Idálio de Jesus Contreiras Viegas

Licenciado em Biotecnologia

N6-methyladenosine, a new modification in *T. brucei* epitranscriptome

Dissertação para obtenção do Grau de Mestre em

Genética Molecular e Biomedicina

Orientador: Luísa Miranda Figueiredo, PhD, Universidade do Porto

Júri:

Presidente: Doutora Paula Maria Theriaga Mendes Bernardo Gonçalves

Arguente: Doutor Diogo Pinto da Cruz Sampaio e Castro

Vogal: Doutora Luísa Miranda Figueiredo

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

September 2014

Idálio de Jesus Contreiras Viegas

Licenciado em Biotecnologia

N6-methyladenosine, a new modification in *T. brucei* epitranscriptome

Dissertação para obtenção do Grau de Mestre em

Genética Molecular e Biomedicina

Orientador: Luísa Miranda Figueiredo, PhD, Universidade do Porto

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

N6-methyladenosine, a new modification in *T. brucei* epitranscriptome

Copyright Idálio Viegas, FCT/UNL, UNL

A Faculdade de Ciências e Tecnologia e a Universidade Nova de Lisboa têm o direito, perpétuo e sem limites geográficos, de arquivar e publicar esta dissertação através de exemplares impressos reproduzidos em papel ou de forma digital, ou por qualquer outro meio conhecido ou que venha a ser inventado, e de a divulgar através de repositórios científicos e de admitir a sua cópia e distribuição com objectivos educacionais ou de investigação, não comerciais, desde que seja dado crédito ao autor e editor.

Agradecimentos

"Um dia sem rir é um dia desperdiçado"

Charles Chaplin

Muito obrigado Luísa. Muito obrigado por todo o apoio, toda a dedicação, paciência e por me acompanhares em todos os momentos desde que cheguei ao teu grupo. Agradeço-te pela chefe que és, e acima de tudo, pela pessoa que és, sabes ensinar e motivar, estimulas que os teus alunos tenham opinião, sejam criativos, críticos e cresçam como cientistas e pessoas. Foi um privilégio aprender contigo.

Obrigado Francisco (Xico), foste o coorientador deste trabalho, sendo incansável na dedicação e esforço. Obrigado pela dedicação, pela disponibilidade, empenho, e acima de tudo pela paciência para ensinar e ajudar, estiveste sempre presente ao longo do projeto tornando-o possível.

Ao longo do ano, dia após dia de convivência, os elementos da UPAR estiveram sempre disponíveis para ajudar, sempre disponíveis para apoiar, sempre disponíveis para ensinar. Mais que convivência profissional estiveram disponíveis para realmente conviver, falar e sorrir, passaram de colegas a amigos. Por tudo isto, o meu muito obrigado a todos, agradeço á Pena, ao Daniel, á Margarida, ao Fabien, à Filipa, á Sandra, á Leonor, á Mafalda, á Helena e ao novo reforço o Fábio.

Como sempre faço, agradeço a todos os meus amigos, não sendo necessário estar listar nomes, eles sabem quem são. Obrigado.

Um obrigado a toda a minha família, que me acompanha e apoia desde sempre. Um obrigado muito especial aos meus pais, a quem devo tudo, mais que tudo, que com grande esforço e sacrifício possibilitaram que eu crescesse e seja a pessoa que sou. Obrigado. Um obrigado ao meu irmão, o miúdo mais importante á face da Terra.

Fui breve nas palavras, mas acreditem, quando digo obrigado estou mesmo grato!

Por fim, quero dedicar o todo meu esforço e dedicação a quem já vou tarde para agradecer, gostava de puder agradecer, estou muito grato. Diz a nossa cultura que e função do padrinho é estar presente e ajudar nos momentos mais difíceis, o meu padrinho esteve presente desde que nasci e quando eu precisei ajudou-me sem hesitar. Sem essa ajuda não estaria aqui neste momento, a escrever os agradecimentos deste trabalho, não estaria a perseguir um sonho de miúdo. Obrigado padrinho.

Resumo

A doença do sono em humanos é causada pelo *Trypanosoma brucei,* um parasita eucariota unicelular. Neste parasita, quase todos os genes são transcritos constitutivamente e, portanto, a sua regulação é sobretudo por mecanismos pós-transcricionais. A N6-metiladenosina (m^6A) é uma modificação presente no RNA que tem sido associada à regulação da expressão génica ao nível pós-transcricional em vários eucariotas. Com base nestas observações propus que esta modificação existe no transcriptoma de *T. brucei* e é um mecanismo de regulação da expressão génica ao nível pós-transcricional. Neste trabalho, pela primeira vez, detetou-se esta modificação no RNA. Também encontrei esta modificação no DNA, sendo esta a primeira descrição de um organismo em que esta modificação existe nos dois tipos de ácidos nucleicos. A modificação m^6A no RNA parece ser dinâmica: verifiquei que os níveis variam em diferentes condições biológicas, nomeadamente aumentam durante a diferenciação entres dois estadios do ciclo de vida e quando os parasitas são colocados em condições de stress provocadas por alta densidade celular. Bioinformaticamente foi procurado no genoma deste parasita, genes candidatos que codificam possíveis enzimas que catalisam a formação e a remoção desta modificação. Encontrou-se uma possível metiltransferase e seis possíveis demetilases. Para testar a sua possível função, foram geradas linhas celulares knockouts da possível metiltransferase (Tb927.7.6620) e de duas possíveis demetilases (Tb927.4.460 denominada TbALKBH1 e Tb927.5.980 denominada TbALKBH2). A quantificação dos níveis de m^6A no RNA dos knockouts da metiltransferase e de uma demetilase (TbALKBH1) não revelou evidência que suportasse a possível função proposta. No entanto, um aumento nos níveis de m^6A no RNA do knockout da possível demetilase TbALKBH2 indica que poderá ser uma demetilase de m^6A no RNA. A evidência apresentada nesta tese levanta a possibilidade de um novo mecanismo de regulação pós transcricional em *T. brucei* através desta modificação no epitranscriptoma do parasita.

Palavras-chave: *T. brucei* – modificações no RNA - N6-metiladenosina- Regulação génica

Abstract

Trypanosoma brucei is a unicellular eukaryote parasite that causes human sleeping sickness. In this parasite, transcription is mainly constitutive and gene expression regulation occurs essentially at post-transcriptional level. N6-methyladenosine (m^6A) is an RNA modification associated with post-transcriptional gene regulation in eukaryotes. These observations led to the proposal that this modification occurs in *T. brucei* transcriptome and is involved in post-transcriptional gene regulation. In this thesis, m^6A was detected for the first time in *T. brucei* RNA and additionally in DNA, from bloodstream and procyclic life stages. As far as I know, this is the first description of an organism in which has m^6A is found in both type of nucleic acids. In RNA, I observed that the levels are regulated in different biological circumstances, namely, it increases during differentiation from bloodstream to procyclic life-cycle stages and it also increases when parasites are stressed by being placed at high cell density. *T. brucei* genome was searched with bioinformatics tools to find enzymes that catalyse the formation and the removal the of m^6A modification in RNA. One putative RNA m^6A methyltransferase and six putative demethylases were found. Knockout cell lines of the putative methyltransferase (Tb927.7.6620) and of two putative demethylases (Tb927.4.460 named TbALKBH1 and Tb927.5.980 named TbALKBH2) were generated to test their putative functions. Quantification of m^6A levels in RNA from the knockout cell lines did not reveal evidence that support the putative function of the methyltransferase and one demethylase (TbALKBH1). However, knockout of TbALKBH2 resulted in a slight increase in m^6A levels, suggesting that this candidate could be an RNA m^6A demethylase. The evidence presented in this thesis raises the possibility of post-transcriptional gene regulation mediated by the presence of m^6A modification in *T. brucei* epitranscriptome.

Keywords: *T. brucei* - RNA modifications - N6-methyladenosine- Gene regulation

Contents

Index of figures

Index of tables

Abbreviations

AAT: Animal African trypanosomiasis

ALBA: Acetylation lowers binding affinity

ALKBH: AlkB homologue

BLAST: Basic Local Alignment Search Tool

Bp: Base pair

BSF: Bloodstream form

CNS: Central nervous system

DALYs: Disability-adjusted life years

DNA: Deoxyribonucleic Acid

DNase I: Deoxyribonuclease I

ECL: Enhanced chemiluminescence

EP: EP procyclins

E-value: Expect value

FTO: Fat and mass associated protein

GPEETS: GPEET procyclins

HAT: Human African trypanosomiasis

HMM: Hidden Markov Model

IME4: Inducer of Meiosis 4

KO: Knockout

lncRNA: long non-coding RNA

m^5C: 5-methylcytosine

m^6A: N6-methyladenosine

m^7G: 7-methylguanosine

Me-RIP: m^6A-specific methylated RNA immunoprecipitation

METTL14: Methyltransferase like 14

METTL3: Methyltransferase like 3

miRNA: Micro RNA

mRNA: Messenger Ribonucleic Acid

MSA: Multiple sequence alignment

MTA: mRNA adenosine methylase

ncRNA: Non-coding RNA

NEB: New England Biolabs

NPC: Nuclear pore complex

ORF: Open reading frame

PC: Procyclic form

PCR: Polymerase Chain Reaction

Poll I: RNA polymerase I

Poll II: RNA polymerase II

RBP: RNA binding protein

RNA: Ribonucleic Acid

RNase A: Ribonuclease A

RRM: RNA recognition motif

rRNA: Ribosomal Ribonucleic Acid

SAM: S-adenosylmethionine

SL: Spliced leader

S_N2: Nucleophilic Substitution bi-molecular

TDB: Trypanosome dilution buffer

tRNA: Transfer Ribonucleic Acid

TSS: Transcription start site

TTS: Transcription termination site

UTR: Untranslated region

VSG: Variable surface glycoprotein

WTAP: Wilms' tumor 1-associating protein

YLL: Years of life lost

YTHF: YTH domain family

1.Introduction

1.1 Trypanosoma brucei

1.1.1. Human African trypanosomiasis and human infective subspecies

Trypanosoma brucei (*T. brucei*) is a unicellular protozoan parasite. This parasite causes the Human African trypanosomiasis (HAT), also known as sleeping sickness (Brun et al., 2010). This disease occurs in 36 sub-Saharan countries, where 6314 new cases were reported in the year of 2013 (Franco et al., 2014). The impact caused by HAT in the population was estimated in 1.6 million disability-adjusted life years (DALYs) per year and 27 years of life lost (YLL) per death (WHO, 2012) . Clinically, HAT presents two stages: first the haemolymphatic stage, characterized by the presence of parasites in the blood and in the interstitial space of diverse tissues. Second the meningoencephalitic stage, characterized by infiltration of parasites in the central nervous system (CNS) (Brun et al., 2010; Kennedy, 2004). During the first stage the main symptoms are fever, pruritus, lymphadenopathy and hepatosplenomegaly. Sleep disturbances that result from dysregulation of the circadian rhythm is the major symptom of the second stage. These sleep disturbances are responsible for the attribution of the name sleeping sickness. If untreated, HAT ultimately leads to coma and death (Brun et al., 2010; Kennedy, 2004).

Two subspecies of *T. brucei* can cause HAT, *T. brucei rhodesiense* (East Africa) and *T. brucei gambiense* (West Africa). The distribution is represented in figure 1.1. The infection caused by *T. brucei rhodesiense* is acute, leading to death in weeks or months. As for *T. brucei gambiense*, the disease has a progressive course with a chronic infection and death occurs after several years (3-7 years) (Brun et al., 2010; Kennedy, 2004). Parasites are transmitted to humans by the tsetse, a fly from the *Glossina* species, during its blood meal (Brun et al., 2010; Dyer et al., 2013). Besides HAT, this parasite also causes Animal African trypanosomiasis (AAT, or nagana) in other mammals like cattle (Brun et al., 2010; Steverding, 2008).

