whole Trypanosoma genome, forms one of the most elegant as well as effective evasive systems known in biology (8).

Therefore, there is a great need for new and inexpensive drugs that can be implemented in field conditions outside of the sporadic urban hospitals. Treatment of HAT currently relies on drugs that have undesirable side effects and/or problems with efficiency, administration and compliance (9). First stage drugs, suramin and pentamidine (Sanofi-Aventis), are efficient with moderate adverse effects; however, they must be administered parenterally. Two drugs are approved for treatment of the second stage of HAT. The arsenical melarsoprol (Sanofi-Aventis) is administered intravenously and causes severe adverse reactions that can be life threatening. The second drug eflornithin (Sanofi-Aventis) is effective only against *T. brucei gambiense* and is used in combination with nifurtimox, which is a drug with limited efficiency when used alone. Furthermore, the number of treatment failures following the administration of drugs is increasing due to the growing drug resistance among the *T. brucei* population.

1.3 Trypanosoma brucei brucei as a model organism

The term “model organism” is narrowly applied to those species that facilitate experimental research and can bring us essential information for further development of drugs, diagnostic techniques, or new treatment approaches. Model organisms are usually chosen because of their properties such as short generation time, easy cultivation, safe manipulation and relatively simple experimental design. Examples of such species can be found throughout the whole tree of life, for example *Escherichia coli* (bacteria), *Saccharomyces cerevisiae* (fungi), *Caenorhabditis elegans* (roundworm), *Drosophila melanogaster* (fruit fly), and

Arabidopsis thaliana (plant). Every model organism is ideally equipped with an extensive toolkit consisting of molecular, genetic, and biochemical methods of study. *Trypanosoma brucei brucei* has been one of the most studied protozoan organisms in last 50 years. During this period, it compiled a considerable repertoire of experimental tools (in Table 1) that qualify it as a model organism.

Table 1. A summary of genetic tools in *T. brucei*; (+ - possible method, ++ - frequently used method, +++ - method of choice) **(10)**.

Genetic tool	Level of establishment
In vitro culture models	++
In vivo disease models	++
Transient transfection	+++
Stable transfection	+++
Expression vectors:	
Episomal	+
Integrating	+++
Regulatable	+++
Selectable markers:	
Positive	6
Negative	1
Gene knockouts	+++
Sexual crossing	+
Positional cloning	Possible
RNAi	+++
Functional rescue	+
Transpozon mutagenesis	+

Various reverse genetic tools are available for *T. brucei* including well-established methods for DNA manipulation *in vivo*, stable transfection and an array of expression vectors that allow the expression of heterologous proteins. Thus, *T. brucei* represents a powerful system for the characterization of novel proteins with unknown function. In addition, unlike the Leishmania species (10), *T. brucei* possesses all the protein components necessary for RNA interference (RNAi). Furthermore, it is possible to generate gene knock-outs (double knock-out and conditional double knock-outs) in *T. brucei* to determine the function of a particular protein of interest. Compared to all of these established reverse genetic methods, there have been fewer examples of forward genetics used in the field of Trypanosomes. However, one excellent example of mutagenesis used in the past involves the generation of Trypanosomes lacking their kinetoplast DNA when treated with high levels of ethidium bromide and acriflavine (11). When forward genetics are applied experimentally and mutants are available, genetic rescues can be achieved by the transfection of wild-type DNA libraries.

The *T. brucei* toolkit grows even more considerably with the fact that the genome of this species was sequenced in 2005 and the whole genome database is accessible through the TriTrypDB (www.tritryp.org) (12). This web page represents an integrated and functional genomic database for pathogens of the Trypanosomatida family, including the fully sequenced

species of *Leishmania major* and *Trypanosoma cruzi* (13), (14). Since the RNAi is the one of the most widely used manipulation tools in *T. brucei,* we will take a closer look at its methodology.

2 Scientific background

2.1 RNA Interference

One of the most widely used manipulations in the field of Trypanosomes is the method of RNAi to determine the function of a protein of interest. This powerful technique was first observed in *Caenorhabditis elegans* in 1998 (15), when researchers were using it to assess the functional importance of a particular gene in the development of the organism (16). Since this discovery, much has been learned about the molecular mechanism of RNAi and it is now routinely used to quickly and specifically knock-down the mRNA levels of a specific gene in many organisms.

To initiate the process of RNAi in a cell, you must first over-express a dsRNA region of your gene of interest. There are many tet-inducible vectors available for this in trypanosomes, utilizing either head-to-head T7 promoters or stem loop vectors that contain a stuffer region between two inverted repeats of a gene. Once the dsRNA is overexpressed by adding Tetracycline to your media the cellular enzyme Dicer binds to the dsRNA and cleaves it into short pieces of ~ 20 nucleotides in length, which are known as small interfering RNA (siRNA). These siRNAs then interact with the RNA-induced silencing complex (RISC), which binds a single strand of the siRNA to the complementary sequence of a single stranded mRNA molecule. The nuclease activity of

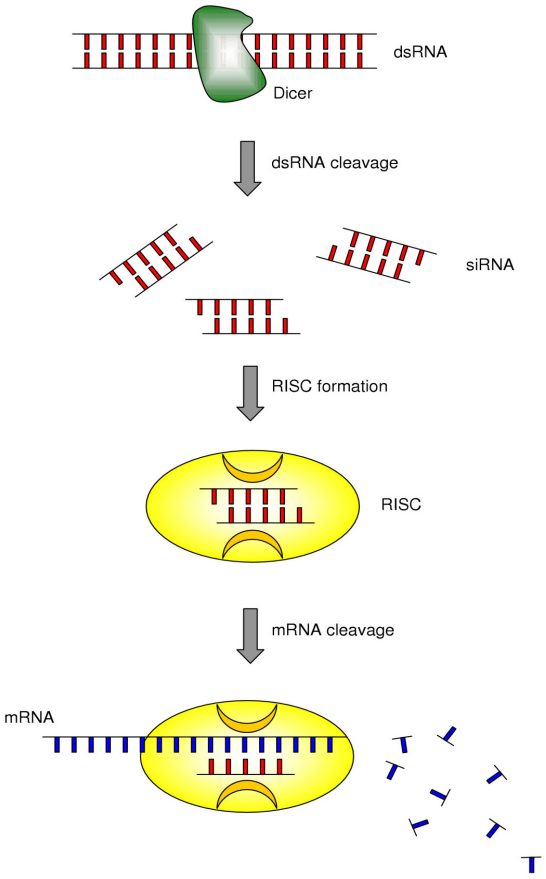

Figure 3, Mechanism of RNA interference (RNAi). RISC – RNA-induced silencing complex;dsRNA – double stranded RNA, siRNA – small interferring RNA. Adapted from **(54)**.

RISC then degrades the mRNA, thus silencing the expression of the corresponding gene (scheme in Figure 3). Since it was discovered that *T. brucei* contains all the components necessary for RNAi, this methodology has become the method of choice to quickly screen for the function of unknown gene products.

2.2 Transcription and mRNA splicing

As mentioned previously, Trypanosomes diverged quite early from higher eukaryotes and thus retain some properties that more resemble their prokaryotic ancestors. For example, Trypanosomatid genomes are organized into long arrays of open reading frames, which are transcribed into long, polycistronic pre-mRNAs. While this eliminates most of the need for the tight transcription regulation, we see in higher eukaryotes, it is not exactly homologous to the highly ordered operon organization found in bacteria as each polycistronic pre-mRNA in *T. brucei* contains some genes from very different metabolic pathways. Moreover, the polycistronic pre-mRNA is further processed by coupled *trans*-splicing and polyadenylation reactions to generate mature monocistronic mRNA. In *T. brucei, trans*-splicing entails the transfer of the 39 bp long sequence of spliced leader RNA (SL-RNA) to the 5' end of all mRNAs (in Figure 4) (17). The SL-RNA sequence is ubiquitously found and is highly conserved among Trypanosomatids. Since the trans-spliced SL sequence also provides the cap structure, trans-splicing can be also viewed as a capping reaction. The cap structure of Trypanosomatid mRNA is unique in the eukaryotic kingdom with regard to the extent of modified nucleotides.

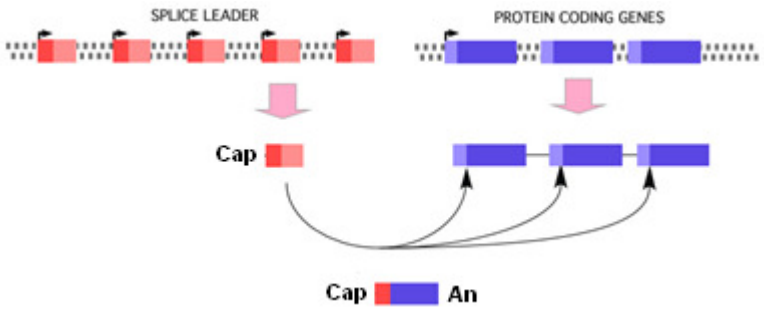

Figure 4, The general scheme of SL-RNA, pre-mRNA generation and trans-splicing. Cap – capped 5' end; An – poly adenine 3' end. Adapted from **(53)**.

mRNA capping, as well as the mRNA polyadenylation, is a vital process for the formation of a mature mRNA molecule that is subsequently translated. Similarly to eukaryotic mRNA transcript, *T. brucei* pre-mRNA is capped on its 5' end by the addition of a 7-methylguanosine (m7G) capped SL RNA. This modification of the SL RNA is followed by further methylations of the first few nucleotides. According to the type and amount of methylated nucleotides, we can distinguish four types of cap structures. All trypanosomatid SL RNA possess a cap

structure known as cap 4, the most complex cap found in nature. The cap 4 structure is unique in the eukaryotic kingdom in terms of its high content of modified nucleotides – three nucleotides following the m7G are methylated and these four nucleotides combined carry seven methyl groups in total. The specific methylation pattern of this cap 4 structure is represented in the following scheme [m^7**G**-ppp-N^6,N^6,2'-O-trimethyl**A**-p-2'-O-methyl**C**-p-N^3,2'-O-dimethyl**U**] (18), where the first nucleotide is methylguanosine, followed by an adenine that is methylated at three distinct positions, succeeded by a methylcytosine, and terminating with a uracil methylated at two positions.

One possible role of the Kinetoplastid cap 4 involves the recognition and assembly of components of the trans-spliceosome, which is a complex of proteins and specialized RNA that cleaves the polycistronic pre-mRNA into monocistronic mRNAs and ligates them to the SL (19). Another possible role may be the recruitment of the small and large subunits of the ribosome to the mRNA to initiate translation (20). No matter which function the methylation directs, it has been shown that some of the methyl groups of this cap 4 structure are essential for *T. brucei*.

2.3 Methylation

Methyltransferases (MTases) are abundant enzymes that catalyze the addition of a methyl group to a number of various substrates that include DNA, RNA and proteins (in Figure 5). Methylation is involved in the regulation of gene expression by modifying DNA and histones, as well as the metabolism of RNA as the mRNA cap acts to stabilize a transcript. In eukaryotes, gene expression is often regulated by the changes that do not alter the primary sequence but rather modify the chemical structure of either the DNA or the histones that interact with the DNA. These epigenetic modifications are not transient, but persist throughout the cell divisions and can regulate tissue specific gene expression (21).

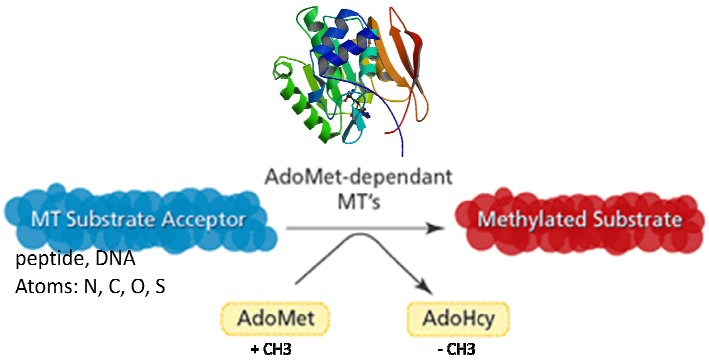

Figure 5, General scheme of methylation by S-adenosyl methionine dependent MTase.

DNA methylation is a crucial part of the normal development and cellular differentiation in higher organisms as the DNA methylation alters the gene expression pattern in cells with the transfer of a methyl group from S-adenosylmethionine (SAM) to the pyrimidine ring of cytosine or adenine in a DNA molecule. DNA methylation is a very common modification observed in most of the organisms throughout the tree of life; for example *Escherichia coli*, *Oryza sativa* and *Homo sapiens*. All DNA MTases share distinct conserved features in their primary protein sequence. Specifically the N-6 Adenine-specific DNA MTases signature motif (PROSITE accession number PS00092). Since this domain is highly conserved, it carries potential to predict homologous DNA MTases from proteins with unknown function in other species (22).

Besides DNA MTases, there are also several mRNA cap MTases that have been characterized to date. The first was the Vaccinia virus MTase VP39, a protein responsible for methylation of the 5' mRNA cap at 2'-O- position (23). In 2007, another cap MTase was reported. TbMTr1was found in *T. brucei* and it is responsible for the 2'-O methylation of a cap 1 structure, which is a transitional state of the process of cap 4 formation (24). *In-silico* homology searches of the Trypanosomatid genome database revealed two hypothetical proteins with significant homology to the *Vaccinia* virus MTase VP39 (23), (25). Both of them were examined, independently, by two groups who showed that while these MTases are not essential for *T. brucei* growth, their depletion does result in a defected cap 4 structure, specifically at the 2'-O methylation position (20).

One of the proteins suspected in the methylation of cap 4 is also the protein of interest (Tb10.6k15.0440, MT420) for my Master thesis. However, further computational analysis of MT420 showed a significant homology to protein N5-glutamine MTases and the HemK MTase family.

2.4 PrmC/HemK methyltransferase

The PrmC/HemK MTase plays an important role in the translation machinery in *E. coli* (26). The prokaryotic translation is a well studied process comprising of the ribosome assembly on an mRNA strand, translation initiation by a Met-tRNA molecule with the help of initiation factors (IF), elongation of the newly synthesized peptide, and the termination facilitated by the binding of release factors (RFs). Bacterial RFs recognize the stop codon in an RNA sequence and are able to release the complete polypeptide and disassemble the

ribosome. While there are three classes of RFs that recognize different stop codons, class I RFs in prokaryotes recognizes the codons UAA and UAG. Class I RFs are methylated on the glutamine residue of a conserved tripeptide motif "Gly Gly Glu" (GGQ). This methylated amino acid then interacts with the peptidyl transferase center of the large ribosomal subunit, triggering hydrolysis of the ester bond in peptidyl-tRNA and releasing the newly synthesized polypeptide from the ribosome (26).