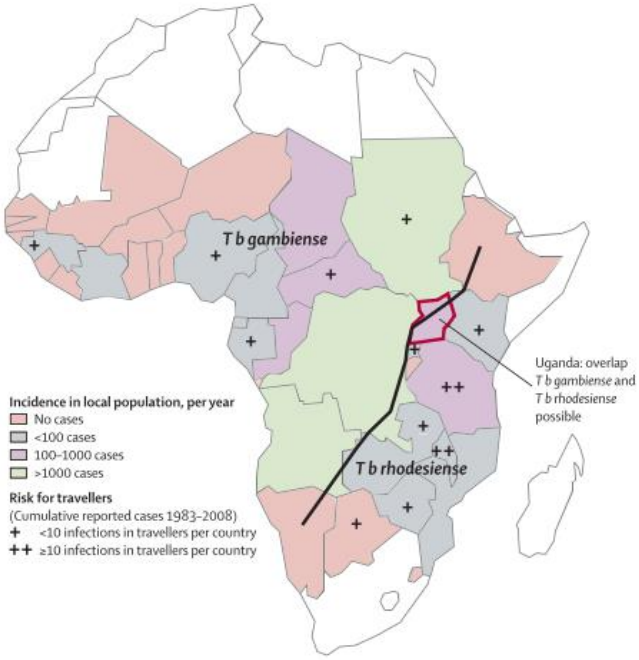


Figure 1.1. HAT distribution with incidence and risk for travellers indicated. Black line separate the *T.b. gambiense and T. b. rhodesiense* distribution. Adapted from (Brun et al., 2010).

1.1.2. *T. brucei life cycle*

The life cycle of *T. brucei* is divided between the mammalian host and the tsetse vector (Figure 1.2.) (Fenn and Matthews, 2007). In the blood of the mammalian host, parasites proliferate as bloodstream slender forms. When the parasite population increases, cells are able to differentiate into non-dividing stumpy forms, using a quorum sensing mechanism (Fenn and Matthews, 2007; Matthews et al., 2004). Stumpy cells are competent to complete the life cycle if uptaken by the tsetse. When stumpy cells enter the midgut of the transmission vector, the progression of differentiation into procyclic forms takes place. Procyclic cells migrate to proventriculus and differentiate into epimastigote forms. After this event, some epimastigote cells migrate to the salivary glands of the fly, where they differentiate into non-dividing metacyclic forms (Dyer et al., 2013; Fenn and Matthews, 2007). It was proposed that in the salivary glands of the tsetse, *T. brucei* could pass through a meiotic stage in the life cycle, besides the mitotic cell division (Peacock et al., 2011). When the tsetse takes another blood meal from a mammal, metacyclic forms are injected into its bloodstream, where the parasites differentiate into bloodstream slender cells closing the cycle (Matthews et al., 2004).

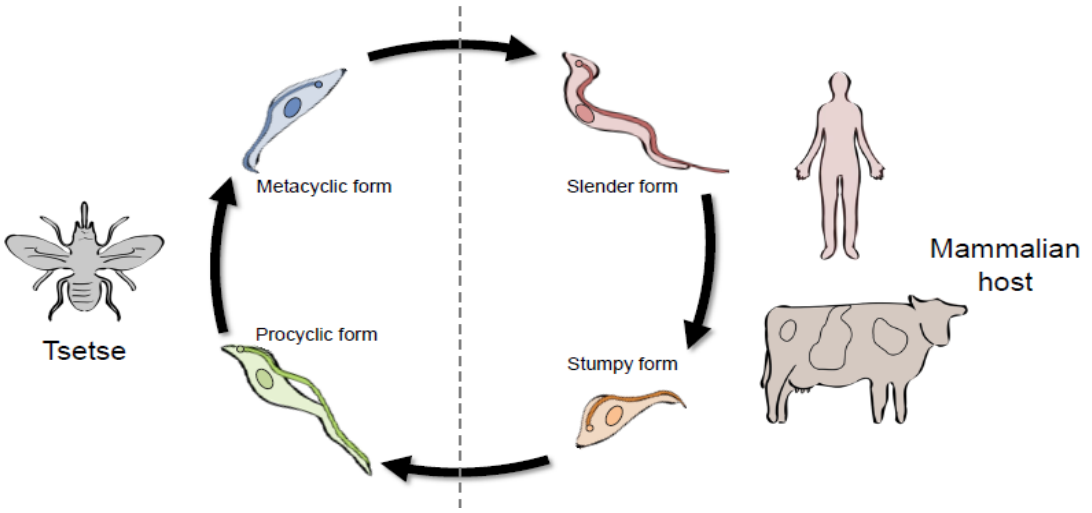


Figure 1.2. *T. brucei* life cycle (courtesy of Daniel Pinto-Neves.

Cell differentiation throughout life cycle implies that parasites can adapt to different environments (Fenn and Matthews, 2007; Matthews, 2005). This adaptation is defined by alterations of gene expression. Comparison of the transcriptome of different life-cycle stages revealed that around 28-40% of genes are differentially expressed between bloodstream slender forms and procyclic forms (Nilsson et al., 2010; Veitch et al., 2010). Consistent with these observations, recent studies comparing protein expression between bloodstream slender forms and procyclic forms revealed that around 33-48% of the proteome is different (Butter et al., 2013; Gunasekera et al., 2012).

Also during differentiation, some genes are co-regulated establishing multiple post-transcriptional regulons. These co-regulated clusters are formed by transcripts from genes involved in diverse functions and some regulons are composed of genes involved in the same biochemical pathway. For example, genes involved in cell division and macromolecular biosynthesis are co-regulated during differentiation, like ribosomal and flagellar proteins encoding genes (Queiroz et al., 2009).

1.1.3. Genome Organization in T. brucei

T. brucei is a diploid organism whose 26 Mb genome is organized in different classes of chromosomes: eleven pairs of megabase chromosomes (around 1 Mb - 6 Mb), one to six intermediate chromosomes (200 Kb – 900 Kb) and around one hundred minichromosomes (50 Kb - 150 Kb) (El-Sayed et al., 2000). The megabase chromosomes have been sequenced and shown to harbour 9 068 genes (Berriman et al., 2005). Recent studies allowed the identification of 1 114 new non-annotated genes (Kolev et al., 2010). Genes are organized in polycistronic units, similar to bacterial operons, but unlike these, genes in the same polycistronic unit do not seem be involved in the same pathway

(Berriman et al., 2005). Unlike megabase chromosomes, which are diploid, the intermediate chromosomes and minichromosomes seem to be haploid (Ersfeld, 2011).

20% of the genes of *T. brucei* encode for variant surface glycoproteins (VSGs) (Cross et al., 2014), which form a dense layer at the surface of parasite. VSGs are transcribed in a monoallelic fashion (Rudenko, 2010; Taylor and Rudenko, 2006), so that only one VSG coat is exposed to the immune system at a time. By a mechanism known as antigenic variation, some cells of the parasite population periodically change the expressed VSG, allowing these cells to escape the immune system and ensure the infection persists (Rudenko, 2010; Taylor and Rudenko, 2006).

1.1.4. Gene expression in T. brucei

Transcription in trypanosomes shows clear differences from "canonical" eukaryotes. Due to the polycistronic organization of the genes, they are transcribed as a polycistronic transcript by RNA polymerase II (Pol II), without any apparent transcriptional control (Clayton, 2002; Palenchar and Bellofatto, 2006). The polycistronic transcript is processed into individual mRNAs through a process called trans-splicing. In this process, a 39 nucleotide RNA cap sequence, termed spliced leader, is added to the 5' end of the newly transcribed gene, while a poly-A tail is added at the 3' end of upstream transcribed gene (Liang et al., 2003; Palenchar and Bellofatto, 2006). RNA polymerase I (Pol I) transcribes not only rRNA (18S, 5.8S and 28S), but also life-cycle stage specific proteins, including VSGs (in bloodstream forms) and procyclins (EPs and GPEETs, in procyclic forms) (Palenchar and Bellofatto, 2006). Unlike Pol II , Pol I is regulated at transcriptional level (Rudenko, 2010).

1.1.4.1. Transcriptional control in T. brucei

In eukaryotes, gene expression is controlled by an interconnected net of diverse mechanisms. These mechanisms involve genetic elements like promoters (Juven-Gershon and Kadonaga, 2010), transcription factors assembly (Lemon, 2000), enhancers and silencers (Kolovos et al., 2012; Ong and Corces, 2011). Beyond genetic elements, gene expression is regulated by molecular mechanisms that do not involve changes in DNA sequence and that are called epigenetics (Goldberg et al., 2007; Jaenisch and Bird, 2003). Epigenetic mechanisms include DNA methylation (Bird, 2002; Jones, 2012), non-coding RNAs (Kaikkonen et al., 2011; Mercer and Mattick, 2013), alterations in chromatin structure mediated by histone modifications (Bannister and Kouzarides, 2011; Kouzarides, 2007),

chromatin remodelling (Clapier and Cairns, 2009; Saha et al., 2006) and organization of chromatin in the nucleus (Fedorova and Zink, 2008; Schneider and Grosschedl, 2007).

In *T. brucei*, Pol II promoters lack well-established genetic elements. The only exception is the promoter of the Spliced Leader gene (Schimanski et al., 2005). Besides the lack of unidentified promoters, very few transcription factors can be found in the genome of this parasite (Iyer et al., 2008). Histone variants are present at transcription start sites (TSS) and transcription termination sites (TTS), suggesting that chromatin may play an important role in defining key functional sites of the chromosomes (Siegel et al., 2009).

A wide variety of chemical modifications has been found in histones of several eukaryotes: acetylation, methylation, phosphorylation, ubiquitylation, sumoylation, ADP ribosylation, deimination, proline isomerization) (Bannister and Kouzarides, 2011). These modifications are dynamic. The enzymes that add the chemical modifications are called "writers" and those that remove are called "erasers" (Jakovcevski and Akbarian, 2012). Histone modifications can function by the recruitment of effectors proteins or complexes. These recruited complexes have histone modifications recognition domains and are called "readers" of the epigenetic code (Kouzarides, 2007; Yun et al., 2011). *T. brucei* has fewer histone modifications, than most other eukaryotes (Figueiredo et al., 2009). Some modifying enzymes have been characterized, including lysine acetyltransferase (MYST family and EPL3 homologues), deacetylases (HDAC 1-2,HDAC 3-4 and SIR2 related histone deacetylases) and lysine methyltransferases (Disruptor of telomerase silencing DOT1 homologues) (Figueiredo et al., 2009).

In mammalian cells, DNA methylation (5-methylcytosine) is an important epigenetic mark that occurs mainly in CG repetitions (CpG islands) located at transcription start sites and is involved in gene repression (Bird, 2002). DNA methylation can also occur in other transcriptional start sites without CpG islands, in gene bodies, at regulatory elements and at repeat sequences (Jones, 2012). 5-methylcytosine has also been detected in nuclear DNA of *T. brucei*, both in bloodstream and procyclic forms (Militello et al., 2008). A putative 5-methylcytosine methyltransferase has been found in the genome (Militello et al., 2008), but the activity was not yet empirically tested.

Besides this DNA modification, bloodstream form of *T. brucei* has another unusual modification in its DNA, the base J (β-d-glucopyranosyloxymethyluracil) (Gommers-ampt et al., 1993). Base J is found mainly in the telomere repeats, repetitive sequences and transcription termination sites. Two base J binding proteins were found, J-binding protein 1 (JBP1) and J-binding protein 2 (JBP2). These base J binding proteins are required for the synthesis of base J. (Borst and Sabatini, 2008). In *Leishmania*, base J is required for proper genome wide Pol II transcription termination (van Luenen et al., 2012), however this genome wide function is not conserved in *T. brucei*, where base J controls transcription termination at specific locations (Reynolds et al., 2014).

1.1.4.2 Post-transcriptional control in T. brucei

Diverse post-transcriptional mechanisms influence gene expression, including nuclear transport (Köhler and Hurt, 2007; Strambio-De-Castillia et al., 2010), RNA decay (Garneau et al., 2007; Wilusz and Wilusz, 2004), and translation regulation. These processes can be mediated by the binding or interaction of diverse RNA binding proteins (Glisovic et al., 2008; Lunde et al., 2007) and non-coding RNAs with mRNAs (Kaikkonen et al., 2011).

Nuclear transport

mRNAs need to be transported from the nucleus to the cytoplasm, where they are translated. mRNAs are exported via nuclear pore complexes (NPCs), cylinder structures composed of many proteins that cross the double nuclear membrane. NPCs are also involved in epigenetic control of gene expression through interactions with the chromatin (Rodríguez-Navarro and Hurt, 2011; Strambio-De-Castillia et al., 2010). The structure of the NPC seems to be conserved in *T. brucei*, suggesting that RNA export in *T. brucei* could be similar to other eukaryotes (DeGrasse et al., 2009).