The MTase responsible for methylation of RF1 in *E. coli* is S-adenosylmethionine (SAM)-dependent MTase PrmC/HemK discovered in *E. coli* and at least two times misidentified as a component of the heme biosynthetic pathway (27). It has been shown that PrmC/HemK mediates the methylation of RFs and thus has an important stimulatory effect on the *in-vitro* release activity of the *E. coli* (28). The inactivation of *PrmC/HemK* in *E. coli* K-12 leads to an increase in the read-through of specific stop codons, an induction of the oxidative stress response, and severely retarded growth indicating the importance of this posttranslational methylation event (29). The discovery that PrmC/HemK actually methylates peptides was unexpected because PrmC/HemK contains the hallmark sequence motif thought to be specific to N^6-adenine DNA MTases and none of the DNA MTases has been previously reported to methylate proteins. Structures of PrmC/HemK from *E. coli* (Figure 6) as well as from *T. maritima* were recently published (30), (31).

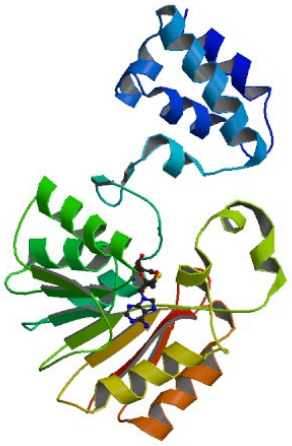

Figure 6, the structure of PrmC/HemK; the SAM- dependent protein N5-glutamine MTase crystallized with SAM, *E. coli*. Adapted from **(30)**.

3 Aims

In this project, we focused on the determination of the fundamental properties concerning a putative MTase (GeneDB ID: Tb10.6k15.0440; Tb927.10.9860; called MT420). We used tandem affinity purification (TAP) tagged MT420 to determine its localization within the cell and to elucidate its potential binding partners. We also created an MT420-RNAi silenced cell line to examine if the protein is essential for the growth of procyclic stage (PS) *T. brucei* cells. Finally, we over-expressed and affinity-purified a recombinant maltose binding protein (MBP) tagged MT420 for future *in vitro* methylation activity assays.

Specific aims of this project are:

- To perform computational sequence analysis of the MT420 protein
- To over-express the TAP-tagged MT420 in *T. brucei* PS cell lines
- To localize MT420 using a TAP-tagged PS *T. brucei* cell line by means of
    - In-situ Immunofluorencence antibody analysis (IFA)
    - Cell compartment fractionation
- To TAP-tag purify MT420 from the PS *T. brucei* cells to identify any of its potential binding partners by LC-MS/MS analysis
- o generate an RNAi cell line to knock-down the expression of MT420 in *T. brucei*
- To over-express and purify MBP-tagged MT420 in *E. coli*

4 Materials and Methods

4.1 Cloning, transfection and generation of the transgenic T. brucei

For the preparation of DNA constructs, standard procedures including the amplification of DNA using synthetic oligonucleotides, ligation and restriction analysis were used.

Genes of interest were amplified by Polymerase chain reaction (PCR) employing rTaq Polymerase (TopBio) and oligonucleotides listed in Table 2. Amplified products were inserted into the pGEM-T easy Vector as recommended by the manufacturer (Promega).

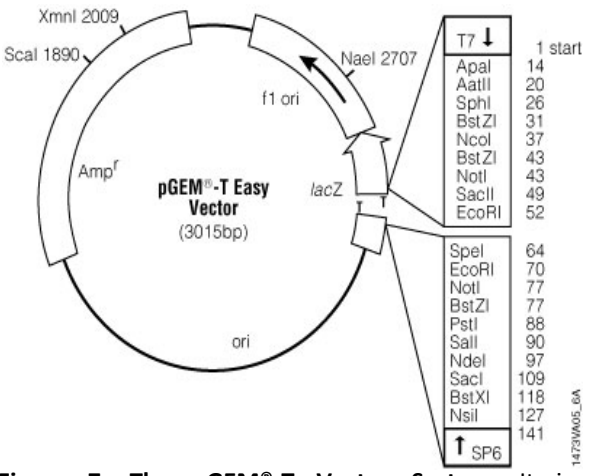

Figure 7, The pGEM®-T Vector System. It is a convenient system for the cloning of PCR products. Vector map adapted from official Promega materials.

Table 2. List of used oligonucleotides, restriction endonuclease sites underlined.

Oligonucleotide name	Sequence 5' -> 3'	Restriction site
RNAi Fw	CAC AAG CTT ACG AGA ACA TTG G	HindIII
RNAi Rv	CAC GGA TCC TTT GGT ACC AC	BamHI
MTtap Fw	AAG CTT ATG CGT CGA TGG CTG	HindIII
MTtap Rv	GGA TCC GTG GGA GGA AGA GG	BamHI

The heat-shock transformation protocol was used with competent bacterial cell strain, *E. coli XL-1 DH5α,* for the cloning of desired vectors. Positive colonies were selected by Blue/White selection and plasmid DNA was isolated using a QIAprep Spin Miniprep Kit (QIAGEN) according to the manufacturer's protocol. The concentration of a DNA sample was measured using the NanoDrop Spectrophotometer (Thermo Scientific) as described by the manufacturer's protocol.

The restriction endonuclease reactions were performed as recommended by the manufacturer (New England BioLabs) to screen clones for the proper insert. The digested

plasmids were run on a 1% DNA agarose gel and the PCR inserts were purified from the gel using QIAquick Gel extraction Kit (QIAGEN). Purified inserts were then cloned into the final vector.

4.1.1 Transformation

Approximately 50 ng of plasmid DNA or 5 µl of the ligation mixture were used for transformation of *E. coli XL-1 DH5α*. DNA and cells were mixed and incubated on ice for 30 minutes. The cells were heat-shocked at 42°C for 45 seconds and then incubated for 2 minutes on ice. SOC medium (2% peptone, 0.5% yeast extract, 0.05% NaCl, 0.25 mM potassium chloride, 10 mM magnesium chloride, 0.04% glucose, 0.02 mM magnesium sulphate, pH 7) was added and cells were incubated for another 1 hour at 37°C, shaking. Then the cells were spread on LB agar plate (1% NaCl, 1% peptone, 0.5% yeast extract, 1.5% agar) containing ampicillin at a concentration of 100 µg/ml. Finally, the plates were incubated at 37°C for 18 hours or over-night. Positive clones were further analyzed.

4.1.2 TAP tagged MT420 construct

To create the vectors for the inducible expression of a C-terminally tagged MT420, its whole ORF was PCR amplified from *T. brucei* strain 29-13 genomic DNA using the MTtap Fw and MTtap Rv oligonucleotides (Table 2). The PCR product was cloned into pGEM-T easy vector (in Figure 7), digested with BamHI and HindIII enzymes and ligated into the pLEW79-MHT vector (Figure 8). The resulting construct pLEW79MHTAP-MT420 was verified by restriction analysis and by sequencing.

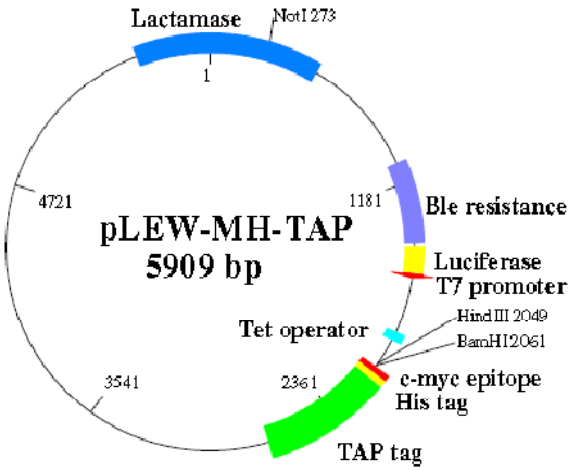

Figure 8, pLEW79-MH-TAP vector used for generation of TAP tagged MT420. Vector map assembled in BioEdit freeware (Tom Hall, Ibis Therapeutics).

4.1.3 RNAi construct

To create the RNAi construct for MT420 repression, part of the *MT420* gene was amplified by PCR using RNAi Fw and RNAi Rev oligonucleotides (Table 2). The PCR product was cloned into the pGEM-T easy vector (Figure 7), digested with BamHI and HindIII enzymes, and ligated into the p2T7-177 vector (Figure 9). The resulting construct (Figure 10) p2T7-177-MT420 was verified by sequencing.

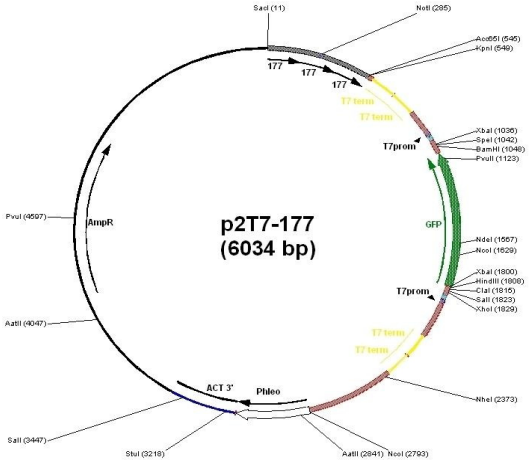

Figure 9, Original p2T7-177 vector map. It contains two T7 promoters. Vector map kindly provided by RNDr. Alena Zíková, PhD.

Figure 10, Scheme of MT420-RNAi construct. This construct was created using RNAi Fw and Rv primers, it consists of 152 bp of MT420 5'UTR and 370 bp of MT420 ORF, creating the 522 bp long construct.

4.1.4 Transfection of T. brucei

Introduction of alien DNA into an organism is a method of DNA manipulation that every model organism has to be able to employ. Due to homologous recombination, *Trypanosoma brucei* is able to incorporate ectopic DNA flanked by homologous *T. brucei* sequences that direct the integration into a predetermined location in its genome. Electroporation of linearized DNA from a suitable vector is generally used for this purpose. We used this technique in order to introduce several different constructs into the *T. brucei* genome, namely the p2T7 177 MT420 vector and the pLEW79-MH-TAP plasmid containing the MT420 ORF.

The 29-13 strain is a transgenic *T. brucei* PS cell strain that contains the ectopic T7 RNA polymerase and tetracycline (tet) repressor. These cells were grown *in vitro* at 27°C in SDM-79 media containing hemin (7.5 mg/ml) and 10% FBS (32). The TAP-plasmid and RNAi plasmid (around 20ug of DNA each) were linearized with NotI enzyme over-night. The linearized plasmids were then precipitated by three volumes of 95% ice-cold ethanol and 1/10

volume of 3M NaOAC (pH 5.4), washed with 70% ethanol and resuspended in sterile MilliQ water. The resulting DNA concentration was measured using the Nanodrop.

Approximately $2x10^7$ *T. brucei* 29-13 PS cells were centrifuged, washed once in phosphate-buffered saline (1xPBS) and then resuspended in 1 ml of cold Cytomix buffer (25 mM HEPES pH 7.6; 120 mM KCl; 0.15 mM $CaCl_2$; 10 mM K_2HPO_4/KH_2PO_4 pH 7.6; 2 mM EDTA; 6 mM glucose; 5 mM $MgCl_2$). 500 µl of cells and 10 – 15 µg of DNA were loaded into the 0.2 cm gap electroporation cuvette (Electroporation Cuvettes Plus Model no. 620, BTX) and electroporated with two pulses (1st pulse 1500 V, 25 Ω, 50 µF; 10 s pause; 2nd pulse 1700 V, 25 Ω, 50 µF) using the ECM630 PrecisionPulse™, BTX-Genetronics, Inc. electroporator. The cells were immediately transferred to a 25 cm^2 cultivating flask containing SDM79 medium without selection antibiotics. After 12 hours, the selectable antibiotic (phleomycin) was added at a concentration of 1 µg/ml and the cells were serially diluted into a 24-well plate.

Synthesis of dsRNAi was induced by the addition of tet at 1µg/ml concentration. The cells were counted using the Z2 Cell Counter (Beckman Coulter Inc.) and growth curves were generated for the RNAi cell lines over a period of 7 days. For TAP-tagged cell lines, the expression of the tagged protein was also induced by 1ug/ml of tetracycline.

4.2 Tandem affinity purification and mass spectrometry analysis

Tandem affinity purification is a highly specific two-step purification method that has, thanks to the mild conditions during the process, the potential to identify possible binding partners for tagged proteins (33). The TAP tag consists of three parts: the Calmodulin binding protein (CBP), the TEV protease cleavage site and the protein A tag that has affinity to IgG (Figure 11). In addition, the T. brucei pLEW79MHTAP vector contains a *c*-myc and His_6-tag.

The *T. brucei* TAP_MT420 cells were induced by tetracycline and ~10^{10} cells were harvested after 48 hours. Washed cells were resuspended in IPP-150 buffer (10 mM Tris-HCl pH 8, 150 mM NaCl, 0.1% Nonidet P40) containing one tablet of a protease inhibitor cocktail (Roche) per 10 ml of buffer. The cells were then lysed with 0.25% Nonidet P40 (NP-40, Roche), incubated for 20 minutes at 4°C and cleared by low speed centrifugation. The resulting supernatant was further lysed with 1.25% NP-40, incubated on ice for another 30 minutes and finally cleared by centrifugation (45000 rpm at 4°C in a Sorvall SW-55 rotor for 45 minutes).

In the meantime, IgG beads were prewashed in a 20 ml chromatography column (BioRad) with IPP-150 buffer containing 0.5% NP40. The cleared lysate was then incubated with the washed IgG beads for 2 hours at 4°C. After that the beads were washed three times with IPP-150 buffer and once with TEV cleavage buffer (TEVCB: 10 mM Tris-HCl pH 8, 150 mM NaCl, 0.5% NP40, 0.5 mM EDTA, 1 mM DTT). The tagged complexes were cleaved from the column by TEV protease (2 hours, 16°C) and eluted into a clean tube. Furthermore, the column was washed with TEVCB to recover the dead volume of the column, which was collected into the same tube as the eluate.