RNA stability

RNA stability is one major factor in the regulation of gene expression. RNA stability depends on the interaction between several molecular mechanisms, which include untranslated regions (UTRs) (Mignone et al., 2002), non-coding RNAs (Kaikkonen et al., 2011) and RNA-binding proteins (Glisovic et al., 2008; Lunde et al., 2007). The presence in the UTRs of different motifs, recognition regions and secondary structures influences the interaction of proteins and non-coding RNAs. In general, the regulatory elements present in 5' UTR are more associated with translation efficiency, while the elements in the 3' UTR are associated with mRNA stability (Mignone et al., 2002). In *T. brucei*, diverse putative regulatory elements were found in UTRs, which could potentially regulate gene expression (Mao et al., 2009). Besides the presence of regulatory elements, transcripts that encode for the same protein could have different UTRs as result of heterogeneity in trans-splicing (Nilsson et al., 2010).

RNA binding proteins

In a cell, RNA molecules are associated to RNA binding proteins, which control RNA transport, decay and translation (Glisovic et al., 2008; Lunde et al., 2007). Therefore, RNA binding proteins are important factors in the regulation of gene expression. There are diverse RNA binding proteins, which are characterized by the presence of one or several RNA binding domains, including for example the RNA-binding domain, K-homology domain, RGG box (Glisovic et al., 2008; Lunde et al., 2007). *T. brucei* has diverse proteins with RNA binding domains, including RNA recognition motif (RRM), "acetylation lowers binding affinity" domains (ALBA), Pumilio domains (RBP) and

Zinc-Finger domains (CCCH) (Clayton, 2013; Kolev et al., 2014). RNA binding proteins are involved in diverse biological processes, for example overexpression of one RNA binding protein, RBP6, in procyclics leads to differentiation to epimastigotes and metacyclics forms (Kolev et al., 2012). Another example is the response to heat shock mediated by the RNA binding protein ZC3H11. This protein binds and stabilizes diverse mRNAs encoding heat shock proteins (Droll et al., 2013).

RNA decay

Steady state levels of RNAs in a cell are affected by their rate of decay (Garneau et al., 2007; Wilusz and Wilusz, 2004). The usual mechanisms that lead to RNA decay can be divided in three pathways: the deadenylation-dependent mRNA decay, in which the first step is the removal of poly A tail, followed by the decapping and degradation of the RNA from 5' to 3' end (5'→3' decay); alternatively, after the removal the poly A tail the degradation can start at the 3' end (3'→5' decay); The other pathway is the deadenylation-independent mRNA decay, in which the decay starts with mRNA decapping, followed by degradation from the 5' end; The endonuclease-mediated mRNA decay is the third pathway that starts with an internal cleavage of RNA by an endonuclease, followed by the degradation of the RNA fragments (Garneau et al., 2007; Wilusz and Wilusz, 2004). In *T. brucei* the majority of RNAs are probably degraded by a deadenylation-dependent mRNA pathway. (Clayton, 2014) *T. brucei* has deadenylation enzymes (NOT complex) (Färber et al., 2013) and enzymes involved in 5'→3' decay (XRNA) (Manful et al., 2011). However, no enzymes have been identified that could be responsible for mRNA decapping.

Regulation by non-coding RNAs

Diverse small non-coding RNAs regulate gene expression at the post-transcriptional level (Ghildiyal and Zamore, 2009). The most studied small RNAs are the micro RNAs (miRNA) and small interfering RNAs (siRNA), which are very similar in their biochemical properties and pathways of action (He and Hannon, 2004). *T. brucei* has an intrinsic interference RNA pathway that leads to RNA degradation (Ngô et al., 1998). Several components of the pathway have been identified: two DICER proteins (TbDCL1 and TbDCL2) (Patrick et al., 2009) and the argonaute protein (AGO1) (Shi et al., 2004). The small RNAs of *T. brucei* have around 23-26 nt and originate from diverse genomic sources, including natural antisense transcripts, tRNAs, rRNAs and transposable elements. (Zheng et al., 2013b). Bioinformatic analysis have proposed a group of putative miRNAs that could target VSGs (Mallick et al., 2008).

Long non-coding RNAs (lncRNAs) can also regulate gene expression (Kung et al., 2013; Wilusz et al., 2009). These are RNAs which are typically longer than 200 nt and do not encode for proteins (Kung et al., 2013). lncRNAs can be classified based on the genomic localization from which they are transcribed, including long intronic ncRNAs (from introns), natural antisense transcripts (transcribed in the complementary strand of the ORF) and stand-alone ncRNAs (transcribed from

transcription units independent of ORFs) (Kung et al., 2013). The functions of lncRNAs are diverse, ranging from regulation of transcription, chromatin structure and nuclear organization (Wilusz et al., 2009). At the post-transcriptional level, lncRNAs typically interact with mRNAs and this interaction modulates their processing, stability and translation (Kung et al., 2013).

Cytoplasmic storage of mRNAs

Another means of regulating gene expression is by storing and/or degrading RNA in cytoplasmic structures (Balagopal and Parker, 2009; Eulalio et al., 2007). RNAs can be stored without degradation, delaying translation, a process called translational repression. RNAs can be stored in constitutive structures, called P-bodies, or in temporary stress granules (Balagopal and Parker, 2009; Eulalio et al., 2007). *T. brucei* has P-bodies in the cytoplasm where RNAs are stored as part of the RNA processing pathway (Cassola, 2011). Stress granules are formed under heat shock and they are associated with the storage of transcripts. When normal grow conditions are established, stored mRNAs are released again to the translating pool. Starvation conditions in *T. brucei* lead to the combination of P-bodies with several ribonucleic complexes forming granules called mRNA granules, where transcripts are stored (Cassola, 2011).

Translation regulation

Protein translation is also subject to regulation. This can happen globally at the level of translation initiation. Translation of specific mRNAs is also dependent on the binding of proteins that recognize regulatory elements in the UTRs (Gebauer and Hentze, 2004). Translation efficiency is also important in *T. brucei*, varying among transcripts, in a range around 117 fold in the procyclic forms and around 64 fold in bloodstream form (Vasquez et al., 2014).

1.2. N6-methyladenosine (m^6A) in RNA

1.2.1. RNA modifications

More than one hundred different post-transcriptional chemical modifications have been found in RNA molecules, either in the four nitrogen bases or in the RNA backbone (Cantara et al., 2011; Machnicka et al., 2013). These chemical alterations could be the ligation of a chemical group (small or bulky) or an isomerization, as summarized in figure 1.3., and the same nucleotide could have more than one modification at the same time. Different RNA types including mRNAs, tRNAs, rRNAs, snRNAs and miRNAs are modified and RNA modifications are present in the three domains of life (Cantara et al., 2011; Machnicka et al., 2013). Potentially these modifications could modulate the

function, stability and information content of the RNA molecule, however the role of these modifications are, in general, not understood (Helm and Alfonzo, 2014; Li and Mason, 2014). In some cases the function of specific modifications starts to be revealed, including that of 7-methylguanosine, pseudouridine and 5-methylcytosine.

7-methylguanosine (m^7G, guanosine methylated in N7 position) is present in tRNA and rRNA in bacteria and, additionally to these RNAs types, in eukaryotes is also present in mRNA (Cantara et al., 2011; Machnicka et al., 2013). In eukaryotes mRNA, m^7G is added to the 5' end of the primary transcript through a triphosphate bridge forming a structure called cap (Cowling, 2010). These reactions occurs co-transcriptionally and are catalysed by capping enzymes. First, two enzymes (RNA 5' triphosphatase and guanylyltransferase) promote the ligation of a guanosine cap, that is methylated by another enzyme, the RNMT (RNA guanine-7- methyltransferase) forming the m^7G cap (Cowling, 2010). Presence of cap in 5' end of mRNA is essential to mRNA processing, and is involved in diverse steps including transcription, polyadenilation, splicing, transport, stability and translation. For example, in splicing 5' cap is bound by a protein complex, (cap binding complex) which interacts and recruits splicing complex components. To be translated, the majority of mRNAs requires the presence of m^7G cap structure. The process occurs by the binding of eIF4F (eukaryotic initiation factor 4) to cap structure promoting the recruitment of ribosomal components and initiator tRNA (Cowling, 2010).

Pseudouridine (ψ; 5-ribosyluracil) results from isomerization of uridine. The process involves breaking the N1-C1'covalent bound between the nitrogenous base and the ribose, followed by a 180º rotation of the nitrogenous base and formation of one C5-C1' ligation between the ribose and the nitrogenous base (Ge and Yu, 2013). This alteration generates an additional hydrogen bond donor. Isomerization reaction is catalysed in specific places by box H/ACA RNAs, with some rare cases where it is catalysed by protein pseudouridylases (Ge and Yu, 2013). Pseudouridine is found in tRNAs, rRNAs, snRNAs and mRNAs, usually located in functional regions of the molecules, for example the peptidyl transferase center (PTC), the decoding center, the A-site finger, and subunits interaction sites in rRNAs (Ge and Yu, 2013). Presence of pseudouridine increases the affinity of rRNAs and tRNAs promoting efficient translation. In mRNAs this modification affects the codon specificity and, therefore, the coding potential. The three stop codons (UAA, UAG and UGA) present in the standard genetic code, have uridine. When the uridine is modified to pseudouridine, translation does not stop, an aminoacylated tRNAs binds the modified codon and translation continues, process denominated nonsense suppression. Presence of pseudouridine in tRNAs anticodons promotes the recognition of alternate codons, and potentially the pseudouridines in codons could have a similar effect (Ge and Yu, 2013).

5-methylcytosine is found in bacterial rRNA and eukaryotic tRNA, rRNA and mRNA (Cantara et al., 2011; Machnicka et al., 2013). RNA m^5C methyltransferases form a large protein family composed of sub-families that include the RsmB family, RsmF/YebU family, Dnmt2 family, RlmI family, and Ynl022 family. Only some members identified in these families were empirically verified to 5-methylcytosine methyltransferase activity, like for example the *E. coli* RsmB and yeast Trm4 (Motorin et al., 2010). *E. coli* RsmB catalyse the methylation in naked 16S rRNA, but not in assembled 30S subunits. Yeast Trm4 catalyse the methylation in specific position of tRNA, namely positions 34, 40, 48 and 49. Positions 34 and 40 are only methylated in the precursors of the tRNA that carry the amino acids leucine and phenylalanine (Motorin et al., 2010). Biological function of m^5C is not completely understood, although, in tRNAs, the methylated cytosine appears in specific positions and seems to be involved in structural conformation and stability. Degradation of tRNA is apparently increased in molecules lacking some methylated positions. In rRNA and mRNAs the function of m^5C is not yet understood (Motorin et al., 2010).

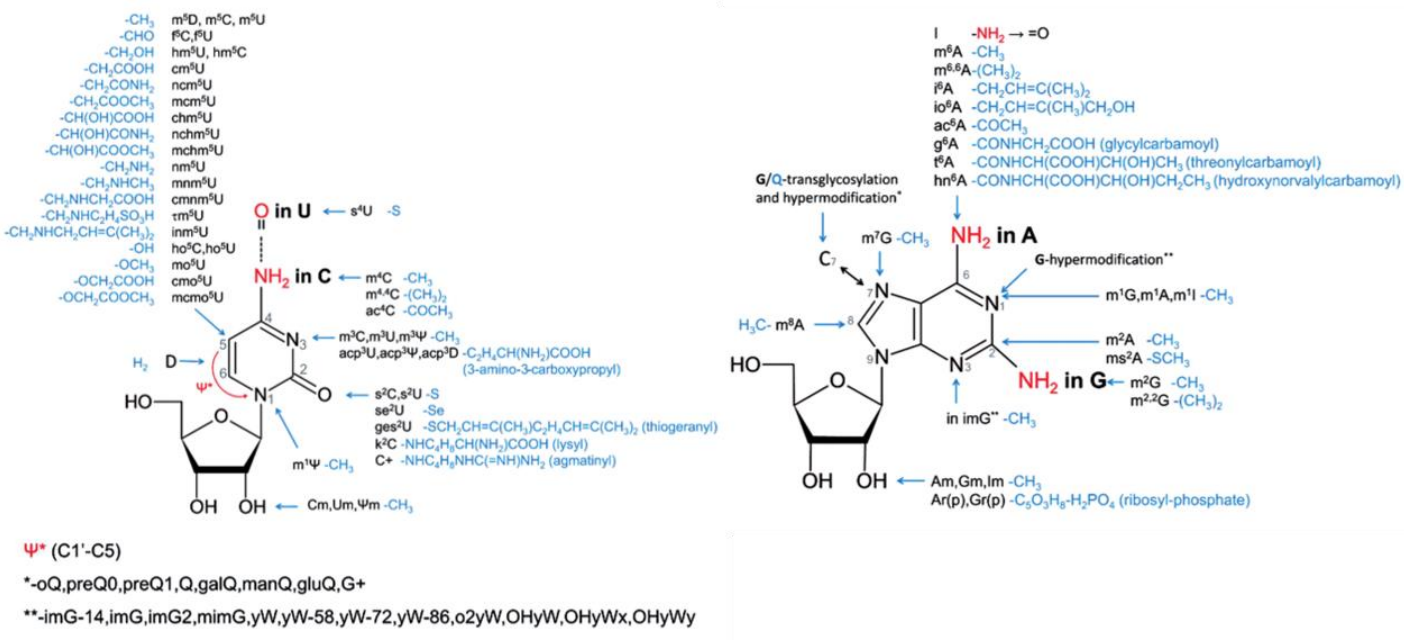


Figure 1.3. Representation of the ribonucleoside modifications found in RNA, the nomenclature is according MODOMICs database. Adapted from (Machnicka et al., 2013).