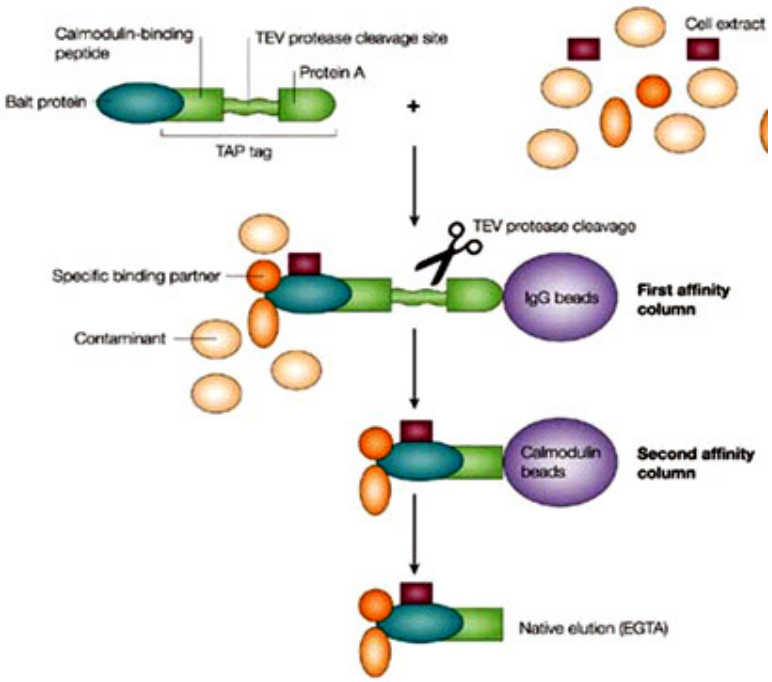

Figure 11, Tandem affinity purification scheme. Scheme adapted from Nature Reviews.

The TEV eluate was diluted with Calmodulin Binding Buffer (CBP: 10 mM Tris-HCl pH 8, 150 mM NaCl, 0.5% NP40, 1 mM MgOAc, 1 mM imidazole, 2 mM $CaCl_2$, and 10 mM β-mercaptoethanol) and 3 volumes of 0.1 M $CaCl_2$. This mixture was combined with calmodulin beads (Stratagene) that were prewashed with CBB buffer into a 10 ml chromatography column (BioRad) and incubated over-night at 4°C. The next day the beads were washed with 5 ml of CBB buffer five times before the tagged complexes were eluted with Calmodulin Elution Buffer (10 mM Tris-HCl pH 8, 150 mM NaCl, 0.5% NP40, 1 mM magnesium acetate, 1 mM imidazole, 2 mM EGTA, 10 mM β-mercaptoethanol), collecting each fraction into a separate tube.

The protein samples from each purification step were fractionated by SDS-PAGE, blotted onto a PVDF membrane, probed with anti-His_6 mAb (1:2000, Invitrogen), and developed using the ECL system (Roche). Elution fractions were stained with SYPRO Ruby gel staining and analyzed by LC-MS/MS.

4.3 SYPRO Ruby gel staining

Polyacrylamide gels prepared for SDS-PAGE electrophoresis can be analyzed using staining with variously sensitive dyes. One of the most sensitive dyes is the SYPRO Ruby stain (Molecular Probes). Moreover, this staining procedure is compatible with further analysis by LC-MS/MS.

After electrophoresis, the gel was placed into strictly very clean Petri dish of a suitable size and then fixed twice in fixing solution (50% methanol, 7% acetic acid), slowly agitating at room temperature for 20 minutes. The protein was stained with SYPRO Ruby stain (20 times the gel volume) over-night, washed in a solution containing 10% methanol and 7% acetic acid and visualized on a UV transluminator equipped with CCD camera.

4.4 Preparation of samples for Mass Spectrometry analysis

Here we use mass spectrometry to detect tryptic peptides, which can then identify proteins in a gel slice or in a mixed protein solution. Trypsin is a pancreatic serine endoprotease that hydrolyzes peptide bonds specifically at the carboxyl side of arginine and lysine. It is routinely used in proteomics for peptide mapping and protein sequencing due to the high frequency of arginine and lysine residues in most proteins, thus providing enough peptides of a detectable size to get good coverage of the proteins in a sample.

4.4.1 In-gel digest

In-gel digestion is used to analyze protein bands from SDS-PAGE gels. The visible protein bands were cut out using a clean razor blade, transferred to a clear glass plate and cut into pieces – app. 1 mm. Next, the gel pieces were transferred to a new tube, dehydrated twice, with acetonitrile (nanopure acetonitrile for LC/MS, Fisher Scientific) for 10 minutes, and then dried by speed vac. The gel pieces were then rehydrated in 50 mM NH_4HCO_3 (Sigma-Aldrich) containing 12.5 μg/μl of trypsin (proteomics grade, Sigma-Aldrich). The gel pieces were covered with a minimal volume of the trypsin-containing buffer and the samples were incubated for 45 minutes on ice. An additional 10 μl of the trypsin-containing buffer was added and the samples were left at 37°C over-night. On the second day, the gel pieces were centrifuged for 5 minutes at maximum speed and the supernatant containing the tryptic peptides was collected into a new tube. Peptides were further extracted with one change of 20 mM NH_4HCO_3, followed by two changes of a solution containing 5% formic acid (Formic acid optima LC/MS, Fisher Scientific) and 50% acetonitrile, incubating each wash step for 10

minutes at room temperature. The final sample was then centrifuged (maximal speed for 5 minutes) and the supernatant was collected using PAGE-loading tips and deposited into the same tube containing the first. The collected liquid was evaporated in a speedvac for 2 hours and the remaining pellet containing the tryptic peptides was resuspended in 0.1% formic acid and analyzed by mass spectrometer Q-Tof Premiere MS/MS (Waters).

4.5 Protein immunoblot – Western blot

The western blot is a protein analysis method employing specific antibodies paired with a secondary antibody conjugated with horseradish peroxidase, which works as a reporter enzyme. First, the proteins are transferred from a PAGE gel onto a PVDF (Immobilon-P PVDF Transfer Membrane, Millipore) membrane that acts as a durable structure to stably display immobilized proteins that can be analyzed with various antibodies.

The PVDF membrane is activated in methanol for 40 seconds, rinsed in MilliQ water and then equilibrated in transfer buffer (39 mM glycine, 48 mM TRIS, 20% methanol) for 5 minutes. The supporting Whatman papers and the SDS-PAGE gel itself are also allowed to equilibrate with the transfer buffer for 5 minutes. Blots are assembled according to the scheme at the right (in Figure 12), submerged into cooled transfer buffer and then run in a Cleaver Scientific Wet Transfer aperture for 90 minutes at 90 V.

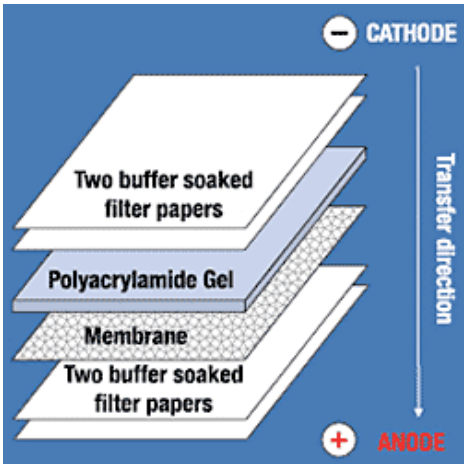

Figure 12, Scheme showing the assembly of Western blot.
Adapted from en.wikipedia.org.

After the transfer, the PVDF membrane was blocked over-night at 4°C in the blocking solution 1xPBS-T/milk (phosphate buffered saline – PBS, 0.05% Tween, 5% skim milk). After washing, the blot is incubated for 1 hour on a shaker with a specific primary antibody dissolved in 1xPBS-T/milk. After the incubation, any excess primary antibody is removed by washing the membrane again with 1xPBS-T. The membrane is then incubated with a secondary antibody (conjugated with Horseradish peroxidase/HRP) for 1 hour on shaker at room temperature. The membrane is washed again and then incubated with the substrate for HRP (Pierce®ECL Western Blotting Substrate, Thermo Scientific) for 1 minute. The resulting signal was detected using autoradiograph films exposed to the membrane for various amount

of time (Kodak BioMax Maximum Sensitivity film, 18x24 cm, Sigma-Aldrich). The exposed film was processed using two different chemical solutions (GBX developer and GBX fixer, Sigma-Aldrich) and water.

4.6 Sedimentation of MT420-TAP in glycerol gradient

When complex protein mixtures are spun at a high speed in a linear glycerol gradient, native proteins and complexes will sediment in the gradient based on their buoyancy and density. This method allows us to fractionate proteins and protein complexes according to their mass under native conditions, which can be used to analyze what other proteins sediment at the same value and therefore might be associated with a protein of interest. This method was performed using mild lysis conditions, protease inhibitors, and ultracentrifugation in linear glycerol density gradient to analyze the sedimentation properties of the MT420-TAP.

A linear gradient from 10 - 30% glycerol was prepared using a gradient maker (Gradient Mixer, Cleaver Scientific) to mix together buffers consisting of 10% and 30% glycerol (10 mM Tris-HCl pH 7.2, 10 mM $MgCl_2$, 100 mM KCl, 10% and 30% glycerol, respectively). Furthermore, 1 mM DTT and protease inhibitors were added to the gradients, which once prepared were kept on ice.

A 400 ml culture *T. brucei* 29-13 MT420-TAP cells induced for 48 hours were harvested by centrifugation at 4°C for 10 minutes at 1300 x g. Cells were washed once in 1xPBS, resuspended in Lysis buffer (10 mM Tris-HCl pH 7.2, 10 mM $MgCl_2$, 200 mM KCl, protease inhibitors) and then lyzed using 1% Triton X-100. The lysate was incubated for 30 minutes on ice and then centrifuged for 30 minutes at maximum speed at 4°C. The resulting supernatant was then transferred to a new tube and spun down again for 15 minutes. This cleared lysate was loaded onto the surface of the gradient in the ultracentrifuge tube.

The samples were spun down in a Sorvall SW-40 rotor (L8-M Ultracentrifuge, Beckman) for 12 hours, at 38,000 rpm. 500 µl fractions were sequentially collected from the surface of the gradient using a cut pipette tip. These fractions were all flash-frozen in liquid nitrogen and then stored at -80°C. Odd fractions were subsequently analyzed by immunoblot and subjected to immunoprecipitation.

4.7 Immunoprecipitation

To analyze further the glycerol gradient fractions, we employed immunoprecipitation with magnetic beads bound with IgG molecules pre-bound on their surface. Since the immunoblot signal corresponding to the TAP-tagged MT420 was detected throughout the whole glycerol gradient (from 10S up to 50S), immunoprecipitation analysis was used on various pooled glycerol gradient fractions in order to determine different possible binding partners of MT420. The goat anti-rabbit IgG magnetic beads (Dynabeads M-280, Invitrogen) were incubated overnight with rabbit anti-c-myc primary antibody. The beads were washed with IP200 buffer (10 mM Tris pH 7.2, 10 mM $MgCl_2$, 200 mM KCl and 0.1% Triton X-100) for 5 minutes. Glycerol gradient fractions 3 – 7, 11 – 15, 19 – 24 roughly corresponding to the 10S, 30S, and 50S regions of the gradient were pooled, mixed with IP200/1% BSA to a final volume of 500 µl and then incubated with the magnetic IgG anti-*c*-myc beads for 2 hours at 4°C. The beads were then washed and resuspended in 40 µl of IP200 buffer. SDS-PAGE Dye was added to the beads, which were boiled to release all bound proteins. Samples were resolved using SDS-PAGE, and visualized with SYPRO Ruby gel stain. The visible protein bands were excised from the gel and prepared for LC-MS/MS analysis.

4.8 Sub-cellular fractionation

One method to determine the specific localization of the protein of interest within a cell involves sub-cellular fractionation. This method is used to fractionate the cell into different cellular compartments that are resolved by SDS Page and further analyzed by immunoblotting with a specific antibody that recognizes the protein of interest. Sub-cellular fractionation employs differential centrifugation after using different concentrations of mild detergents to gradually dissolve all eukaryotic membranes. Te two-step fractionation technique we used allows us to obtain four separate sub-cellular fractions representing the cytosol, whole mitochondrion, mitochondrial matrix and insoluble pieces of membrane. First, the cellular membrane is disrupted with 0.015% digitonin, which releases a soluble cytosol fraction and a mitochondrial pellet. It is important to note that in this crude fractionation, there will be some contamination in the mitochondrial pellet, mostly intact glycosomes. The mitochondrial pellet is then lysed with 1% Triton X-100 to separate the soluble mitochondrial matrix from the insoluble membrane fraction.

$2x10^8$ induced MT420-TAP cells were harvested by centrifugation, washed once with 1xPBS containing 0.1% glucose and then resuspended in SoTE buffer (20 mM Tris-HCl pH 7.75, 2 mM EDTA pH 8, 0.6 M Sorbitol). A final concentration of 0.015% digitonin was added to the sample, which was incubated on ice for 5 minutes. The sample was spun down and the supernatant, representing the cytosolic fraction of the cell, was collected. The pellet was then resuspended in HBSS buffer (140 mM sodium chloride, 5.5 mM potassium chloride, 0.8 mM magnesium sulphate, 1.3 mM calcium chloride, 0.4 mM potassium phosphate, 0.3 mM disodium phosphate, 4.2 mM sodium bicarbonate, 5.5 mM glucose, pH 7.3) with 1% Triton X-100 and incubated for 30 minutes on ice. After a centrifugation for 15 minutes at maximal speed, the supernatant was kept isolated as the mitochondrial matrix fraction, while the pellet was resuspended in HBSS buffer and kept as the membrane fraction.

All fractions were analyzed by SDS-PAGE and immunoblotting. Antibodies against MT420-TAP and against different proteins that are known to localize in specific cellular compartments (i.e., cytosol, mitochondrion and glycosomes) were used to identify each fraction. As a marker for the cytosolic fraction, we used anti-enolase polyclonal antibody in a 1:1000, while we utilized the anti-MRP1 antibody diluted 1:1000 to distinguish the mitochondrion. Alternatively, we used the anti-hexokinase antibody in a 1:1000dilution to identify the glycosomal fraction. The MT420-TAP was detected using the anti-*c*-myc antibody in a 1:2000 dilution, utilizing the *c*-myc epitope tag, which is built into the TAP tag.

4.9 In-situ immunofluorescence antibody analysis – IFA

The IFA technique enables us to visualize our protein of interest *in-situ* with respect to the different cellular compartments of o fixed cell. The co-localization of our protein of interest with known organelle markers can be achieved through the binding of a specific primary antibody followed by the staining of a fluorescent secondary antibody. Moreover, staining of the DNA using an intercalating fluorescent stain such as DAPI (4',6-diamidino-2-phenylindole) provides information about the cell division and cell cycle status of the individual cell.

Induced and non-induced PS *T. brucei* MT420-TAP cells were harvested, washed once with 1xPBS and then resuspended in 1xPBS. A drop of the cell suspension containing$2x10^6$ cells was added on a round glass cover-slip in 24-well plate and incubated for 20 minutes with the lid to the 24-well plate on to prevent evaporation. The cells were fixed for 10 minutes with

4% paraformaldehyde (for 10 ml: 1.25 mM sodium hydroxide, PBS, 4 g of paraformaldehyd), washed twice with 1x PBS and permeabilized using 0.2% Triton X-100 in 1xPBS. Then the cells were washed once with 0.1% Triton X-100 in 1xPBS (PBS-T) and the exposed protein epitopes were blocked with 5.5 % fetal bovine serum in PBS-T for 1 hour at room temperature. This incubation helps to reduce the non-specific binding of the primary antibody to any abundant proteins and is followed by repeated washing with PBS-T.

The primary antibody solutions were prepared in 3% BSA and 1xPBS-T. The polyclonal anti-*c*-myc primary antibody (Invitrogen) specific for the *c*-myc epitope present in the TAP tag was added at the ratio of 1:200. The monoclonal anti-Hsp70 antibody specific for the mitochondrial heat shock protein 70 (Hsp70) was added at the ratio of 1:5. Glass cover-slips containing the fixed trypanosomes were incubated with 80µl of the primary antibody solution for 3 hours in room temperature. To remove excess primary antibody, the cells were washed three times with 1xPBS-T.

To visualize MT420-TAP, the cells were incubated with a secondary antibody conjugated with FITC (diluted 1:500), which recognizes the primary anti-*c*-myc rabbit antibody. In addition, an anti-mouse antibody conjugated with Texas RED was used to visualize Hsp70 (dilution 1:400). Cells were incubated with secondary antibodies for 1 hour at room temperature in the dark. The glass cover-slips were washed with 1xPBS-T and then once with 1x PBS only. The cover slip was then mounted on a glass slide using VECTASHIELD Mounting Media (Vectorlabs), which contained DAPI to stain the DNA of the nucleus and kinetoplast. The slides were observed in a Yeiss Axioplan 2 microscope (Carl Yeiss, Jena, Germany) equipped with a standard fluorescence filter set. Black and white images were recorded with a cooled F-View CCD camera, which separately captures each fluorescent dye with the AnalySIS software, version 3.2 (Soft Imaging System GmbH, Munster, Germany) Then the images were pseudo-colored (light blue for DAPI, green for FITC, and red for Texas RED) and superimposed with Adobe Photoshop, version CS4.

4.10 Overexpression and purification of recombinant MT420

Escherichia coli is an appreciated overexpression system used for the generation of large amounts of protein suitable for antibody production, *in-vitro* activity assays, or X-ray crystal structure analyses. This system is generally used because of its high productivity, low-cost and simple use.

The entire ORF of MT420 was cloned into the expression vector pOPINM (NCBI accession number: EF372396) containing the maltose binding protein (MBP) domain and the His_6 tag on its N-terminus. The expression of the protein of interest is controlled by the *lac* promoter, which is inducible using Isopropyl β-D-1-thiogalactopyranoside (IPTG).

pOPINM containing the MT420 ORF was transformed into *E. coli* BL21 cell strain (BL21-CodonPlus® (DE3)-RIL Competent Cells, Stratagene). These bacterial cells were grown in LB media (1% Bacto-Tryptone, 0.5% Yeast extract, 0.5% sodium chloride, pH 7.25), for 2 hours at 37°C. Induction of the *lac* promoter was performed using 0.5 mM IPTG for three hours. Bacterial cells were then harvested by centrifugation, washed once with 1xPBS, and then incubated for 30 minutes in BugBuster (BugBuster Master Mix, Novagen), a proprietary reagent capable of solubilizing most bacterially expressed proteins. The lysis was then spun for 20 minutes at 16,000 x g at 4°C and the resulting supernatant and pellet fractions were resolved by SDS-PAGE and followed with coomassie staining or western blot analysis to determine if the overexpressed protein is soluble or insoluble.

4.10.1 Solubilization of insoluble protein with lauroylsarcosine

A recent paper by Tao et al. describes a method to purify natively folded proteins from inclusion bodies using the anionic detergent, lauroylsarcosine sodium salt (34). Since the MT420-MBP protein was present mostly in the insoluble fraction (pellet) even after the BugBuster treatment, the pellet was resuspended in ST buffer (50 mM Tris-HCl pH 8, 300 mM NaCl, 5 mM $ZnCl_2$, 10 mM β-mercaptoethanol, 10% N-lauroylsarcosine sodium salt) and incubated over-night at 4°C while rotating. To increase the binding characteristics of the His_6 tagged protein, the concentration of sarcosyl in the solution was lowered from 10% to 2% using either the soluble fraction (column 1) obtained from the BugBuster treatment or the native binding buffer (50 mM $NaHPO_4$, 0.5 M NaCl, pH 8) (column 2). Solubilized material

was further purified by utilizing the ability of Ni-NTA agarose to bind the His_6 epitope present on the N-terminus of MT420-MBP.

4.10.2 Ni-NTA column purification

Ni-NTA purification was performed using a disposable chromatography column (Poly-Prep Chromatography column, Bio-Rad) and Ni-NTA agarose (Qiagen). The agarose beads were resuspended in MilliQ and washed with native binding buffer. The solubilized material containing 2% sarcosyl was loaded onto the column and incubated for 2 hours at 4°C while rotating. The flow through was removed from the column and the agarose beads were washed with native wash buffer (50 mM $NaHPO_4$, 0.5 M NaCl, 20 mM imidazole, pH 8). Purified protein was finally eluted with elution buffer (50 mM $NaHPO_4$, 0.5 M NaCl, 250 mM imidazole, pH 8) and the 500 µl elution fractions were analyzed by SDS-PAGE and immunoblotting.

4.11 Isolation of RNA from T. brucei PS cells

Total cellular RNA from MT420 *T. brucei* PS cells was isolated and further analyzed by Northern blot. The mid-log culture of induced and uninduced MT420-RNAi cells were harvested by centrifugation. The pellet was washed with 1xPBS containing 0.1% glucose and resuspended in RNA Blue reagent (Top-Bio). RNA was isolated using the protocol recommended by the manufacturer. Briefly, samples were incubated at room temperature for 5 minutes, then chloroform was added and the samples were vigorously vortexed for 15 seconds. Samples were incubated for another three minutes at room temperature and then spun down for 15 minutes. The aqueous phase was transferred to fresh tubes and an equal volume of Phenol: Chloroform: Isoamyl alcohol (25:24:1) pH 6 was added. The sample was vortexed for 15 seconds and then spun down for 10 minutes. This extraction was repeated twice to thoroughly remove all protein contamination. The purified RNA was precipitated using isopropanol and the RNA pellet was washed once with pre-chilled 70% ethanol before it was air-dried. The RNA pellet was resuspended in RNAse free water warmed to 65°C. The concentration of each RNA sample was measured using the NanoDrop 1000 Spectrophotometer (Thermo Scientific).

4.12 Northern blotting

Northern blotting is a technique that allows us to determine expression of the specific genes in various conditions. The method uses agarose electrophoresis to resolve the RNA separation according to the size of each molecule. The RNA is then transferred from the gel onto a nitrocellulose membrane employing capillary action through a wick underneath the gel and absorbing papers placed on top of the membrane. Finally, the RNA molecule of interest is visualized by using a radioactively labeled hybridization probe that is complementary in sequence to a part of the target sequence.

RNA samples were separated on a 1% agarose gel containing 1.1% formaldehyde and 1 x MOPS solution (0.42% 3-(N-morpholino) propanesulfonic acid, 0.068% NaOAc, 0.037% EDTA, pH 7). The blot was assembled as shown in Figure 13 and the RNA was transferred over-night to a Nitrocellulose membrane (BioTrace NT, Pall Corporation). After the transfer, the RNA was immobilized using a UV cross linker (UV Stratalinker, Stratagene) according to manufacturer.

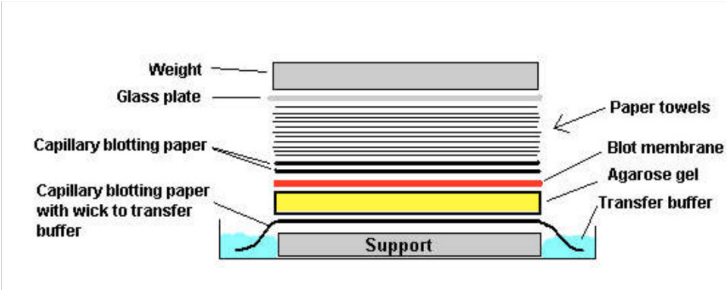

Figure 13, Scheme of the Northern blot assembly. The capillary blotting system setup for the transfer of RNA from an electrophoresis gel to a blotting membrane.
Adapted from en.wikipedia.org.

As a template for the radioactively labeled probe, we used a 500 bp fragment of the *MT420* gene that was used for the RNAi construct (Figure 10). 100 ng of this template DNA was mixed with nuclease-free water and random deca-nucleotides supplied in a reaction buffer supplemented by Fermentas. The mixture was incubated at 100°C for 10 minutes and then the DNA was cooled on ice. To this sample, Mix A (from labeling kit, Fermentas), radioactively labeled [α-^{32}P]dATP (Institute of Isotopes Co. Ltd., Budapest) and the polymerase Klenow fragment (labeling kit, Fermentas) were added. The reaction was incubated for 5 minutes at 37°C. Then the nucleotide mix (dNTP Mix, Fermentas) was added and whole reaction was incubated for another 5 minutes at 37°C. The polymerization reaction was finally stopped by the addition of 0.5 M EDTA pH 8.

Methods

Unincorporated nucleotides were removed by size exclusion chromatography on a Sephadex column (Microspin G-50 columns, Illustra) as described by manufacturer. The specific activity of the radio labeled DNA was measured using the particle detector.

The membrane bound with RNA was incubated in a hybridization oven for 2 hours at 55°C with 10 ml of a prehybridization buffer (7% SDS, 0.5 M $NaHPO_4$ pH 7.2, 1 mM EDTA pH 7, 1% BSA). The purified radioactively labeled probe was heat denatured at 100°C for 5 minutes to remove any secondary structure. It was then rapidly cooled in an ethanol ice bath for 2 minutes before it was added to the hybridization solution and incubated over-night at 55°C. The following day the membrane was washed twice with 1xSSC with 0.1% SDS (0.9% NaCl, 0.4% sodium citrate, 0.1% SDS) at room temperature and once by 0.5xSSC with 0.1% SDS (0.45% NaCl, 0.2% sodium citrate, 0.1% SDS) at 55°C. The signal was visualized using a GE Healthcare Phosphoimager screen and the Typhoo 9400. The autoradiograms were analyzed by densitometry using ImageQuant™ software.

5 Results

5.1 Bioinformatics' studies of MT420

MT420 is annotated as a hypothetical protein in the *T. brucei* genome database (www.genedb.org). This annotation implies that the function of this protein is unknown and that this protein does not possess any significant homology to any other proteins with known function. The *MT420* gene encodes a 420 amino acid protein (47.6 kDa) with a pI = 5.2.

To specify intracellular localization of MT420, we performed a signal sequence analysis using four different on-line software tools (TargetP1.1 (35), Mitoprot (36), Psort II (37), and Predotar (38)) to identify a possible mitochondrial localization signal. Indeed, all four predictions showed a high probability that MT420 is a mitochondrially targeted protein (Table 3).

Table 3. Software prediction of the cell targeting signals.

Software[a]	Probability score[b]
TargetP 1.1	0,937
MITOPROT	0,978
PSORT II	0,739
Predotar	0,840

[a] Four different software tools (first column) were used.
[b] Probability score represents the probability of the mitochondrial import.

Furthermore, we performed a protein BLAST search against a "non redundant" protein database along with domain/motif searches using different databases e.g. CDD, PFAM, PROSITE and InterPro. Interestingly we found that MT420 shares motifs with other proteins of known function with varying degree of similarity. A PROSITE domain search identified that the MT420 sequence contains the N-6 Adenine-specific DNA MTase signature motif that is typical of enzymes that methylate the amino group at the C-6 position of adenines in DNA. The consensus sequence pattern follows the scheme: [LIVMAC]-[LIVFYWA]-{DYP}-[DN]-P-P-[FYW], (PROSITE acc. no.: PS00092) (Figure 14). This motif was derived from 33 individual DNA MTase sequences. It is found in at least 13 000 different protein sequences and probably plays an important role in the binding of the methyl group from the donor molecule, S-adenosyl methionine (SAM), which is a derivative of the amino acid methionine.

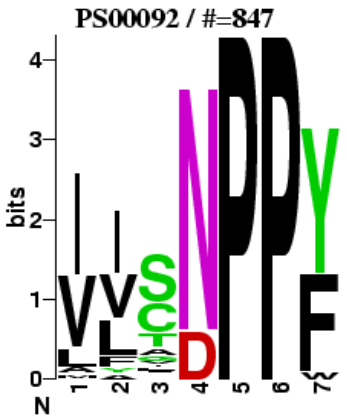

Figure 14, N-6 adenine-specific DNA MTase signature motif logo, PROSITE accession number PS00092.
Adapted from www.ExPASY.org.

Additionally, the MT420 protein displays similarity to bacterial methyltransferases that belong to the modification methylase HemK family, which possess the SAM-dependent methyltransferase activity. One of the proteins of this family, the *E. coli* PrmC/HemK protein was thought to be a protoporphyrinogen oxidase; however, the functional analysis of the nearest homolog in *Saccharomyces cerevisiae* (39) showed that it is not a protoporphyrinogen oxidase, but that it methylates a translation release factor. The alignment between *E. coli*, *T. maritima*, *S. cerevisiae,* and *T. brucei* shows a high degree of conservation in several regions, the region marked with an arrow contains the conserved motif for HemK MTases. (Figure 26 in Supplement).