1.2.2. N6-methyladenosine

One RNA modification is N6-methyladenosine (m^6A), which differs from canonical adenosine by the presence of a methyl group (CH_3) in the nitrogen atom linked to carbon six (N6) (Figure 1.4.). The presence of m^6A on RNA molecules was detected in the 1970's, in the polyadenylated RNA fraction. (Desrosiers et al., 1974; Perry and Kelley, 1974) and was measured to be around 0.1 to 0.4% of total adenosines (Darnellt, 1975; Perry and Kelley, 1975). The methylated adenosine occurs mainly in the consensus sequence [G/A/U]-[G/A]-A-C-[U/A/C], (where the underlined A correspond to the m^6A) in a non-stoichiometric ratio. (Carroll SM, Narayan P, Rottman FM, 1990; Csepany et al., 1990; Narayan et al., 1994). In bacteria, this modification is present in rRNA, tRNA (Cantara et al., 2011) and adenines in DNA can also be methylated in N6 position (N6-methyladenine) (Wion and Casadesús, 2006).

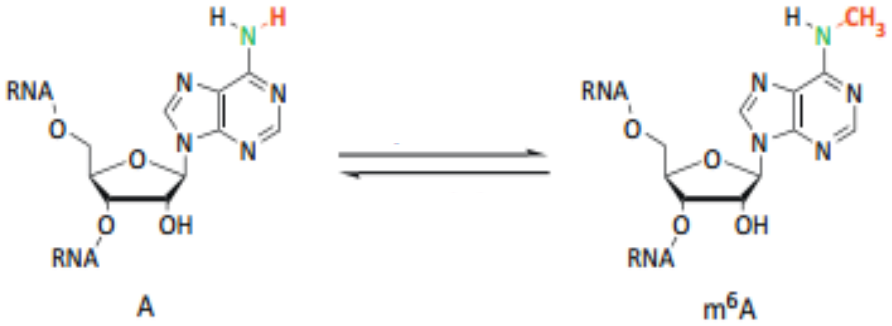


Fig 1.4. structures of the canonical adenosine (left) and the N6-methyladenosine (right). Adapted from (Jia et al., 2013).

1.2.3. Chemical reactions

This modified nucleoside is formed by the transfer of a methyl group from S-adenosylmethionine (AdoMet or SAM) to adenosine. This reaction is catalysed by RNA m^6A methyltransferases (discussed below) (Bokar et al., 1997; Liu et al., 2014). These enzymes are very similar to the DNA m^6A methyltransferases (Bujnicki et al., 2002), therefore it is likely that the catalytic mechanism is preserved. Based on the crystal structure of M.TaqI, (DNA m^6A methyltransferase from *Thermus aquaticus*) the reaction involves the interaction between the N6 of adenine to be methylated with one conserved motif (IV) in these enzymes (Goedecke et al., 2001). This interaction leads to the formation of hydrogen bounds between the hydrogens of amine group and the motif IV amino acids, promoting a hybridization change from sp^2 to sp^3 in the nitrogen atom. This leaves a free electron pair, not conjugated to the aromatic system, that attacks the methyl group in SAM, resulting in a nucleophilic substitution (S_N2) (Goedecke et al., 2001). The reverse reaction, N6-

methyladenosine to adenosine, is catalysed by RNA m^6A demethylases (discussed ahead) (Jia et al., 2011; Zheng et al., 2013a). The mechanism proposed involves the hydroxylation (ligation of one OH) to the methyl group (N6 position) forming one intermediate, N6-hydroxymethyladenosine. This intermediate can oxidize directly to adenosine (releasing a formaldehyde molecule). Besides direct oxidation of N6-hydroxymethyladenosine to adenosine, this intermediate can be oxidized to a second intermediate, N6-formyladenosine. The adenosine is produced by the oxidation of this second intermediate. These two intermediates are stable at physiological conditions and were detected in mammalian RNA, opening the possibility that they could have additional biological functions (Fu et al., 2013).

1.2.4. Enzymes and Reversibility

From HeLa cells nuclear extracts, it was possible to partially purify a multi-subunit protein complex, that catalyses the RNA adenosine methylation *in vitro*, which is dependent of SAM (Bokar et al., 1994). One of the subunits of the complex, which binds to SAM, was identified as MT-A70 (or METTL3) (Bokar et al., 1994). Cloning of MT-A70 gene (Bokar et al., 1997) allowed bioinformatics analysis by phylogenetic inference, with similar sequences found in databases being clustered in four subfamilies (A to D). Altogether they form the MT-A70 protein family (Bujnicki et al., 2002). Fold recognition analysis results indicates a structure with a consensus SAM - dependent MTase fold, characterized by α/β/α "sandwich" with seven central β-strands. In this consensus fold, several amino-methyltransferases characteristic motifs were identified. One of the most conserved is motif IV (N/D/S)-P-P-(F/W/Y/H) (Bujnicki et al., 2002). The MT-A70 protein family is very similar to the m^6A DNA Methyltransferase families as they share the SAM - dependent MTase fold and the conserved motifs (Malone et al., 1995).

Besides MT-A70, two other subunits of the m^6A RNA methyltransferase complex were revealed: METTL14 (methyltransferase like 14) and WTAP (Wilms' tumor 1 (WT1)-associating protein) (Liu et al., 2014; Ping et al., 2014; Wang et al., 2014b). METTL14 is a member of the MT-A70 family, containing the characterized motifs and possesses m^6A RNA methyltransferase activity (Liu et al., 2014; Ping et al., 2014; Wang et al., 2014b). MT-A70 and METTL14 form a heterodimer in a 1:1 ratio and the catalytic activity of the dimer is higher than the subunits alone (Liu et al., 2014; Wang et al., 2014b). Knockdown of MT-A70 or METTL14 leads to a decrease in m^6A levels and the stability of each subunit depends of the presence of the other (Liu et al., 2014; Ping et al., 2014; Wang et al., 2014b). The subunit WTAP does not have any m^6A RNA methyltransferase domain (Ping et al., 2014), and it does not have catalytic activity as an independent subunit either (Liu et al., 2014). However, WTAP interacts with the MT-A70-METTL14 heterodimer and affects its catalytic activity

in vivo (Liu et al., 2014; Ping et al., 2014). This suggests that this subunit has a regulatory role in the m^{6}A RNA methyltransferase complex. Besides these subunits, another protein that interacts with the methyltransferase complex and results in a decrease of methylation after knockdown is KIAA1429 (Schwartz et al., 2014).

Enzymes responsible for removing m^{6}A from RNA have also been recently identified. The first gene identified was the human FTO gene (which encodes for the fat and mass associated protein) (Jia et al., 2011). This protein demonstrates oxidative demethylation activity *in vitro* in dsDNA, ssDNA and ssRNA (Fu et al., 2013). Three dimensional crystal structure of FTO revealed that this protein is composed of two domains, a catalytic N-terminal AlkB-like domain and a C-terminal domain (now named FTO C-terminal domain). (Han et al., 2010). The observation that N6-methyladenosine in RNA is a main substrate of FTO (Jia et al., 2011) was a breakthrough, because for the first time it was demonstrated that this RNA modification is reversible, a key feature of regulatory mechanisms. Recently a second RNA m^{6}A demethylase, ALKBH5, was identified in humans (Zheng et al., 2013a). This protein is a member of the ALKBH family, a family of homologues of the bacterial AlkB. The demethylase catalytic activity of ALKBH5 is comparable to the FTO activity (Zheng et al., 2013a).

The known RNA methylation process is summarized in figure 1.5. The demonstration of reversibility raised the possibility of post-transcriptional gene regulation mediated by m^{6}A and other RNA modifications and leads to the introduction of the concepts epitranscriptome and RNA epigenetics (He, 2010; Saletore et al., 2012).

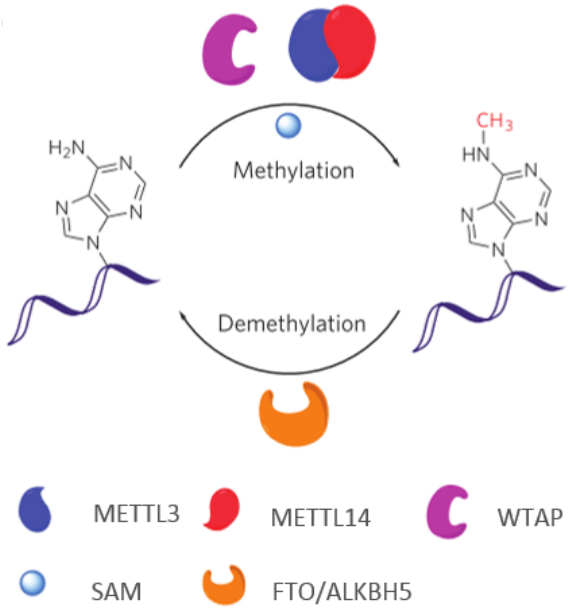


Figure 1.5. RNA m^{6}A methylation/demethylation pathway. Methylation is SAM dependent and catalysed by METTL3-METTL14 heterodimer. WTAP is a regulatory subunit of the methyltransferase dimer. Demethylation is catalysed by FTO or ALKBH5. Adapted from (Liu et al., 2014)

1.2.5. Biological functions

The relevance of this modification in biological systems can be addressed through the resulting phenotypes when the enzymes are disturbed through knockout, knockdown, mutations or overexpression. In HeLa cells, the lack of the methyltransferase MT-A70 results in apoptosis (Liu et al., 2014). Knockdown of the mouse homologue in embryonic stem cells affects the self-renewal capability (Wang et al., 2014b). In *S. cerevisiae,* m^6A RNA methylation is only found in the sexual life stage, and the MT-A70 homologue, *Ime4*, is involved in the induction of meiosis and sporulation. Mutations that result in IME4 loss of function lead to defects in sporulation (Clancy et al., 2002). Besides induction of sporulation, lineage restriction is partially dependent on the activity of this gene (Agarwala et al., 2012). *A. thaliana* homologue of MT-A70, MTA is highly expressed in dividing tissues and essential to seed development (Zhong et al., 2008). The homologue found in *D. melanogaster*, DmIme4, is essential and affects oogenesis through notch signalling (Hongay and Orr-Weaver, 2011).

The FTO gene has been associated with metabolic disorders (Wang et al., 2012) and obesity (hence the name, at **f**a**t** mass and **o**besity-associated) (Dina et al., 2007). Overexpression of FTO demethylase leads to a food intake increase and obesity in mice (Church et al., 2010) and affect hepatic metabolism in liver cell lines (Bravard et al., 2014). Also, the activity of the dopaminergic signalling in the midbrain is regulated by FTO (Hess et al., 2013). The lack of ALKBH5 demethylase in mice affects fertility due to the occurrence of apoptosis in spermatogenesis (Zheng et al., 2013a).

Besides the interference of the methyltransferases/demethylases, the use of methylation inhibitors allows to understand the biological effects of m^6A RNA methylation. Using this strategy, it was recently demonstrated that the circadian rhythm is affected by this modification (Fustin et al., 2013). Circadian rhythm are biological activities, for example activity/rest behaviour, that follow a cycle of around twenty four hours. This rhythmic behaviour in time is due to an endogenous self-sustained molecular clock that is synchronized with environmental stimulus and the time that the cycle takes is called period (Merrow et al., 2005). The inhibition of methylation as well as MT-A70 knockdown elongates the period, while the opposite effect was observed with MT-A70 overexpression, demonstrating that RNA m^6A regulates the speed of the circadian clock (Fustin et al., 2013)

The fact that the perturbation of RNA methylation/demethylation balance affects diverse biological process in different organisms suggests that this modification has an important role in biological systems influencing a wide range of pathways.

1.2.6. Molecular mechanisms and targets

To understand the molecular mechanisms and functions of m^6A RNA modification, two independent groups have identified the RNA molecules harbouring the modification in humans and mice. This was achieved by immunoprecipitation of m^6A containing RNAs, followed by high throughput sequencing (Dominissini et al., 2012; Meyer et al., 2012). In one study, m^6A was identified in 5,768 and 8,843 human (from HEK293T cell line), and mouse brain transcripts, respectively (Meyer et al., 2012). In the other study, m^6A was identified in 7,240 and 3,442 transcripts in humans (HepG2 cell line) and mice respectively (Dominissini et al., 2012). These two studies demonstrated that this RNA modification is widely distributed throughout the transcriptome (Dominissini et al., 2012; Meyer et al., 2012). Both studies revealed that m^6A is more enriched in the 3' UTR near the stop codon, although peaks could also be detected within other regions of the transcripts (Dominissini et al., 2012; Meyer et al., 2012). Moreover, the frequent motifs found in these enriched regions agree with the previous biochemically determined motif (Carroll SM, Narayan P, Rottman FM, 1990; Dominissini et al., 2012; Meyer et al., 2012; Narayan et al., 1994). Besides human and mouse, the methylated transcriptome was analysed in yeast, which identified peaks in 1183 transcripts (Schwartz et al., 2013).

In addition to the identification of the targets, the detection of m^6A in several mouse tissues, and the increase in its levels from embryo to adult development suggest that this modification is widespread and dynamic (Meyer et al., 2012). Perturbation of the RNA methylation through knockdown of RNA methyltransferases (MT-A70, Mettl14), RNA demethylase (ALKBH5) or the methyltransferase regulator subunit (WTAP) followed transcriptome analysis thought RNA-Seq or microarray reveal diverse alterations in gene expression (Dominissini et al., 2012; Ping et al., 2014; Wang et al., 2014b; Zheng et al., 2013a). These observations suggests that RNA m^6A methylation could be an epigenetic modification that regulates gene expression at post-transcriptional level.

What is the molecular mechanism of action of m^6A modification? This question does not have a unique answer. Multiple studies have suggested multiples mode of action:

1. The effect of m^6A in gene expression regulation could be mediated by readers of the epigenetic code, effector proteins that recognize and bind m^6A in RNA leading to the consequence in the transcript (for example alteration in stability). So far, three m^6A binding proteins were found in humans, YTHF1, YTHF2 and YTHF3 (Dominissini et al., 2012; Wang et al., 2014a). The YFTH2 knockdown leads to an increase in the average lifetime of target transcripts and reduced translation efficiency, suggesting that this interaction destabilizes RNAs. In accordance with this observation, this protein co-localizes with markers of P-bodies (complexes involved in mRNA storage and decay) suggesting an involvement in mRNA decay (Wang et al., 2014a). Similar effect of m^6A in reducing RNA stability was observed in another study, however the molecular mechanism is different. The

effect is not due to a recruitment of a reader protein, but through blocking the binding of HuR, a protein that stabilizes RNA (Wang et al., 2014b).

2. This modification has also been proposed to be important in splicing. The first evidence is that m^6A methyltransferases and demethylases co-localize with nuclear speckles (complexes involved in pre-mRNA processing and splicing) (Jia et al., 2011; Liu et al., 2014; Zheng et al., 2013a). Second, some introns also contain the m^6A modification (Dominissini et al., 2012; Meyer et al., 2012).

3. Knockdown of ALKBH5 demethylase leads to an increase in the rate of mRNA export from the nucleus to the cytoplasm (Zheng et al., 2013a). Together with the observation that methylation inhibitors lead to an increase in the nuclear retention of some mRNAs, namely the circadian RNAs (Fustin et al., 2013), it has been proposed that m^6A modification may be involved in the nuclear transport of RNAs.

4. Localization of m^6A in transcripts occurs mainly in 3' UTR, near the stop codon, therefore localizes close to the typical binding sites of microRNAs (3' UTR). This suggest that m^6A could interfere with the action of microRNAs (Meyer et al., 2012).

Although much remains to be studied on the exact role of m^6A in different organisms, it is clear that this RNA modification is involved in regulation of gene expression at a post-transcriptional level (Fu et al., 2014; Meyer and Jaffrey, 2014). Molecular mechanisms and functions proposed that lead the effects in transcriptome are summarized in figure 1.6.

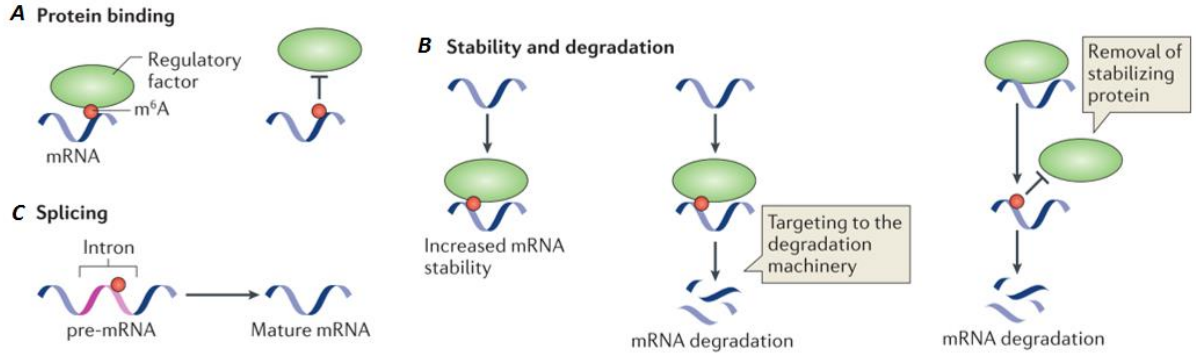


Figure 1.6. Molecular mechanisms and functions of m^6A. A) to promote or block RNA–protein interactions; B) to regulate RNA stability through diverse mechanisms. C) to influence splicing efficiency. Adapted from (Meyer and Jaffrey, 2014)

1.3. Objectives

In several eukaryotes recent evidence indicates that the presence of N6-methyladenosine (m^6A) in RNA is associated to processes of post-transcriptional gene regulation. Because in *T. brucei*, gene expression is mainly regulated at the post-transcriptional level, it is reasonable to hypothesise that m^6A RNA may be a novel mechanism of gene regulation in this parasite. In this thesis, I had four main objectives. First, I tested if this modification was present in RNA of trypanosomes. Then I determined if the levels of m^6A were regulated in different biological conditions. Third, I searched the parasite genome to identify putative enzymes that may add or remove this modification. Finally, I used genetic tools to test the role of some of these enzymes.

2. Methods

2.1. Parasite culture

T. brucei brucei SMOx (Single Marker from Oxford), a modified version of the pleomorphic strain AnTat1.1E, which contains T7 polymerase and TET repressor, was cultured at 37°C in 5% CO_2 in HMI-11 medium (HMI-9 medium (Hirumi, 1989) without serum plus) supplemented with puromycin at 0,1 µg/mL (Invitrogen cat: ant-pr-1). Cell density was maintained below 1 x 10^6 cells/mL (usually cultures were passed when they reached around 0,5 x 10^6 cells/mL, unless stated otherwise). Knockout cell lines were cultured under the same conditions, with the additional drugs supplemented to the medium (G418 at 2,5 µg/mL (Invitrogen cat: ant-gn-5) and hygromycin at 5 µg/mL (Invitrigen cat: ant-hm-1). Procyclic culture was cultured at 27°C in 5% CO_2 in DTM medium (Vassella and Boshart, 1996). Cell density was maintained between 1 – 10 x 10^6 cells/mL.

2.2. Differentiation

Bloodstream culture at a density around 1-2 x10^6 cells/mL were centrifuged at 1800 rpm, for 10 minutes at room temperature. Cells were re-suspended in DTM (Vassella and Boshart, 1996), at a cell density of 1-4 x10^6 cells/mL with freshly prepared cis-aconitate (Sigma-Aldrich cat: A3412) at a final concentration of 6 mM. Culture was incubated at 27°C and 5% CO_2.

The efficiency of differentiation was confirmed by the expression of procyclin through flow cytometry. 0,5 million cells were centrifuged at 2800 rpm for 10 minutes at 4°C. The cells were transferred to an eppendorf tube and centrifuged again at 2800 g for 4 minutes at 4°C. Cell pellet was re-suspended in FITC conjugated anti-procyclin antibody (Cedarlane's cat: CLP001F) diluted in HMI-11 in a 1:500 dilution and incubated 15 minutes at 4°C. After incubation, the cells were spun at 2800 g for 4 minutes at 4°C and washed three times with HMI-11. After washing, the cells were re-suspended in HMI-11 and analyzed with BD LSR Fortessa (Becton Dickinson cytometers).

2.3. RNA extraction

About 50 million parasites were centrifuged at 1800 rpm for 10 minutes at 4°C. Cells were transferred to eppendorf tubes and washed twice with TDB (5 mM KCl, 80 mM NaCl, 1 mM $MgSO_4$, 20 mM Na_2HPO_4, 2 mM NaH_2PO4, 20 mM glucose, pH 7.4), except for procyclic cells that were washed with PBS (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 2 mM KH_2PO_4, pH 7,4). Cell pellets were re-suspended in TRIzol (Life Technologies, cat: 15596) and RNA was extracted according to the manufacturer's instructions. RNA samples were quantified with NanoDrop 2000 Spectrophotometer (Thermo Scientific). 0,5U of DNaseI (New England Biolabs, NEB, cat: M0303S) were added per µg of RNA in DNaseI buffer (NEB cat: B0303S) and incubated at 37°C for 45 minutes. After incubation, EDTA was added to a final concentration of 5 mM and heat inactivated at 75°C for 10 minutes.

2.4. DNA extraction

About 50 million parasites were spun at 1800 rpm for 10 minutes at 4°C. Cells were transferred to eppendorf tubes and washed twice with TDB, except for procyclics which were washed with PBS. Cell pellets were re-suspended in DNAzol (Life Technologies, cat: 10503-027) and DNA was extracted according to the manufacturer's instructions. DNA samples were quantified with NanoDrop 2000 Spectrophotometer (Thermo Scientific). DNA was treated with RNase A (Carl Roth cat: 7156) (100 µg/mL) at 37°C for 2 hours.