Interestingly, the PrmC/HemK family is predominantly found in prokaryotes, while the only eukaryotic exceptions are found in Kinetoplastida and yeast. In order to better understand the evolutionary origins of this gene, we constructed a phylogenetic tree containing the closest homologues of MT420. Although the BLAST search implied that MT420 has significant homology to bacterial PrmC/HemK MTases, our phylogenetic study shows that this protein is probably of eukaryotic origin and was only kept only in the genomes of Kinetoplastida and yeast while it was lost from most other eukaryotes (Figure 27 in supplement).

In summary, these bioinformatic predictions revealed that MT420 contains a conserved motif typical for N6-DNA MTases, possesses a putative SAM-dependent methyltransferase domain, shares a significant similarity with the PrmC/HemK protein family, and might be targeted to the mitochondrion. To experimentally verify the intracellular localization of this protein, we utilized a *T. brucei* cell line expressing a tagged MT420.

5.2 In-cell localization of MT420

To confirm the predicted localization of MT420 in the mitochondrion, we performed a sub-cellular fractionation analysis using a mild detergent and *in-situ* immunofluorescence. First, we used digitonin to fractionate the cell into fractions representing the cytosol, the enriched mitochondrion, the mitochondrial matrix and the mitochondrial membrane fraction. Since a specific antibody against MT420 is not available, we used a *T. brucei* cell line that overexpresses the MT420-TAP protein.

Results

Using a polyclonal anti-*c*-myc antibody, western analysis of the sub-cellular fractions showed that the majority of the MT420-TAP protein was present in the mitochondrion-rich fraction, specifically in the mitochondrial matrix, whereas there was almost no signal observed in the cytosol (Figure 15, second row). The 21 kDa mitochondrial MRP1 protein (40), the 47 kDa cytosolic enolase (41) and the 51 kDa glycosomal hexokinase (42) were used as marker proteins to demonstrate the efficiency of the digitonin fractionation. The immuno-reactive signal for the cytosolic enolase was predominantly detected in the cytosolic fraction, while a small amount of the protein signal was also detected in the mitochondrion-rich fraction, suggesting that a small proportion of the of the *T. brucei* cells were not lysed. As expected, the majority of the glycosomal hexokinase signal overlaps with the mitochondrial fractions as the glycosomes tend to co-purify with the mitochondria in this sub-cellular fractionation protocol (Figure 15, fourth row).

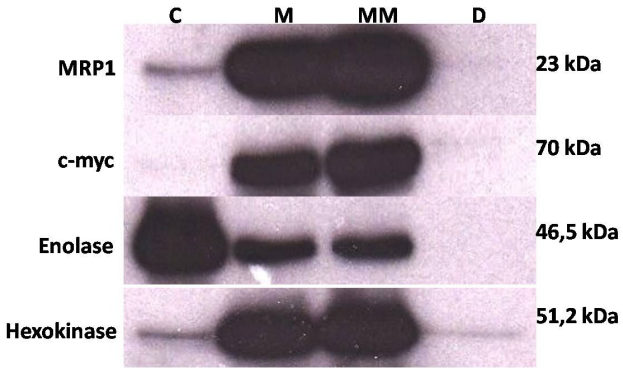

Figure 16, Subcellular fractionation and Western blotting of T. brucei MT420-TAP cell line. Anti *c*-myc antibody indicates that MT420 is localized to the mitochondrion, specifically to mitochondrial matrix. The 21 kDa mt MRP1 protein, 47 kDa cytosolic enolase and 51 kDa glycosomal hexokinase were used as a marker proteins. C, cytosol; M, whole mitochondria; MM, mitochondrial matrix; and D, membranes;

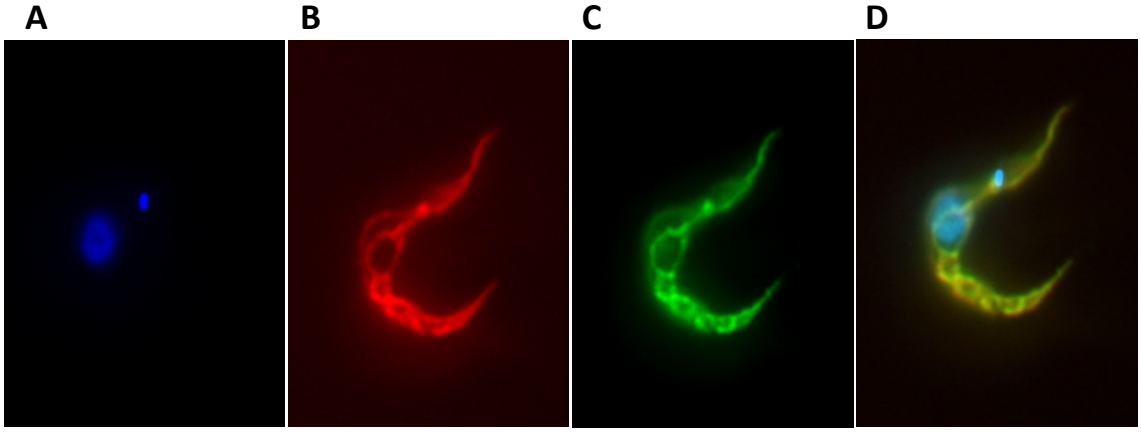

Figure 15, MT420-TAP localizes to the mitochondrion. The tagged MT420 was visualized by fluorescence microscopy using polyclonal anti-c-myc antibody coupled with FITC-conjugated secondary antibody. Co-localization immunofluorescence was performed with monoclonal antibody mAb78 against the mt heat shock protein 70. Panel A, 4,6-diamino-2-phenylindole (DAPI)-staining of nuclear and kinetoplast DNA; panel B, staining of mitochondrial hsp70; panel C, localization of tagged protein; panel D, merged fluorescence.

As a second approach, we employed an immunofluorescence assay (IFA) that allowed us to localize the MT420 *in-situ*. Using an anti-*c*-myc polyclonal antibody, we showed that the MT420-TAP was evenly distributed throughout the reticular mitochondrion (Figure 16, panel C) and co-localized with the mitochondrial protein Hsp70 (Figure 16, panel B). These results confirm the mitochondrial localization of the MT420 protein and indicate that the TAP tag did not interfere with its mitochondrial import, which further provides the validity for purifying the MT420 tagged protein (see below).

5.3 Tandem affinity purification of MT420-TAP

The TAP (tandem affinity purification) tag is a versatile tool used to characterize the composition of complexes or identify novel binding proteins of the tagged protein. Furthermore, the tag sequence can be used for western analysis using commercially available antibodies against different parts of the tag sequence (e.g. anti-His_6, anti-c-myc, anti-protein A).

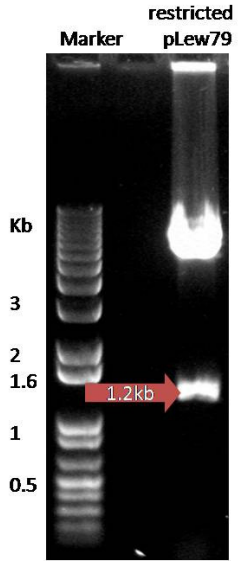

Figure 17, Restriction analysis of pLEW79-MH-TAP containing MT420. The arrow depicts the cleaved fragment that is corresponding to MT420 ORF.

We cloned the MT420 ORF into the pLEW79-MH-TAP vector and verified the correct insertion by restriction analysis (Figure 17). This vector was transfected into the *T. brucei* 29-13 strain. After the phleomycin-resistant *T. brucei* cell lines were established, we performed a pilot induction of the TAP-tagged MT420 and analyzed the selected clones by immunoblotting (Figure 18 A). An immunoblot signal corresponding to MT420-TAP was detected only in induced cells, when no signal was detected in the uninduced cells, indicating that there was tight regulation of the tagged MT420 expression. Te cell line designated as TAP4 was used for further experiments.

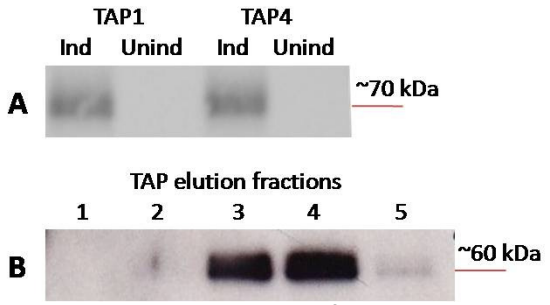

Figure 18, Western analysis of the MT420-TAP tag cell lines. A, Immunoblotting analysis of the induced (Ind) and uninduced (Unind) cell lines TAP1 and TAP4. Whole cell lysate was resolved on SDS-PAGE, blotted and probed with anti-His antibody. B, Western analysis of the elution fractions after a two-step tandem affinity purification.

We purified MT420-TAP from 600 ml of cell culture by tandem affinity

purification and analyzed the final elution fractions by SDS-PAGE and immunoblotting (Figure 18 B). The majority of the MT420-TAP protein was detected in elution fractions 3 and 4 (please note, that the size of the MT420-TAP is smaller due to the TEV cleavage of the protein A tag). We further analyzed elution fraction four by SDS-PAGE followed by SYPRO Ruby staining (Figure 19). Two visible bands, most likely representing tagged MT420, were cut from the gel, digested with trypsin, and analyzed by LC-MS/MS. The results from this analysis are shown in Table 4.

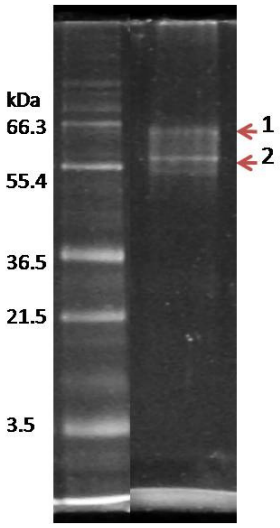

Figure 19, SYPRO Ruby staining of the TAP elution fraction 4. Arrows depict protein bands that were submitted to LC/MS-MS analysis.

Both protein bands contain the MT420 protein, which was detected by 3 unique peptides. There was one additional protein detected in protein band 1 a hypothetical conserved *T. brucei* protein, Tb11.02.3350. This hypothetical protein has a predicted mass of ~103 kDa and should migrate at a much lower rate than the observed SYPRO Ruby bands isolated for LC-MS/MS. Furthermore, this protein was detected by only 1 peptide, which indicates that it is most likely a spurious contaminant rather than a *bona fide* binding partner to MT420. Considering the MS data and the sizes of the two visible protein bands on SDS PAGE gel, we can assume that the lower band represents the endogenous MT420 and the higher band represents the tagged-MT420. This implies that MT420 may function as a dimer.

Table 4. Proteins identified by LC-MS/MS analysis in TAP elution fraction 4. (MT420 in bold).

Band	Accession number	Protein annotation	kDa[a]	Pept.[b]
	Tb10.6k5.0440	**MT420**	**47.463**	**3**
1	Tb11.02.3350	hypothetical protein conserved, *T. brucei*	102.706	1
2	**Tb10.6k5.0440**	**MT420**	**47.463**	**3**

[a] number indicates predicted molecular weight of the protein in kDa
[b] numbers indicate number of unique peptides identified by LC-MS/MS

5.4 Glycerol gradient sedimentation

In order to analyze the sedimentation profile and the putative binding partners of MT420, we performed a density-gradient ultracentrifugation followed by a western blot analysis of all the odd gradient fractions using an anti-*c*-myc antibody. Interestingly, the MT420-TAP protein was detected throughout the gradient, with the majority of the protein sedimenting between fractions 3 – 13 (Figure 20 A). This sedimentation profile of the tagged MT420 protein throughout the entire gradient suggests that MT420 might be a subunit of various putative complexes, each with a distinctive peak in the gradient. Another alternative is that the MT420-TAP is bound to nucleic acids and thus it forms protein-nucleic acid complexes that sediment at various positions throughout the gradient. To investigate these possibilities, we further purified the tagged MT420 protein using immunoprecipitation of pooled gradient fractions corresponding to ~10S (3 – 7), ~30S (11 – 15) and ~50S (19 – 24). Magnetic IgG beads bound with the anti-*c*-myc antibody were incubated with the pooled glycerol gradient fractions. The immunoprecipitated samples were analyzed by SDS-PAGE followed by SYPRO Ruby staining (Figure 20 B). Protein bands corresponding to 55 kDa and 20 kDa are presumably the light and heavy chain of the IgG molecules, which were released from the IgG beads together with the sample. We decided to further analyze the three most prominent protein bands marked as number 1, 2 and 3 by LC-MS/MS. Protein band number 1 was cut out only from the lane corresponding to 10S, while protein bands 2 and 3 were isolated from all three lanes, pooled together and analyzed by LC-MS/MS.

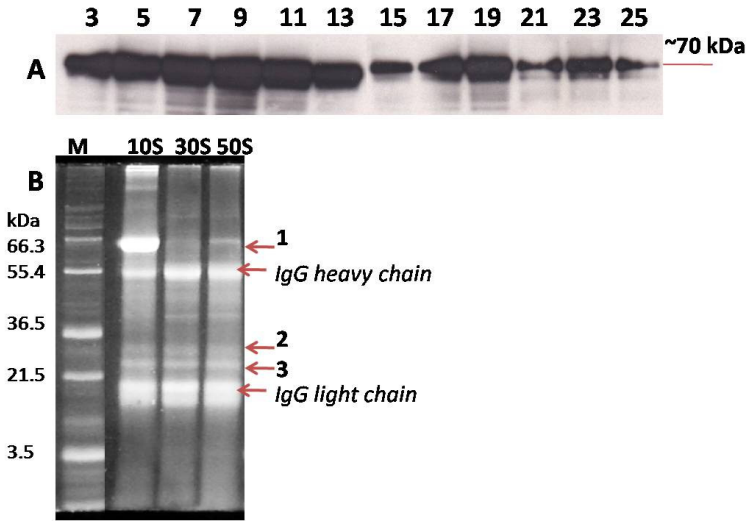

Figure 20, Glycerol gradient sedimentation of MT420-TAP followed by immunoprecipitation analysis. A. Sedimentation analysis of MT420, the cleared whole cell lyzate was loaded onto a 10 to 30% glycerol gradient and fractionated. The odd fractions were screened for the presence of the tagged protein using anti-His antibody. B. SYPRO Ruby-stained SDS-PAGE protein profile of immunoprecipitated 10S, 30S, and 50S pooled glycerol gradient fractions using the c-myc monoclonal antibody. Protein bands corresponding to immunoglobulin heavy and light chains as well as positions of size standards are indicated. The numbered arrows depict the proteins bands, which were submitted to LC-MS/MS analysis.

Results

The MS analysis is summarized in Table 5. MS analysis of band 1 detected MT420 and three conserved proteins from *T. brucei*, each with significant peptide coverage. The protein designated as Tb10.70.3050 contains a domain consistent with the SMC (structural maintenance of chromosomes) proteins that bind DNA and act in the organization and segregation of chromosomes during cell fission. The other two proteins, designated as Tb927.3.1010 and Tb927.6.1440, are hypothetical with no known motifs. Unfortunately, the analysis of band 2 did not reveal any significant matches. The tagged MT420 protein was also identified from band 3, which may imply cellular proteolysis of the overexpressed MT420 or contamination during the MS analysis. A few additional proteins (e.g. putative mitochondrial carrier protein from *T. brucei,* Tb10.389.0690) were identified by only one unique tryptic peptide, which most likely represents a contaminating peptide (data not shown).

Table 5. Proteins identified in immunoprecipitated samples.

Band[a]	Accession nr.	Name	kDa[b]	Pept.[c]
	Tb10.70.3050	protein conserved, *T. brucei*	99.809	5
	Tb927.3.1010	protein conserved, *T. brucei* chr 3 TIGR	64.301	5
	Tb10.6k5.0440	**hypothetical protein conserved, *T. brucei***	**47.463**	**5**
1	Tb927.6.1440	protein conserved, *T. brucei* chr 6 TIGR	34.781	4
	Tb10.6k5.0440	**hypothetical protein conserved, *T. brucei***	**47.463**	**5**
3	Tb927.3.1840	4 dehydrogenase putative, *T. brucei* chr 3 TIGR	33.333	3

[a] indicates a number of the analyzed protein band
[b] number indicates predicted molecular weight of the protein in kDa
[c] numbers indicate number of unique peptides identified by LC-MS/MS
The tagged protein MT420 is written in bold

5.5 RNAi silencing of MT420 in T. brucei

To investigate if MT420 is crucial for cell survival, we prepared a transgenic cell line of *T. brucei* 29-13 cells containing the p2T7-177-MT420 vector, previously described in the methods. We selected for phleomycin-resistant RNAi cell lines and then triggered the production of dsRNA using tetracycline. The cell growth was measured for 7 days using a cell counter device (Z2 Coulter – Particle count and size analyzer, Beckman Coulter) (Figure 21).

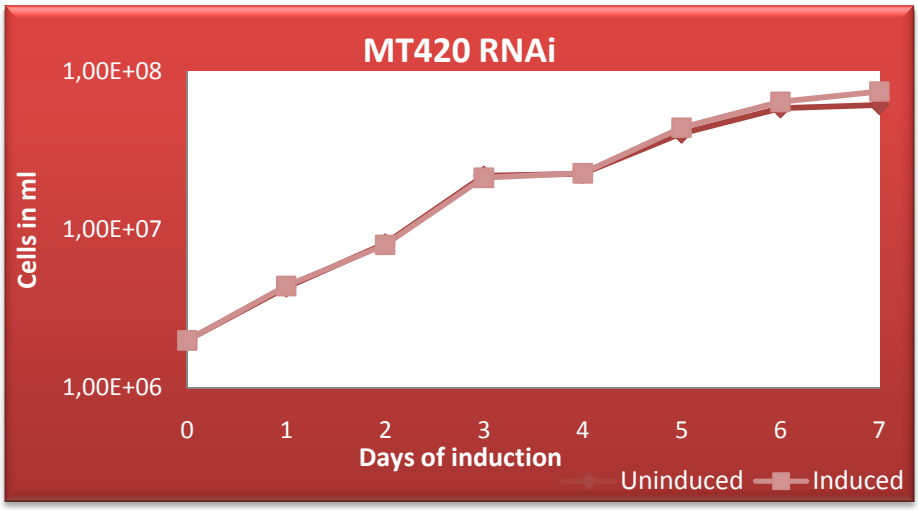

Figure 21, Growth curve of RNAi-induced (light red) and uninduced (darker red) *T. brucei* MT420. Cells were maintained in the exponential growth phase (between 106 and 107 cells/ml) and cumulative cell number represents the normalization of cell density by multiplication with the dilution factor. Every point presents two independent measurements of one culture.

The induced cells grew almost exactly at the same rate as the non-induced cells, suggesting that MT420 is not essential for the growth of PS. To assess the efficiency of RNAi, we performed a Northern analysis of the RNA isolated from uninduced and RNAi cells induced48-hours.

We analyzed the total cellular RNA to determine the abundance of the MT420 transcript. This experiment was only partially successful. We were able to visualize the radioactive signal corresponding to the dsRNA transcribed from the RNAi vector in the induced cell line, which migrates at a position that corresponds to its size.

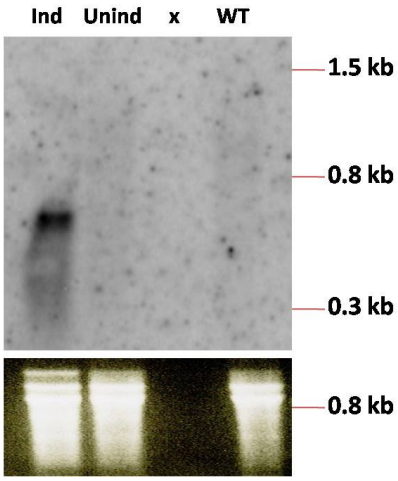

Figure 22, Northern blot analysis of total RNA 29-13 (wt), RNAi induced (Ind) and uninduced (Undin) cells. Lower panel shows the ethidium-bromide stained gel of rRNAs that serves as a loading control.

On the other hand, we were not able to detect any signal corresponding to the full-length transcript of MT420 even in the uninduced cell line and the wild type cells (Figure 22). From this result, we can only conclude that the cells are producing the dsRNA corresponding to *MT420* gene and thus the RNAi machinery is working. However, we cannot conclude the efficiency of the silencing of the transcript. One possible explanation for the lack of signal in the wild-type sample is that the probe consists of a large region of the 5'UTR, which may not be present in the mature MT420 transcript. This would depend on the location of the splice leader sequence, but if this 151bp of the 5'UTR is not present for the annealing of the probe, it could be easily removed from the membrane during the stringent washes. To address this issue, we will need to design a probe that is complementary to just the 5' region of the MT420 ORF. This would recognize both the dsRNA and the mature MT420 transcript.

5.6 Overexpression of recombinant protein

To confirm the proposed methyltransferase activity of MT420, we need to purify a sufficient amount of native and active protein.

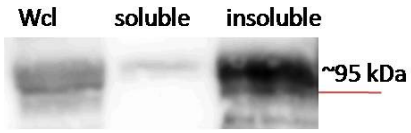

Figure 23, Pilot induction and solubility assay of MT420-MBP. Wcl – whole cell lyzate, soluble fraction, insoluble fraction.

For this purpose, MT420 was overexpressed in an *Escherichia coli* expression system. From a wide range of expression vectors, we obtained the in-fusion expression vector pOPINM and inserted the MT420 open reading frame, which is then C-terminally fused with a cleavable MBP protein and a His_6 tag. We transformed this vector into the BL21(DE3)RIL competent *E. coli* strain, as described in the methods. We performed a pilot induction and solubility assay to confirm the expression and solubility of the MT420-MBP protein. As shown in Figure 23, the solubility of MT420-MBP is insufficient as more than 95% of protein is present in the insoluble fraction. Thus we were challenged by the fact that MT420 is insoluble when over-expressed at 37°C, even if it contains the MBP fusion protein that is supposed to improve the

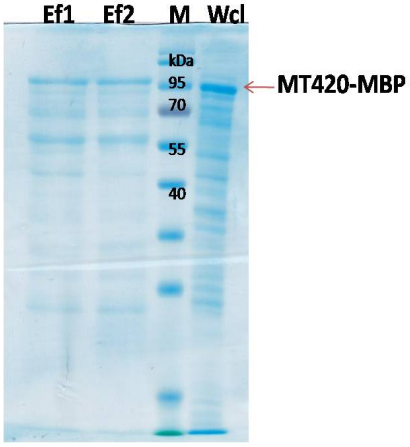

Figure 24: Purification of MT420-MBP under denaturing conditions. Ef1,2 – elution fractions 1 and 2 from Ni-NTA affinity column, Wcl – whole cell lyzate.

solubility of a protein of interest (43). We subsequently changed a number of variables, namely IPTG concentration, temperature, and type of media to find an optimal setting to produce enough of the soluble protein. We were able to optimize the protocol, which included using 5 mM IPTG for induction and growing the bacteria at 20 °C for at least 20 hours in Terrific Broth medium, but its efficiency was still too low, especially compared to the expenses incurred from this protocol (data not shown).

Since ~97% of the tagged protein was detected in insoluble fraction, we decided to purify the MT420-MBP protein under denaturing conditions, followed by stepwise dialysis in an attempt to obtain soluble protein. This approach again proved not efficient enough to provide us with a satisfactory amount of purified MT420 to even continue with dialysis. Moreover, both elution fractions from the Ni-NTA column contained a significant amount of contaminating proteins (Figure 24).

At this point, we decided to apply a new solubilization protocol based on the presence of lauroylsarcosine sodium salt, an ionic surfactant derived from sarcosine that helps to solubilize inclusion bodies when overexpressing a protein in bacteria (44). To solubilize the recombinant MT420, we added 10% sarcosyl to the insoluble fraction of the cell lysate as described in the methods. Once the proteins were solubilized with 10% sarcosyl, the sample was split into two fractions that were treated in two different ways to reduce the amount of sarcosyl in the sample so as not to interfere with the binding of the tagged MT420 to a Ni-NTA column. The

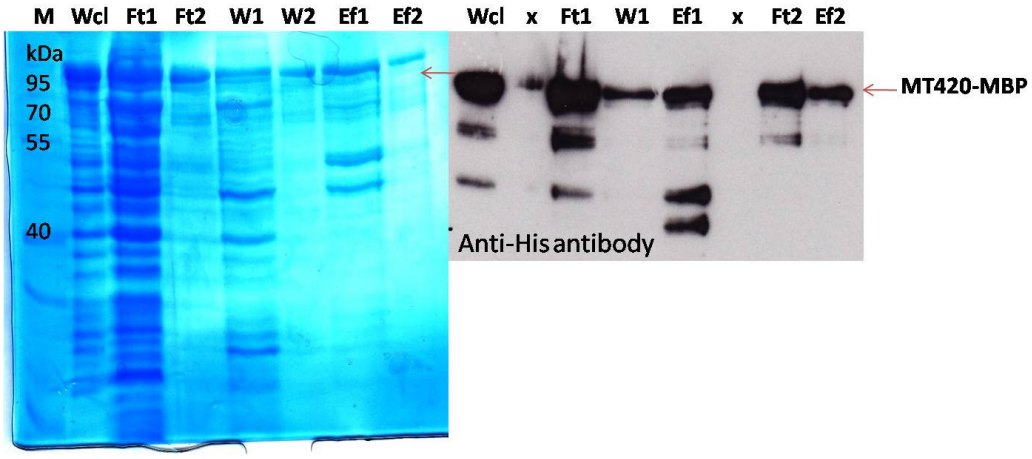

Figure 25, Ni-NTA purification of MT420-MBP using a sarcosyl solubilization protocol. Left panel, Coomassie stained gel illustrating the profile of the purification. Right panel, Immunoblotting of selected fractions from purification. Wcl - whole cell lyzate, Ft1 – flow trough from sample C1, Ft2 - flow through from sample C2, W1 – wash of sample C1, W2 – wash of sample C2, Ef1 – elution fraction from sample C1, Ef2 – elution fraction from sample C2. Protein marker positions are depicted on the left. Arrow shows the recombinant MT420-MBP protein.

first sample was diluted with the soluble fraction resulting from the initial BugBuster treatment (sample C1) that lysed the bacterial cell pelet, while the second sample was diluted with native binding buffer (sample C2). Both protein samples were loaded onto separate Ni-NTA purification columns, washed and then the recombinant protein was eluted with various concentrations of imidazole.

Though a significant amount of the MT420-MBP protein did not bind to the column (Figure 25, Ft1 and Ft2), elution fraction 1 and 2 contained a fair amount of the soluble native protein (Figure 25, Ef1 and Ef2). Unfortunately, there are some contaminating proteins present in elution fraction 1, but if these interfere with downstream activity assays, the recombinant MT420 can be further purified on a gel exclusion column. In addition, the western analysis showed a slight degradation of the tagged protein, which may be caused by the robust overexpression in bacteria, as some of these bands are detected in the whole cell lysate. Bradford analysis of both elution fractions indicated that we are able to purify 60 μg of the recombinant MT420 from sample C1 and 370 μg from sample C2. In total, we purified 450 μg of the recombinant protein from 200 ml of bacterial culture. In summary, we were able to purify the native MT40-MBP protein in a sufficient amount for the initial tests for methyltransferase activity.

6 Discussion

6.1 Bioinformatics

MT420 is a hypothetical protein, conserved among Trypanosomatids. MT420 also shares homology with the PrmC/HemK MTases from bacteria and according to Panther categorization, MT420 belongs to the HemK methyltransferase family (PTHR18895:SF7).

Sequence analysis of PrmC/HemK revealed the presence of a NPPY motif, presumably restricted to the members of the adenine-N6 and cytosine-N4 DNA amino MTases. This led to the suggestion that PrmC/HemK was itself an AdoMet-dependent DNA MTase; however, no evidence could be found that it methylates DNA (45). Later studies proved that PrmC/HemK methylates bacterial release factors (RF) and thus they are protein methyltransferases (28). Moreover, the structures of PrmC/HemK from *E. co*li and *Thermotoga maritima* were recently elucidated. Interestingly, these results showed that the putative substrate-binding domains of HemK from *E. coli* (30) and *T. maritima* (46) are structurally similar, despite the fact that they share very little sequence similarity.

The eukaryotic homologue of PrmC/HemK, containing highly conserved motifs typical for N5-glutamine S-adenosyl-L-methionine-dependent MTases was recently reported (47). This yeast Mtq1 MTase methylates the mitochondrial translation release factor 1 (Mrf1), which is a homologue of the *E. coli* peptide release factor (RF). Mrf1 contains a highly conserved GGQ motif, originally found in *E. coli* RFs, and this motif is methylated at the conserved glutamine position by MTase Mtq1. Although the methylation of the release factors is not an essential process in yeast, the lack of Mtq1 activity causes multiple phenotypes, including sensitivity to translation-fidelity drugs such as paromomycin and geneticin (47).