2.5. Cloning

All pIV plasmids possess a common backbone, with an ampicillin resistance gene and an origin of replication for bacteria, which was obtained by digestion of pFAB2 (from Luisa Figueiredo Lab) with NotI-HF (NEB cat: R3189S) and KpnI-HF (NEB cat:R3142S) (CutSmart™ Buffer at 37°C, 5 Units per µg of DNA, for 4 hours). Inserts were amplified with Phusion High-Fidelity DNA Polymerase (Thermo Scientific, cat: F5305) according to the manufacturer's instructions. Inserts that correspond to genomic recombination sites were amplified from *T. brucei* genomic DNA. G418 and hygromycin resistance genes were amplified from pLF13 plasmid and p2T7TA, respectively (Luisa Figueiredo Lab). Primer sequences are described in a table in annexes. Digestion and PCR amplification products were purified through gel extraction with QIAquick Gel Extraction Kit (Quiagen cat: 28706) according to the manufacturer's instructions. After extraction, a sample of each

of the fragments was run in agarose gel to confirm its purity. Fragments designed for each plasmid were cloned with the In-Fusion® HD Cloning Kit (Clontec cat: 639649). Briefly, 100 ng of vector backbone was mixed with 100 ng of an equimolar mixture of inserts. The reaction was incubated at 50 °C for 15 minutes. This mixture was then diluted 1:5 with TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8.0) and used to transform competent JM109 *E. coli* bacteria (Promega cat: L2001). 50 µL of *E. coli* were mixed with 5µL of diluted In-Fusion reaction and incubated at 42°C for 45 seconds and immediately put on ice for 2 minutes (heat shock). Then, 950 µL of SOC medium at room temperature was added and incubated at 37°C for 1 hour, under gentle agitation. Bacteria were centrifuged at 8000 rpm for 2 minutes, re-suspended in 150 µL of SOC medium, plated in LB agar supplemented with ampicillin (100µg/mL) and incubated at 37°C overnight. Colonies obtained were grown in LB liquid medium overnight and the plasmid extracted with Fast-n-Easy Plasmid Mini-Prep Kit (Jena Bioscience cat: PP-204L) according to the manufacturer's instructions. Isolated plasmids were digested to confirm if the cloned plasmids correspond to the desired plasmids. All enzymes used to digest the plasmids were from NEB. pIV1 was digested with EcoRI-HF (cat: R3101S) and KpnI-HF (cat: R3142S), pIV2 was digested with BamHI-HF (cat: R3136S) and SacI-HF (cat: R3138S), pIV3 was digested with BsaI-HF (cat: R3535S) and KpnI-HF (cat: R3142S), pIV4 was digested with EcoRI-HF (cat: R3101S) and NotI-HF (cat: R3189S), pIV5, was digested with BamHI-HF (cat: R3136S) and KpnI-HF (cat: R3142S) , and pIV6 was digested with EcoRI-HF (cat: R3101S) and NotI-HF (cat: R3189S). All reactions were performed in CutSmart™ Buffer at 37°C, 5 Units per µg of DNA, for 4 hours. To confirm that the constructs were correctly cloned, the plasmids were sequenced in STAB VIDA (primer table in annexes).

2.6. Transfections

pIV plasmids were digested with NotI-HF (cat: R3189S) and KpnI-HF (R3142S) (from NEB), in CutSmart™ Buffer at 37°C, 5 Units per µg of DNA, for 4 hours (with the exception of pIV1 that was digested with KpnI-HF and NdeI (cat: R01115) in CutSmart™ Buffer in the same conditions). Digestion products were purified by ethanol precipitation. 1/10 volume of sodium acetate (3M, pH 5,2) and 2,5 volumes of ice cold ethanol were added to each digestion reaction and incubated for 1 hour at -80°C. After incubation, DNA was centrifuged at 13 200 rpm for 30 minutes at 4°C. DNA pellets were washed with 70% ethanol and centrifuged at 13 200 rpm for 10 minutes. Supernatants were discarded inside the flow chamber for sterility maintenance. When pellet was dry, DNA was ressuspended in 10 µL of mili-Q water (inside the flow chamber).

Transfections were done by electroporation (Burkard et al., 2007). Briefly, 30 million cells were spun at 1800 rpm for 10 minutes at room temperature and re-suspended in 90 µL of "Roditi

buffer" (90mM Na-PO_4, 5mM KCL, 50mM HEPES, 0.15mM $CaCl_2$, pH 7,3). Purified DNA sample (5-10 μg in 10 μL) was added to the cells and the mixtures were transferred to 2mm gap cuvettes (BioRad Gene Pulser/MicroPulser Cuvettes). Electroporations were performed with X-001 program in the Amaxa Nucleofector (Lonza Cologne AG, Germany) and immediately parasites were diluted in HMI-11 at 37°C and plated in three plates of 24 wells, with 10-fold dilution to each plate. Plates were incubated at 37 °C and selection drugs were added to the wells 8-16 hours later. Obtained clones were genotyped through amplification of the locus and the recombination regions, with Taq DNA Polymerase (Thermo Scientific), or Phusion High-Fidelity DNA Polymerase (Thermo Scientific) according to the manufacturer's instructions (primers' table in annexes).

2.7. Immunoblot

RNA samples were denatured by incubation at 50°C for 20 minutes and placed immediately on ice. DNA samples were denatured by boiling for 10 minutes and placed immediately on ice. One volume of 20X SSC (3 M NaCl; 0,3 M sodium citrate; pH 7) was added to each sample. Nucleic acid samples were spotted in a positive charged nylon membrane (Amersham Hybond™-N+) and fixed by UV crosslink (Stratalinker® UV Crosslinker, autocrosslink mode: 120 mJ/cm2 with 254 nm). Membranes were stained with methylene blue (0,02% in 0,3 M sodium acetate pH 5,5) and washed with milli-Q water to reduce background. Staining was removed by washing with distaining solution (0,2X SSC, 1% SDS) and washed twice with PBS/Tween 0.1% for five minutes each wash. After that, the membrane was blocked in 5% milk in PBS/Tween 0.1% for one hour. Blocked membrane was incubated with anti-m^6A antibody (Millipore, rabbit polyclonal cat: ABE572) in a 1:1.000 dilution in 3% milk in PBS/Tween 0.1% overnight at 4°C. Membrane was washed three times with PBS/Tween 0.1% for 5 minutes each wash and incubated 45 minutes with secondary anti-rabbit IgG antibody (HRP-Linked, GE Healthcare) in a 1:10.000 dilution. Membrane was washed three times with PBS/Tween 0.1% for 5 minutes each wash and was developed by incubation for 4 minutes with Plus-ECL reagents (enhanced chemiluminescence, PerkinElmer). Luminescence was captured in ChemiDoc XRS System (Bio-Rad) for 30 minutes.

2.8. EpiQuick m^{6}A Quantification

Detection and quantification of m^{6}A was done with EpiQuik™ m^{6}A RNA Methylation Quantification Kit (EPIGENTEK cat: P-9005) according to the manufacturer's instructions. Briefly, nucleic acids were incubated in individual wells with binding solution at 37°C for 90 minutes, washed three times with diluted wash buffer, and incubated with the capture antibody (1:1000 dilution) for one hour at room temperature. The wells were washed three times with diluted wash buffer, and incubated with detection antibody (1:2.000 dilution) for 30 minutes at room temperature. After four washes with diluted wash buffer, wells were incubated with enhancer solution (1:5.000 dilution) for 30 minutes at room temperature. Wells were washed five times as before and developed for 5 minutes, when the reaction stopped. Absorbance was measured at 450 nm in a plate reader (Tecan, model Infinite M200).

2.9. Bioinformatics

Sequences search of putative *T. brucei* RNA m^{6}A methyltransferases and demethylases was performed with BLAST in TriTrypDB (Aslett et al., 2010). Putative *T. brucei* RNA m^{6}A methyltransferases and demethylases domains search was done in Pfam database (Finn et al., 2014). Sequences from *T. brucei* putative candidates were obtained from TriTrypDB and the remain sequences (human ALKBH1-8, E. coli AlkB and eukaryotic containing Pfam domain PF10237 sequences) were obtained from UniProt (Apweiler et al., 2004) . Domains from the previous described sequences were identified with Pfam and manually curated. Multiple sequence alignments were done with Clustal Omega (Sievers et al., 2011), manually edited with Jalview (Waterhouse et al., 2009) and colored with Chroma (Goodstadt and Ponting, 2001) with black and white default parameters (table of color and corresponding properties in annexes). Secondary structure was predicted with PsiPred (McGuffin et al., 2000) and Jpred (Cole et al., 2008).

3. Results

In this project, I hypothesized that N6-methyladenosine modification could be an important mechanism to regulate gene expression in *T. brucei*. For that, I first tried detect m^6A with an immunoblot assay (Section 3.1). Next I used a specific kit to detect and quantify this modification in two different stages of the life cycle of *T. brucei* (Section 3.2). In Section 3.3 and 3.4, I characterized the levels of m^6A levels during differentiation from bloodstreem to procyclic forms and in stress conditions respectively. Finally, I screened for putative methyltransferase or demethylase enzymes and I tested their role by generating knockout mutants (Section 3.4).

3.1. Immunoblot detection of m^6A in T. brucei RNA

My first goal was to test if N6-methyladenosine was present in nucleic acids in *T. brucei* of bloodstream forms (BSF) and procyclic forms (PC). For this, I tried to use an immunoblot method that applies an available m^6A specific antibody. The method consists in the direct application of nucleic acids to a membrane (dot blot). After the ligation of the nucleic acids, the membrane is blocked and incubated with an anti-m^6A antibody. Signal of bound anti-m^6A antibody is generated by a secondary antibody linked to horseradish peroxidase and exposed to ECL substrate. Because m^6A is present in RNA of some eukaryotes and the DNA of some prokaryotes, I first tested the presence of this modification in both type of nucleic acids of *T. brucei*. The negative control used consisted in a synthetic oligo that does not contain m^6A. I also used additional RNA and DNA positive controls from biological samples: an RNA positive control extracted from mouse liver (C57BL/6 mouse), described as a tissue rich in m^6A (Meyer et al., 2012), and a DNA positive control from *E. coli* (JM109 strain, Dam^+), which contains this modification in its chromosomal DNA (Ratel et al., 2006). My samples from *T. brucei* were total RNA and genomic DNA. RNA samples were treated with DNaseI to remove possible DNA contamination, and DNA samples were treated with RNase A to remove possible RNA contamination. The presence of the nucleic acids in the membrane was controlled by staining with methylene blue. The result of the immunoblot assay was not conclusive (table 3.1, figure 2.1.).

Table 3.1. Samples spotted in immunoblot membrane to detect m^{6}A in *T. brucei* nucleic acids

SAMPLE	DESCRIPTION
1	Negative control, synthetic oligo
2	Mouse liver RNA, RNA positive control
3	*E. coli* DNA, DNA positive control
4	*T. brucei* bloodstream RNA
5	*T. brucei* bloodstream DNA
6	*T. brucei* procyclic RNA
7	*T. brucei* procyclic DNA

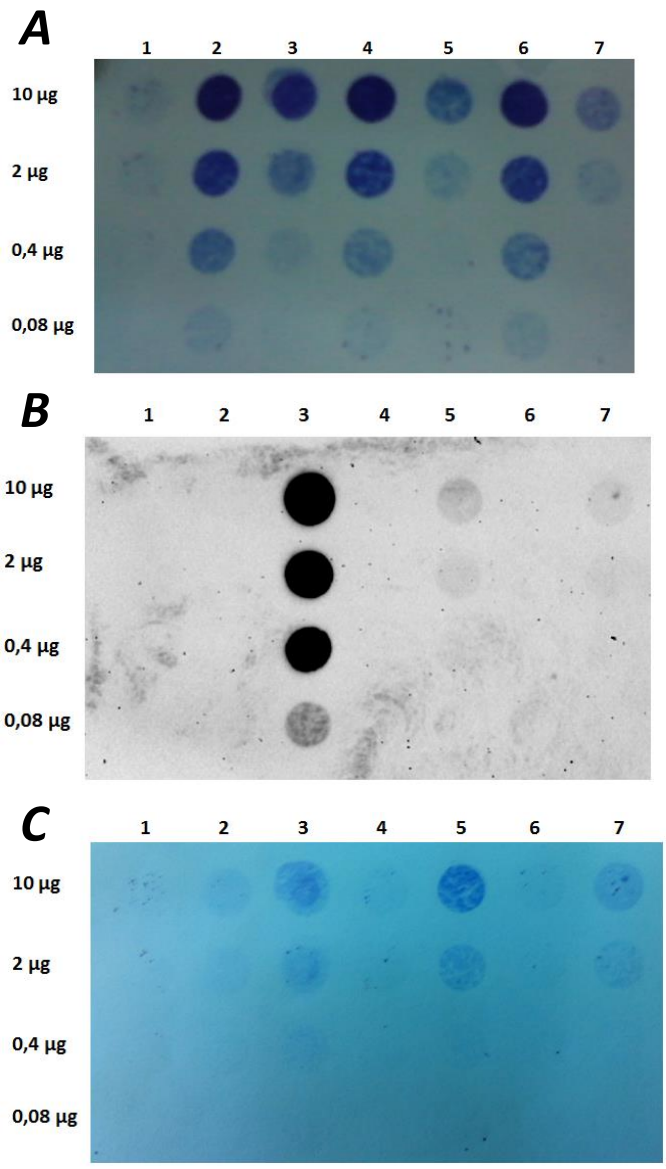


Figure 3.1. Immunoblot to detect m^{6}A A) Methylene blue staining after loading and UV crosslink of nucleic acid samples; B) Anti-m^{6}A labelling; C) Methylene blue staining after development. The number of each sample in the immunoblot is indicated in table 3.1.

Samples were spotted in 5-fold serial dilutions from 10 µg to 0,08 µg. The first picture is methylene blue staining after the samples have been spotted and cross-linked to the membrane (Fig 3.1A). First observations are that the negative control was not attached to the membrane and RNA samples become bound to the membrane more efficiently than DNA samples. Heterogeneity in the intensity of the spots staining with the same amount of sample indicates that the ligation of the nucleic acid to the membrane is not very efficient and part of the sample was lost.

Blot was incubated with anti-m^6A primary antibody, then with a secondary antibody and finally signal was developed by ECL (Figure 3.1B). The negative control did not generate signal, as expected, considering that in the previous methylene blue staining, it was observed that the synthetic oligo did not become bound to the membrane. Therefore, the negative control is not valid as it did not give the nucleic acid background signal. RNA positive control did not generate any signal, therefore it is not possible to draw any conclusion about the *T. brucei* RNA samples that did not generate signal. *T. brucei* DNA samples, both bloodstream and procyclic forms, generated signal, however a much more faint signal than the DNA positive control. Bloodstream form sample generated signal in the titres of 10 µg and 2 µg. Procyclic form only generated signal in the 10 µg titre. Due to the lack of negative control, it is not possible to understand if the faint signal observed in the *T. brucei* DNA samples are the nucleic acid background or a specific signal.

After development the membrane was stained again with methylene blue to control for the possible loss of sample during the development (Figure 3.1C). The staining reveals that all the RNA samples (that are present in the membrane in the first staining) were almost completely lost. This fact explains the lack of signal in the RNA positive sample and possibly in *T. brucei* samples. In DNA samples the intensity of the staining is lower than the methylene blue staining before the development. Therefore, DNA samples were also lost during the development but not completely.

Due to the inefficient binding of nucleic acids to the membrane, this experiment was inconclusive and thus it was not possible to determine if *T. brucei* nucleic acids harbor m^6A. I repeated this experiment several times by changing different parameters. First, I tried to improve the binding efficiency by cross-linking the membrane by UV exposure. UV cross-linking was performed with the auto crosslink default parameters in Stratalinker® UV Crosslinker (254 nm; 120 milijoules/cm^2). I tried to use a different parameter to UV crosslink (254 nm; 70 milijoules/cm^2), which is indicated in the membrane manual as optimal for the membrane. Additionally I tried another apparatus (uvitec UV crosslinker CL-E508) with the same parameters (254 nm, 120 milijoules/cm^2). The result was identical to described above. Another possibility was that the membrane was not in the proper conditions, so I tried a new membrane (identical to the old, but opened recently), however the result was identical. Because the loss of sample is much more pronounced in RNA samples, one possible problem, at least in RNA samples, could be the presence of RNases contamination in the development buffers, even

though the use of RNase free water to make the buffers. I replace all buffers to new ones, and supplement with RNaseOUT (10 U/mL in all buffers and 40 U/mL in the antibody solution). New buffers and the supplement with RNaseOUT did not prevent the loss of RNA from the membrane.

Detection of m^6A in *T. brucei* nucleic acid samples with immunoblot method failed. The method needs to be optimized, it is necessary to solve these problems to allow the detection with the immunoblot technique.

3.2. Detection and quantification of m^6A in T. brucei (in bloodstream and procyclic forms)

To test if *T. brucei* has the m^6A modification in DNA or RNA I next tried an alternative approach: a commercially available kit, *EpiQuik™ m^6A RNA Methylation Quantification Kit*. Briefly, the kit consists in an ELISA-like assay in which nucleic acids are bound to the wells and incubated with an anti-m^6A antibody. After washing to remove unbound antibody, it is incubated with a secondary antibody. The quantification is achieved through incubation with substrate that generates a blue color product. The intensity of the color generated is measured in a plate reader. The intensity of the color is proportional to the amount of m^6A in each well. The kit allows quantitative measurements through comparison of the samples intensity with a standard curve, or alternatively only a semi-quantitative detection, comparing the intensity of the sample with one positive control. The kit provides a negative and a positive control, which consists of synthetic oligos without or with m^6A, respectively. As for the immunoblot assay, I used controls from biological samples: mouse liver RNA (C57BL/6 mouse) and *E. coli* (JM109 strain, Dam^+). *T. brucei* samples consisted of total RNA (DNaseI treated) and DNA (RNase A treated), both from bloodstream and procyclic forms.

In the first preliminary experiment I tested if m^6A was present in the two types of nucleic acids (RNA and gDNA) from the two life-cycle stages (bloodstream and procyclic forms). Signals generated in the four *T. brucei* samples had an intensity that was double or more than the negative control (Table 3.2.). Because in this preliminary assay, the intensity of each sample was compared to only one positive control, and not to a standard curve, the calculation of m^6A/A (%) (fifth column of table 3.2.) is a semi-quantitative estimate.

Table 3.2. Detection of m^{6}A in RNA and DNA samples. OD represents optical density, NC represents negative control.

SAMPLE	LIFE STAGE	OD	OD – NC OD	M^{6}A/A (%)
Negative control	-	0,069	0	0
Positive control	-	1,155	1,086	100
T. brucei RNA	Bloodstream	0,186	0,117	0,22
T. brucei DNA	Bloodstream	0,136	0,067	0,12
T. brucei RNA	Procyclic	0,271	0,202	0,37
T. brucei DNA	Procyclic	0,195	0,126	0,23
Mouse liver RNA	-	0,440	0,371	0,684
E. coli DNA	-	0,572	0,503	0,926

This data provides the first evidence that m^{6}A modification exists in RNA and DNA from *T. brucei*. Together, with the spliced leader cap modifications (Bangs et al., 1992; Zamudio et al., 2009) and the 5-methylcytosine in tRNA (Militello et al., 2014), m^{6}A constitutes a novel modification of the epitranscriptome of *T. brucei*. It has been previously shown that DNA in *T.brucei* contains 5-methylcytosine (Militello et al., 2008) and base J (Gommers-ampt et al., 1993). Here, I present the first evidence of m^{6}A as a new DNA modification in *T. brucei* DNA. This is in agreement with a previous study suggesting that m^{6}A could exist in *T. cruzi* DNA (Rojas and Galanti, 1990).

The proposed hypothesis was that the m^{6}A modification in RNA was associated with the regulation of gene expression at post-transcriptional level, therefore in the remainder of my thesis I focused on the modification in RNA. Future studies will characterize this modification in DNA.

I tested if the levels of m^{6}A were identical in two different stages of the life cycle, the mammalian bloodstream form and the insect procyclic form, measuring the levels in RNA and compared to a standard curve. In bloodstream form the m^{6}A levels are 0,34% (±0,07%) and in procyclic RNA are 0,51% (±0,24%), relative to total adenosines. Although there is a tendency for RNA from procyclic to have higher values of m^{6}A modification, the differences were not statistically different (p= 0,400, Mann-Whitney U test) (Figure 3.2.), perhaps because the number of biological replicates is only three.

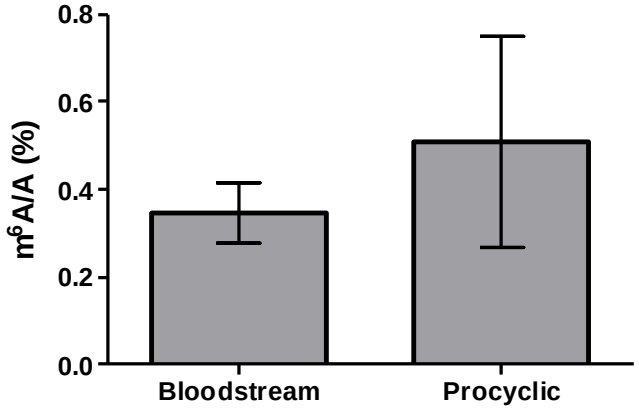


Figure 3.2. Levels of m^{6}A in total RNA from *T. brucei* AnTat 1.1 SMOx in bloodstream and procyclic forms. Measurements result from three independent biological replicates. The data are presented as percentage of m^{6}A relative to total A in RNA.

To test if the signals generated are specific to RNA molecules and not other component of the sample, for example DNA contamination, RNA samples were treated with RNase A and purified over a column (to remove small RNA fragments from RNase A digestion). If the signal in the sample is RNA specific it should disappear after RNase treatment. If is due to other component of the sample (and is not lost in the purification) the signal should remain. After treatment and purification of one RNA sample from bloodstream form and another from procyclic form, the signals were almost completely lost. RNAse-treated bloodstream RNA generated a signal of 0,04% (m^{6}A/A) and the procyclic to 0,06% (m^{6}A/A), around tenfold less intense than the original samples. With this experiment, I conclude that the signal I detect with the anti-m^{6}A antibody in RNA sample is indeed a result from a modification of the RNA, and not of a contaminant.

3.3. Levels of m^{6}A during differentiation

Between bloodstream and procyclic form, 33% of the transcripts are differentially regulated (Nilsson et al., 2010) and they seem to be regulated in regulons, clusters of transcripts with the same expression pattern, during differentiation (Queiroz et al., 2009). I hypothesized that the m^{6}A pattern could change during differentiation as a mechanism to regulate the transcript levels.

Differentiation from bloodstream form to procyclic form *in vitro* is a fair approximation of the natural process that occurs *in vivo*. This process can be induced by adding cis-aconitate to the medium and lowering the temperature to 27ºC (Czichos J, Nonnengaesser C, 1986). Using this methodology, the levels of m^{6}A were quantified at different time points throughout differentiation (figure 3.4.).

Efficiency of the differentiation was controlled by the expression of procyclin, through flow cytometry (figure 3.3.), evaluated 72 hours after induction of differentiation. Negative control correspond to a bloodstream culture and positive control correspond to a previously established procyclic culture. As expected, procyclin staining signal is negative in bloodstream forms and positive in procyclic forms. Considering the intensities of these controls, a threshold to analyze the differentiated culture was defined, where quadrant four defines the negative cells and quadrant three the positive cells. Parasites obtained 72 hr after initiating the differentiation protocol were also subjected to procyclin staining. 93.3% of cells showed positive procyclin staining, indicating that the differentiation protocol worked well.

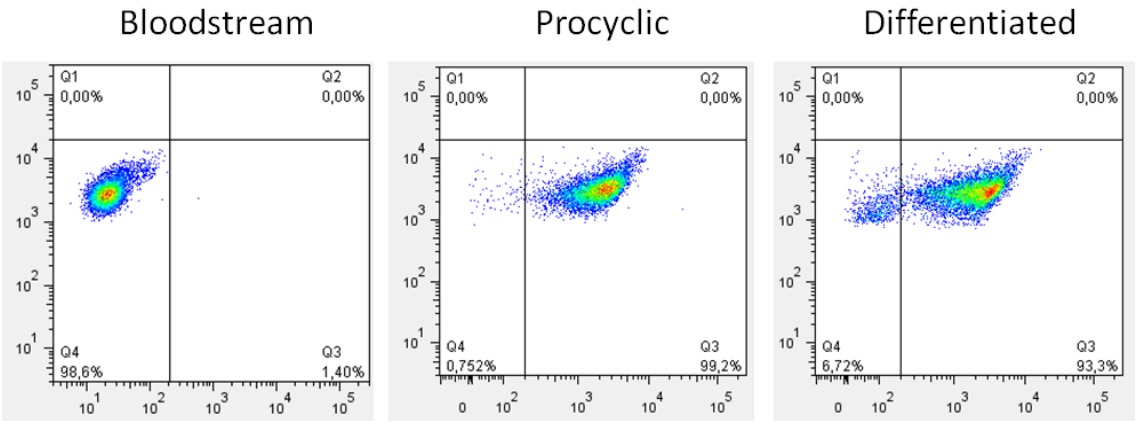


Figure 3.3. Cells were stained with FITC conjugated anti-procyclin antibody and analyzed by flow cytometry. Intensity of procyclin is represented in horizontal axis. The proportion of cells in each quadrant (Q) is indicated as a percentage of total events analyzed. Q4 is considered negative population and Q3 positive population.