According to a phylogenetic study, the MT420 is a conserved protein of probable eukaryotic origin with its nearest relatives found in yeasts. While the prokaryotic proteins do share various similarities with MT420, they are not closely related. This data presents the theory that this MTase was originally an eukaryotic gene, but it was gradually lost in other eukaryotes during the evolution and it was only preserved in yeast and Trypanosomatids.

Since MT420 is closely related to yeast Mtq1, it may possess the same function in trypanosome mitochondria. Our bioinformatic results showed that MT420 contains a

mitochondrial targeting signal and is therefore putatively targeted to mitochondrion. To prove this hypothesis, we performed an intracellular localization study with MT420.

6.2 MT420 is localized into the T. brucei mitochondrion

The predicted mitochondrial localization of the MT420 protein was verified experimentally. Indeed, the analysis of the sub-cellular fractionation data showed enrichment of the signal for MT420-TAP and the mitochondrial protein MRP1 in both the intact mitochondrial fraction and the mitochondrial matrix fraction, particularly when compared to the same signals in the cytosolic fraction. On the other hand, the cytosolic protein enolase was found mostly in the cytosolic fraction, although a weak signal was also observed in the mitochondrial fraction, which may indicate that the initial cell lysis with 0.015% digitonin was less than 100% efficient. Furthermore, an antibody against a specific glycosomal protein was used to detect glycosomes. Glycosomes are single-membrane vesicles, unique for Kinetoplastida, which contain the majority of the glycolytic enzymes. Usually these abundant vesicles co-purify with the mitochondria and their separation during purification is almost impossible (48). Therefore, it is not surprising to see the majority of the glycosome specific hexokinase in the intact mitochondrial and mitochondrial matrix fractions.

The immunofluorescence results further confirmed the previous sub-cellular fractionation results, demonstrating that MT420-TAP is mitochondrially localized. To validate that the observed localization of the tagged-MT420 in induced cells is specific, we also performed an IFA analysis using the non-induced cells. As expected, we observed low background signal throughout the whole cell (data not shown). This low background staining was also seen in ~ 40% of the induced cells suggesting that some cells were not responsive to tetracycline induction. This common phenomenon can be explained by the polyclonal nature of the MT420-TAP cell line, which was never serially diluted to the point that monoclonal cell lines were isolated.

Together, these results suggest that the TAP tags did not interfere with mitochondrial import and MT420 is in fact a mitochondrially localized protein. Interestingly, MT420 is one of the first protein MTases localized to *T. brucei* mitochondrion. Only recently, it was described that there is another methyltransferase, PRMT1 that methylates the mitochondrial protein RBP16 (49).

However, while the activity assays of the mitochondrial extracts suggest that a small amount of PRMT1 may also be present in the mitochondrion, the majority of the PRMT1 activity is found in the cytosol. Furthermore, there is no further experimental data concerning the localization of PRMT1. To my knowledge, except for MT420, there is no other protein MTase with exclusive localization in the *T. brucei* mitochondrion. Most likely, this is due to the current lack of experiments involving protein MTases as they are usually widespread proteins involved in many cell processes.

Apart from the mitochondrial protein MTases, there are several known mitochondrial MTases methylating rRNA. Highly conserved mitochondrial rRNA MTases are known for their role in ribosome function and stabilization (50), (51). Interestingly, when these proteins are mutated or are knocked down, they produce consequences concerning antibiotic resistance or sensitivity (52).

6.3 MT420 is probably not present in complex

The yeast MTase Mtq1, which is homologous to MT420, interacts with the yeast mitochondrial translation release factor Mrf1. To study if *T. brucei* MT420 interacts with any other proteins, we performed a glycerol gradient sedimentation analysis and the TAP purification of MT420. First, we analyzed the sedimentation profile of MT420-TAP in a glycerol gradient. Since MT420 was detected throughout the gradient, it implies that MT420 either associates with other proteins and/or DNA and/or RNA. To further explore these possibilities, we immunoprecipitated the tagged MT420 protein from three different regions of the gradient, that correspond to 10S, 30S and 50S. The immunoprecipitated samples were analyzed by SYPRO Ruby staining followed by LC-MS/MS analysis of the three most prominent protein bands found in samples. As expected, MT420 was identified in the largest protein band visualized on the gel, which corresponds to its predicted size. Interestingly, three hypothetical proteins, Tb10.70.3050, Tb927.3.1010, and Tb927.6.1440 were also identified by more than two unique peptides. Since these proteins were not identified in the TAP-eluate, further experiments are needed to verify the interaction of these proteins with MT420. LC-MS/MS analysis of the remaining two protein bands visualized on the gel did not reveal any other putative binding partners to MT420. Furthermore, we also identify a few very abundant proteins such as tubulins, a dehydrogenase and a carrier protein. However, we have not yet identified the *T. brucei* mitochondrial release factor as it was shown in yeast (47).

In summary, we are not able to conclude if MT420 is found in a protein complex or not. Further experiments are needed to shed light on this issue. It should also be noted that the TAP tag is more than 200 amino acids long and it can interfere with the natural binding partners that might potentially from an active complex with endogenous MT420. Furthermore, we should also consider the possibility of competition between the endogenous MT420 and the over-expressed MT420-TAP. These two forms of a single protein will theoretically compete for their binding partners. To solve this issue, we may consider creating an MT420 double knock-out cell line. Then we can introduce a regulatable ectopic copy of MT420-TAP to this null background to obtain a clearer view of MT420 binding partners.

6.4 MT420 is not essential for the T. brucei cell growth

It has been shown that the loss of PrmC/HemK in *E. coli* or Mtq1 *S. cerevisiae*, does not lead to cell death. Similarly, the *T. brucei* RNAi cell line in which the *MT420* is silenced did not exhibit any growth phenotype after 6 days of RNAi induction. Interestingly, while the loss of PrmC/HemK was not lethal to the *E. coli* cells, it forced them to switch from aerobic respiration to anaerobic respiration (29). In addition, the yeast Mtq1 double knock-out exhibits a moderate growth effect on a non-fermentable carbon source. Considering these results, it would be very interesting to test if the *T. brucei* MT420 RNAi cell line will exhibit any growth phenotype when grown in the absence of glucose.

To confirm the efficiency of RNAi, a northern blot analysis was performed. We observed specific and tightly regulated expression of the dsRNA, however we were not able to detect any MT420 transcript in either the control 29-13 or in the uninduced cells. This result may imply that the MT420 mRNA is not a very abundant transcript in PS *T. brucei* cells. All together, our data suggest that MT420 is not an essential protein for the growth of PS *T. brucei* cells when grown under normal *in vitro* conditions.

6.5 Overexpression of the MT420 is a "long run"

In order to test the methyltransferase properties of MT420, we need to obtain sufficient amount of native protein. We decided to over-express and purify MT420-MBP fusion protein from *E. coli*. Large-scale overexpression of MT420 proved to be a challenge since recombinant MT420 was either insoluble within the cell or precipitated out of any purification solution immediately. Thus, native conditions showed to be insufficient in order to obtain enough protein for further *in-vitro* experiments.

After an unsuccessful effort to purify the native protein, we purified the recombinant MT420 under denaturing conditions and then attempted to refold the enzyme by stepwise dialysis. Unfortunately, the efficiency of this purification was again insufficient. Recently, we decided to integrate a new protocol that employs the ability of lauroylsarcosine salt to solubilize recombinant proteins isolated from inclusion bodies (34). This purification protocol proved to be successful as we were able to obtain 0.45 mg of protein per 200 ml of bacterial culture. We are now planning to use this recombinant protein for *in-vitro* methylation assays (see below).

6.6 Future experiments

An initial bioinformatic analysis performed on the primary sequence of MT420 implies that MT420 possess a methyl-transferase activity. MT420 is a homolog to the protein MTase, PrmC/HemK, which methylates a translation release factor. Using this knowledge, we can express the *T. brucei* mitochondrial translation release factors using a cell-free transcription and translation system and incorporate them in an *in vitro* methylation assay with or natively purified recombinant MT420 to study methylation properties of MT420. Furthermore, we would like to succeed in the creation of a MT420 double knock-out (DKO) cell line. Since MT420 is not an essential protein, we can replace the endogenous alleles using two independent cassettes containing a selectable marker flanked by the MT420 5'and 3' UTRs. The suitable *T. brucei* YTAT cell strain and the DKO plasmids were already obtained from prof. D.A. Campbell (University of Los Angeles, USA). This DKO cell line can be further utilized to introduce a TAP-tagged mitochondrial release factor to study any changes in its methylation pattern compared to wild-type YTAT cells. We hope to conclude soon that MT420 is a *T. brucei* mitochondrial protein MTase that methylates a mitochondrial translation release factor and its function is evolutionary conserved.

7 Literature

1. **Shapiro, TA and Englund, PT.** The structure and replication of Kinetoplast DNA. *Annual Review of Microbiology.* 1995, Vol. 49, pp. 117-143.

2. **van Hellemond, JJ, Opperdoes, FR and Tielens, AGM.** The extraordinary mitochondrion and unusual citric acid cycle in Trypanosoma brucei. *Mechanistic and Functional Studies of Proteins* . 2005, Vol. 50.

3. **Campbell, DA, Thomas, S and Sturm, NR.** Transcription in kinetoplastid protozoa: why be normal? *Microbes and Infection.* 2003, Vol. 5, pp. 1231-1240 .

4. **Vondruskova, E, et al.** RNA interference analyses suggest a transcript-specific regulatory role for mitochondrial RNA-binding proteins MRP1 and MRP2 in RNA editing and other RNA processing in Trypanosoma brucei. *Journal of Biological Chemistry.* 2005, Vol. 280, pp. 2429-2438.

5. **Hannaert, V, et al.** Plant-like traits associated with metabolism of Trypanosoma parasites. *PNAS.* 2003, Vol. 100, pp. 1067-1071.

6. **Matthews, KR, Ellis, JR and Paterou, A.** Molecular regulation of the life cycle of African trypanosomes. *Trends in Parasitology.* 2004, Vol. 20, pp. 40-47.

7. **Liu, YN, Motyka, SA and Englund, PT.** Effects of RNA interference of Trypanosoma brucei structure-specific endonuclease-I on kinetoplast DNA replication. *Journal of Biological Chemistry.* 2005, Vol. 280, pp. 35513-35520.

8. **Stuart, K, et al.** Kinetoplastids: related protozoan pathogens, different diseases. *The Journal of Clinical Investigation.* 4 2008, pp. 1301-1310.

9. **Barret, MP, et al.** Human African trypanosomiasis: pharmacological re-engagement with neglected disease. *British Journal of pharmacology.* 12 2007, pp. 1155-1171.

10. **Beverley, SM.** Protozomics: Trypanosomatid parasite genetis comes of age. *Nature Reviews Genetics.* 1 2003, pp. 11-19.

11. **Riou, GF, Belnat, P and Benard, J.** Complete loss of Kinetoplast DNA sequences induced by ethidium bromide or by acriflavine in Trypanosoma equiperdum. *The Journal of Biological Chemistry.* 1980, Vol. 255.

12. TriTrypDB. *TriTrypDB: The Kinetoplastid genome source.* [Online] An EuPathDB project in collaboration with GeneDB, 3 12, 2010. [Cited: 3 16, 2010.] http://tritrypdb.org/tritrypdb/.

13. **Berriman, M, et al.** The genome of the African trypanosome, Trypanosoma brucei. *Science.* 2005, Vol. 309, pp. 416-422 .

14. **Ivens, AC, et al.** The genome of the kinetoplastid parasite, Leishmania major. *Science.* 2005, Vol. 309, pp. 436-442.

15. **Fire, A, et al.** Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature.* 1998, Vol. 391.

16. **Ngo, H, Tschudi, C and Gull, K.** Double-stranded RNA induces mRNA degradation in Trypanosoma brucei. *PNAS.* 12 1998, pp. 14687-14692.

17. **Liang, XH, et al.** Trans and cis splicing in trypznosomatids: mechanism, factors, and regulation. *Eukaryotic Cell.* 2003, Vol. 2.

18. **Bangs, JD, et al.** Mass-spectrometry of messenger RNA Cap-4 from Trypanosomatids reveals 2 novel nucleosides. *Journal of Biological Chemistry.* 1992, Vol. 267, pp. 9805-9815.

19. **Ambrosio, DL, et al.** Spliceosomal Proteomics in Trypanosoma brucei Reveal New RNA Splicing Factors. *Eukaryotic Cell.* 2009, Vol. 8, pp. 990-1000.

20. **Zamudio, JR, et al.** Complete cap 4 formation is not required for viability in Trypanosoma brucei. *Eukaryotic cell.* 6 2006, pp. 905-915.

21. **Liang, Shoudan.** Genome-Wide Analysis of Epigenetic Modifications. *Journal of Computer Science and Technology.* 25, 1 2010, Vol. 1, pp. 35-41.

22. **Bujnicky, JM and Radlinska, M.** Molecular evolution of DNA-(cytosine-N4) methyltransferases: evidence for their polyphyletic origin. *Nucleic Acid Research.* 11 1999, pp. 4501-4509.

23. **Schnierle, BS, Gershon, PD and Moss, B.** Cap-specific messenger-RNA (nucleoside-O(2)'-)-methyltransferase and poly(A) polmerase stimulatory activities of Vaccinia virus are mediated by a single protein. *PNAS.* 4 1992, pp. 2897-2901.

24. **Zamudio, JR, et al.** The 2 '-O-ribose methyltransferase for cap 1 of spliced leader RNA and U1 small nuclear RNA in Trypanosoma brucei. *Molecular and Cellular Biology.* 2007, Vol. 27, pp. 6084-6092 .

25. **Arhin, GK, Ullu, E and Tschudi, C.** 2 '-O-methylation of position 2 of the trypanosome spliced leader cap 4 is mediated by a 48 kDa protein related to vaccinia virus VP39. *Molecular and Biochemical Parasitology.* 5 2006, pp. 137-139.

26. **Mora, Liliana, et al.** Methylation of bacterial release factors RF1 and RF2 is required for normal translation termination in vivo. *Journal of Biological Chemistry.* 12 2007, pp. 35638-35645.

27. **Nakayashiki, T, Nishimura, K and Inokuchi, H.** Cloning and sequencing of a previously unidentified gene that is involved in the biosynthesis of Heme in Escherichia coli. *Gene.* 1995, pp. 67-70.

28. **Dincbas-Renqvist, V, et al.** A post-translational modification in the GGQ motif of RF2 from Escherichia coli stimulates termination of translation. *EMBO Journal.* 12 2000, pp. 6900-6907.

29. **Nakahigashi, K, et al.** HemK, a class of protein methyl transferase with similarity to DNA methyl transferases, methylates polypeptide chain release factors, and HemK knockout induces defects in translational termination. *PNAS.* 2002, pp. 1473-1478.

30. **Yang, Z, et al.** Structural characterization and comparative phylogenetic analysis of E. coli HemK, a protein (N5)-glutamine methyltransferase. *Journal of molecular biology.* 2004, Vol. 340(4), pp. 695-706.

31. **Agarwal, R, Burley, SK and Swaminathan, S.** A novel mode of dimerization via formation of a glutamate anhydride crosslink in a protein crystal structure. *Proteins.* 2008, Vol. 71, pp. 1038-1041.

32. **Brun, R and Schonenberger, M.** Cultivation and In-vitro Cloning of Procyclic Culture Forms of Trypanosoma brucei in a Semi-defined Medium. *Acta tropica.* 1979, Vol. 36, pp. 289-292.

33. **Rigaut, G, et al.** A generic protein purification method for protein complex characterization and proteome exploration. *Nature Biotechnology.* 1999, Vol. 17, pp. 1030-1032 .

34. **Tao, H, et al.** Purifying natively folded proteins from inclusion bodies using sarkosyl, Triton X-100, and CHAPS. *The International Journal of Life science Methods.* 2010, Vol. 48, pp. 61-64.

35. **Emanuelsson, O, et al.** Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *Journal of Molecular Biology.* 2000, Vol. 300, pp. 1005-1016 .

36. **Claros, MG and Vincens, P.** Computational method to predict mitochondrially imported proteins and their targeting sequences. *European Journal of Biochemistry.* 1996, Vol. 241, pp. 779-786 .

37. **Horton, P and Nakai, K.** Better prediction of protein cellular localization sites with the k nearest neighbors classifier. *Proceedings / ... International Conference on Intelligent Systems for Molecular Biology.* 1997, Vol. 5, pp. 147-152.

38. **Small, I, et al.** Predotar: A tool for rapidly screening proteomes for N-terminal targeting sequences. *Proteomics.* 2004, Vol. 4, pp. 1581-1590 .

39. **Heurgue-Hamard, V, et al.** The glutamine residue of the conserved GGQ motif in Saccharomyces cerevisiae release factor eRF1 is methylated by the product of the YDR140w gene. *The Journal of Biological Chemistry.* 2005, Vol. 280, pp. 2439-2445.

40. **Zíková, A, et al.** Structure and function of the native and recombinant mitochondrial MRP1/MRP2 complex from Trypanosoma brucei. *International Journal of Parasitology.* 2008, Vol. 38, pp. 901-912.

41. **Hannaert, V, et al.** Kinetic characterization, structure modelling studies and crystallization of Trypanosoma brucei enolase. *European Journal of Biochemistry.* 2003, Vol. 270, pp. 3205-3213.

42. **Cárdenas, ML, Cornish-Bowdena, A and Ureta, T.** Evolution and regulatory role of the hexokinases. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research.* 1998, Vol. 1401, pp. 242-264.

43. **Sachdev, D and Chirgwin, JM.** Fusions to maltose-binding protein: Control of folding and solubility in protein purification. *Applications of Chimeric Genes and Hybrid Protteins* . San Diego, CA 92101-4495 USA : Aacademic Press Inc, 525 B Street, SUITE 1900,, 2000.

44. **Kovelman, R and Roeder, RG.** Sarkosyl defines three intermediate steps in transcription initiation by RNA polymerase III: application to stimulation of transcription by E1A. *Genes & Development.* 1990, Vol. 4, pp. 646-658 .

45. **Bujnicky, JM and Radlinska, M.** Is the HemK family of puptative S-adenosylmethionine-dependent methyltransferases a "missing" zeta subfamily of adenine methyltransferases? *IUBMB Life.* 1999, Vol. 48.

46. **Schubert, HL, Philips, JD and Hill, CP.** Structures along the catalytic pathway of PrmC/HemK, an N5-glutamine AdoMet-dependent methyltransferase. *Biochemistry.* 2003, Vol. 42, pp. 5592-5599 .

47. **Polevoda, B, Span, L and Sherman, F.** The yeast translation release factors Mrf1p and Sup45p (eRF1) are methylated, respectively, by the mrthyltransferases Mtq1p and Mtq2p. *Journal of Biological Chemistry.* 2006, Vol. 281, pp. 2562-2571 .

48. **Vertommen, D, et al.** Differential expression of glycosomal and mitochondrial proteins in the two major life-cycle stages of Trypanosoma brucei. *Molecular and Biochemical Parasitology.* 2008, Vol. 158, pp. 189-201.

49. **Pelletier, M, et al.** Arginine methylation of a mitondrial guide RNA binding protein from Trypanosoma brucei. *Molecular & Biochemical Parasitology.* 2001, Vol. 118, pp. 49-59 .

50. **Xu, Z, et al.** A conserved rRNA methyltransferase regulates ribosome biogenesis. *Nature Structural & Molecular Biology.* 2008, Vol. 15, pp. 534-536 .

51. **O'Farrell, HC, et al.** Binding of Adenosine-Based Ligands to the MjDim1 rRNA Methyltransferase: Implications for Reaction Mechanism and Drug Design. *Biochemistry.* 2010, Vol. 49, pp. 2697-2704.

52. **Chow, CS, Larnichhane, TN and Mahto, SK.** Expanding the nucleotide repertoire of the ribosome with post-transcriptional modifications. *ACS Chemical Biology.* 2007, Vol. 2, pp. 610-619.

53. **Ho, CK.** UB Department of Biological Sciences: Kiong Ho. *UB College of Arts and Sciences: Department of Biological Sciences.* [Online] [Cited: 3 17, 2010.] http://www.nsm.buffalo.edu/~kiongho/.

54. **Mocellin, S and Provenzano, M.** RNA interference: learning gene knock-down from cell physiology. *Journal of Translational Medicine.* 2004, Vol. 2, p. 39.

8 Supplement

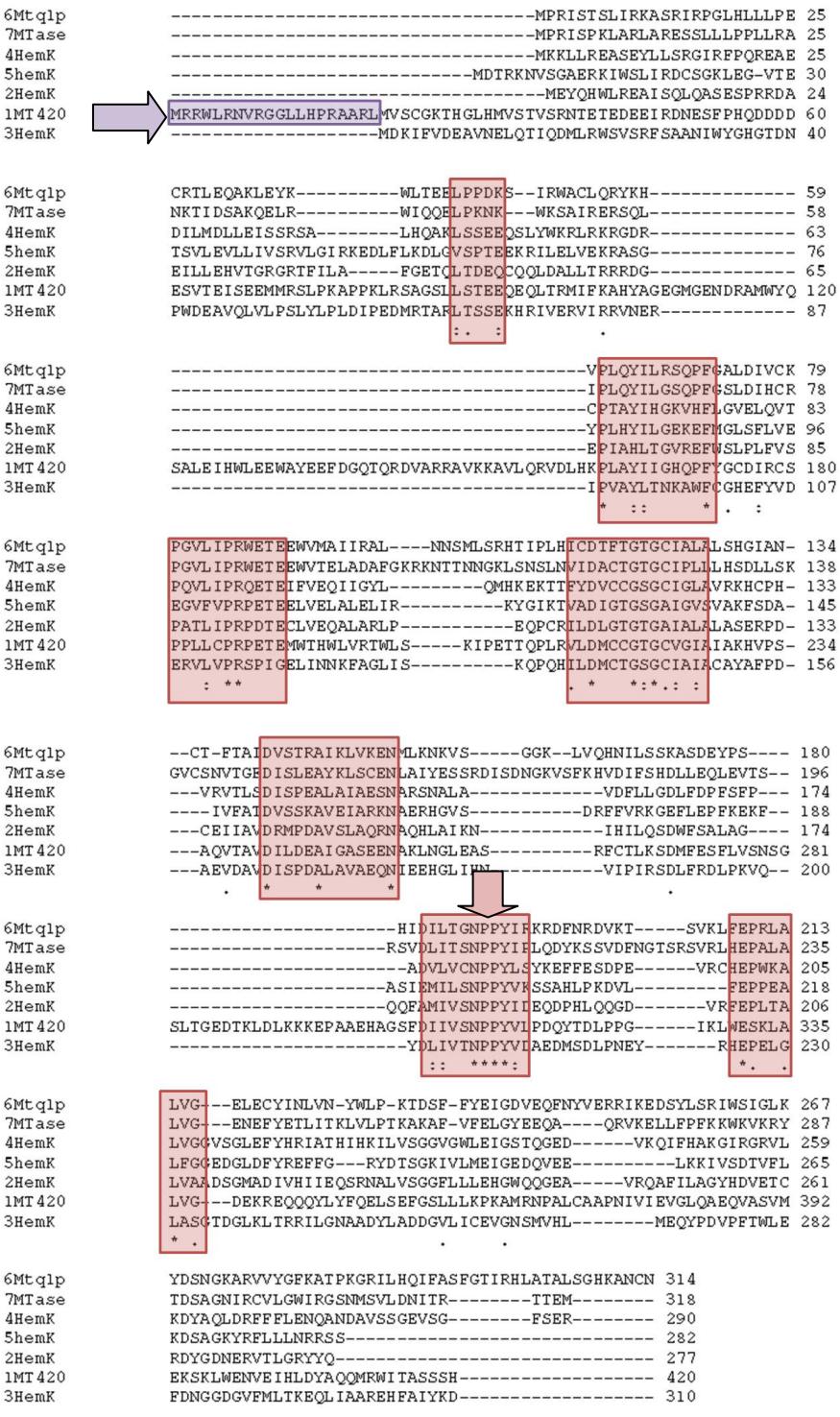

Figure 26, Alignment of homology hits. Predicted targeting sequence highlighted in violet (according to MITOPROT prediction), conserved domains highlighted in red, motif NPPY marked with arrow. Aligned sequences (NCBI accession number): 1 - MT420 (EAN78380.1), 2 – E. coli PrmC/HemK (NP_415730.1), 3 – E. coli PrmC/HemK (NP_416833.4), 4 – Chlamydia HemK (AAT35566.1), 5 – Thermotoga maritima HemK (AAD35573.1), 6 – S. cerevisiae Mtq1p (NP_014336), 7 – Pichia stipitis MTase (ABN65178). Alignment preformed on-line in ClustalW2 (EMBL-EBI).

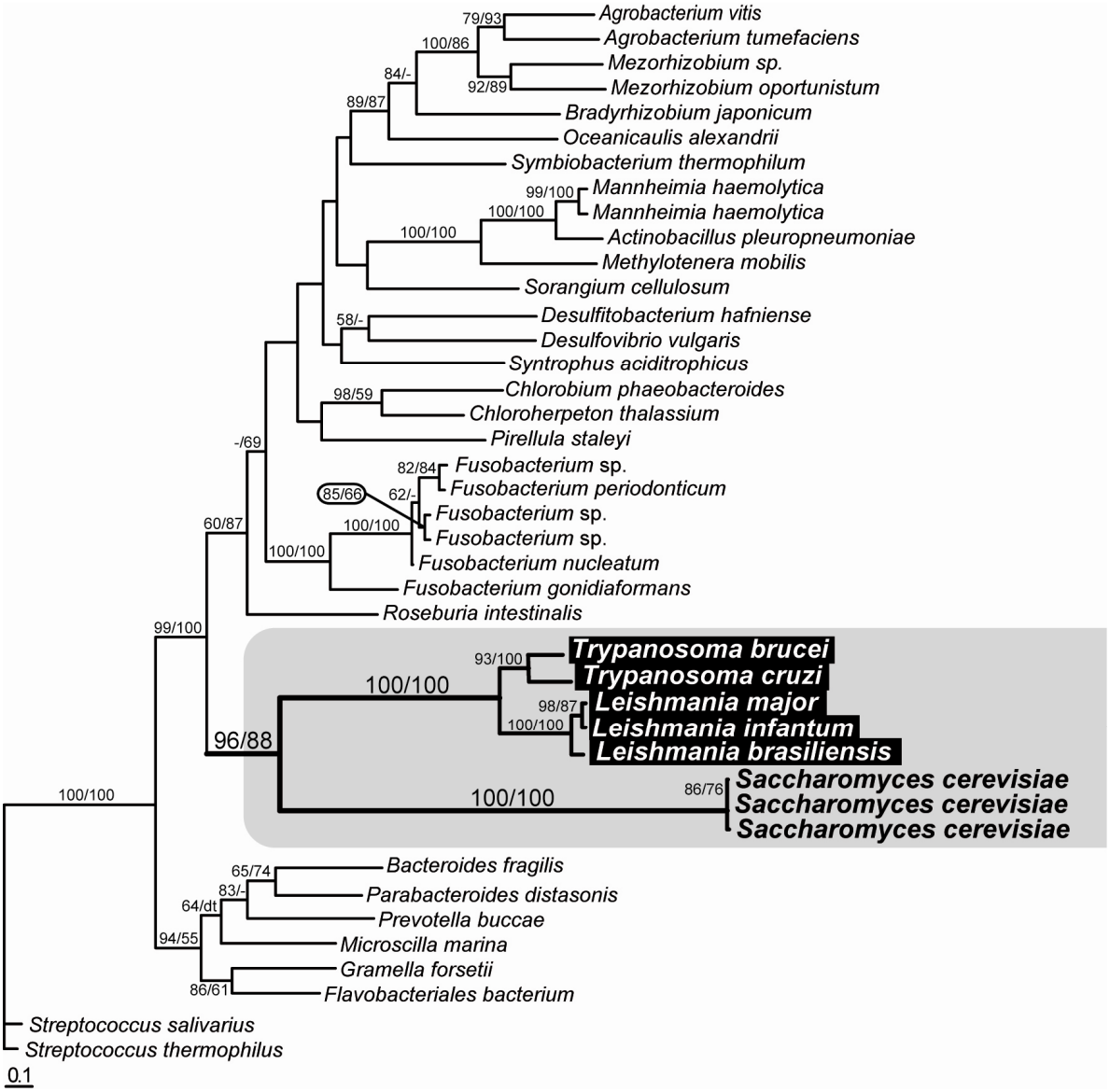

Figure 27, Maximum likelihood phylogenetic tree as inferred from methyltransferase amino acid sequences. Tree was computed using LG model and discrete gamma distribution in 4 categories. Numbers above branches indicate bootstrap support for ML (300 replicates)/NJ (AsaturA; 1000 replicates). "dt" means different topology for particular method.