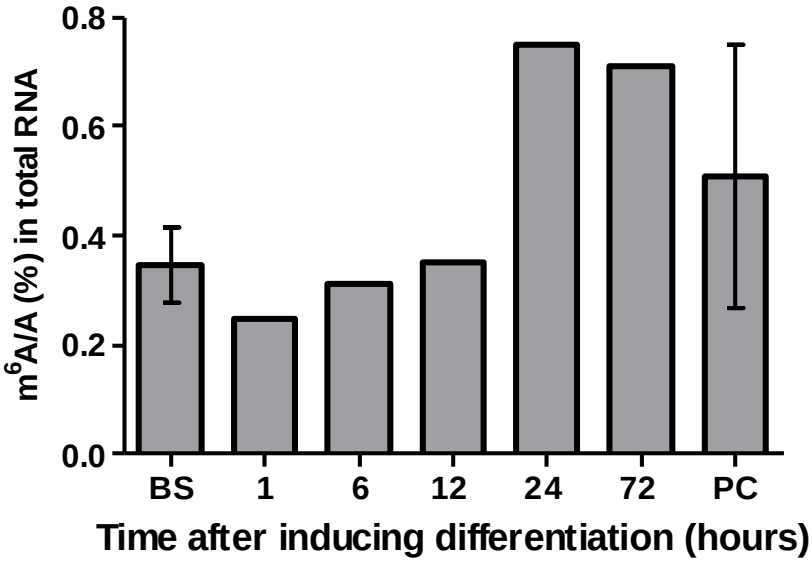


Figure 3.4. Levels of m^{6}A in total RNA during *in vitro* differentiation of AnTat 1.1 SMOx, from bloodstream form to procyclic form. Bloodstream levels and procyclic levels from three independent biological replicates, and differentiation time points result from one experiment.

When m^6A levels were quantified at different time points of differentiation, I observed that within the first 12 hours of differentiation, m^6A levels remain relatively constant and similar to bloodstream form (~0.3%). At 24hr, I observed a sharp increase in m^6A levels to around 0.7% and the levels remained until the end of the experiment at 72hr. To generate more robust evidence, the experiment needs to be repeated.

3.4. Levels of m^6A in density stress condition

T.brucei response to stress conditions involves changes in gene expression, and evidence reveals that this parasite is able to sense and respond to cell density (de Nadal et al., 2011; Reuner et al., 1997). Considering this evidence, I hypothesize the involvement of m^6A in this specific response. This involvement can be determined through an alteration in the levels of m^6A in RNA with different cell densities. For this purpose, *T.brucei* parasites were cultivated at the typical density ($0{,}5\times10^6$ cells/mL) and in higher densities ($1{,}5\times10^6$ cells/mL and $3{,}5\times10^6$ cells/mL) and the m^6A levels quantified in total RNA of these cultures (figure 3.5.).

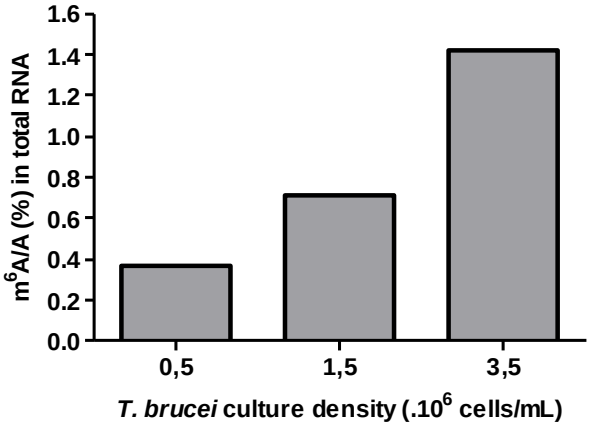


Figure 3.5. Levels of m^6A in total RNA of *T. brucei* bloodstream form from cultures with different cell densities. Result from one single experiment.

This preliminary test shows that the levels of m^6A in total RNA are proportional to cell density. Further measurements should be made to give more robustness and confidence.

3.5. Characterization of putative m^{6}A methyltransferase and demethylases enzymes

The presence of m^{6}A in the RNA of *T.brucei*, indicates that its genome should encode enzymes responsible for RNA adenosine methylation (m^{6}A methyltransferase) and demethylation processes (m^{6}A demethylase). In order to identify candidate enzymes, an *in silico* approach was used. After identification of putative genes, knock out cell lines were generated and the m^{6}A levels in these cell lines were measured.

3.5.1. Identification of putative RNA m^{6}A methyltransferase

In other eukaryotes, RNA m^{6}A methyltransferases belong to the MT-A70 protein family (Bujnicki et al., 2002) . I hypothesized that the *T. brucei* putative m^{6}A methyltransferases could be a homologue of MT-A70 protein-encoding genes. To identify putative candidates, I first used the BLAST tool (Basic Local Alignment Search Tool), a method to compare a query sequence to a database and find similar sequences there (Altschul et al., 1990, 1997) in TriTrypDB (Database of kinetoplastid parasites genomic information, in which several genomic and functional data sets are available (Aslett et al., 2010)). As a first query, I blasted the amino acid sequence of several RNA m^{6}A methyltransferases from different species (*H. sapiens*, *M. musculus*, *A. thaliana*, *D. melanogaster* and *S. cerevisiae*). However, this query did not retrieve any significant hits (arbitrary defined as a hit with e-value less than 1.E^{-3}, e-value correspond to the expected number of sequences with an alignment score identical or bigger than the hit by chance). Due to the modular evolution of proteins, analysis of proteins domains as discrete units (Bhattacharyya et al., 2006; Moore et al., 2008) combined with the application of Hidden Markov Models (HMM), could be more sensitive than BLAST (Eddy, 2011; Madera and Gough, 2002). Pfam database applies this method to organize proteins domains into families (Finn et al., 2014). However, according to this database, MT-A70 family (PF05063) does not have any *T.brucei* protein annotated.

The MT-A70 protein family is very similar to the family of the prokaryotic DNA m^{6}A methyltransferases. Both possess the Rossman fold domain and similar conserved motifs, despite the order of the motifs being different due to sequence circular permutation (Bujnicki et al., 2002; Malone et al., 1995). I hypothesized that genes containing domains typical of DNA m^{6}A methyltransferases, could be putative RNA m^{6}A methyltransferases.

However, neither BLAST in TriTrypDB nor Pfam searches with representative sequences of the three classes of DNA m^{6}A methyltransferases (Dam α group, M.RsrI β group and M.TaqI ϒ group

(Malone. 1996)) lead to the identification of putative sequences in *T.brucei*. The lack of similar candidates detectable by BLAST and HMM searches could be the result of a highly divergent evolution of this species. Alternatively, the methyltransferase reaction in *T. brucei* may be catalyzed by divergent enzymes.

As previously described, the sequences of RNA and DNA m^6A methyltransferases have a set of relatively conserved motifs, among which one of the most conserved is called motif IV (Bujnicki et al., 2002; Malone et al., 1995) .This motif interacts directly with N6 (the - nitrogen atom in m^6A) promoting the transfer of the methyl group (Goedecke et al., 2001). Therefore I search for this motif as an alternative strategy to find putative m^6A methyltransferases genes. The Prosite database has a description of patterns/motifs in proteins families (Hulo et al., 2006), where it is possible to find the motif IV pattern (PS00092) described with the [LIVMAC]-[LIVFYWA]-{DYP}-[DN]-P-P-[FYW] regular expression, and annotated as "N-6 Adenine-specific DNA methylases signature". In TriTrypDB, the gene Tb927.7.6620 has this regular expression in the form of VVADPPF, and it is annotated as "hypothetical protein, conserved". This protein has a domain annotated as "Probable N6-adenine methyltransferase" (PF10237, e-value of $9.E^{-37}$), due to the presence of the conserved motif IV in this family. This family is distinct from the MT-A70, MethyltransfD12, N6_N4_Mtase, and Methyltransf_26 families, which contain the described RNA and DNA m^6A methyltransferases. However, these five families are grouped in the same Pfam clan (CL0063).

I next aligned the sequence of Tb927.7.6620 to other members of its family (Pfam PF10237 family) using Clustal Omega algorithm (Sievers et al., 2011) and predicted the secondary structure with Jpred (Cole et al., 2008) and Psipred (McGuffin et al., 2000) algorithms. Conserved regions are observed in the multiple sequence alignment (MSA) where, as expected, motif IV was identified. However, identification of other motifs was not so obvious, due to the low sequence conservation within the motifs (figure 3.6.). The predicted secondary structure of Tb927.7.6620 corresponds to the Rossmann fold domain, due to the β-α-β repeats observed (Michael G. Rossmann, 1974), this structural domain is found in m^6A methyltransferases (Bujnicki et al., 2002; Malone et al., 1995).

Overall, the *in silico* analysis have allowed me to identify one putative gene, Tb927.7.6620, that may encode for the RNA m^6A methyltransferase.

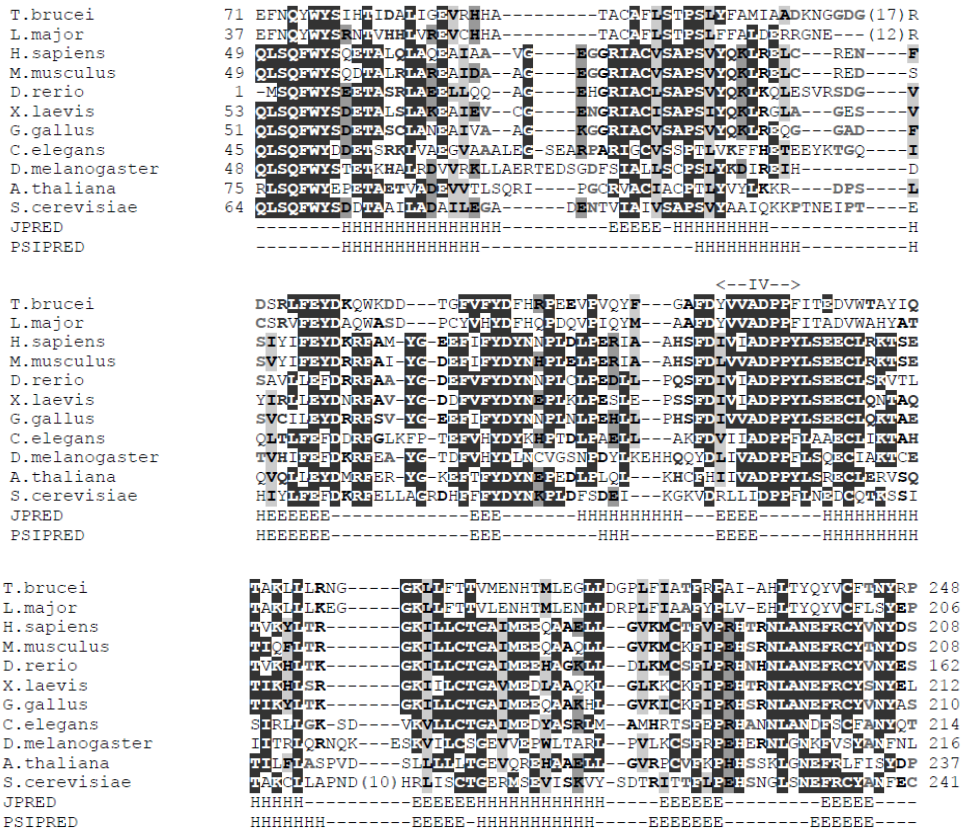


Figure 3.6. Multiple sequence alignment of sequences of the putative N6-adenine methyltransferase family; Conserved motif IV is marked above the corresponding amino acids; in the secondary structure prediction H represents α-helix and E represents β-strand. Numbers in parentheses are positions not represented. Numbers at the beginning and end represent the start and end positions of each sequence on the MSA.

3.5.2. Identification of putative RNA m^{6}A demethylases

To identify putative m^{6}A demethylases, a similar *in silico* strategy was adopted. The enzymes known to catalyze this reaction in other eukaryotes are FTO (Jia et al., 2011) and ALKBH5 (Zheng et al., 2013a) . The search for sequences similar to FTO, through BLAST in TriTrypDB and domains in Pfam did not lead to any significant hit. However, blasting ALKBH5 as a query in TriTrypDB retrieved one significant hit: Tb927.5.980 (e-value of 4E^{-05}), annotated as "hypothetical protein, conserved". This gene encodes a protein that contain a domain from the "2OG-Fe(II) oxygenase superfamily" (PF13532). This family includes the RNA m^{6}A demethylase ALKBH5, the remaining human ALKBH proteins and the *E.coli* AlkB. This protein (AlkB) repairs DNA through demethylation of damaged bases (Falnes and Rognes, 2003) and was the base to find the human ALKBH (1-8) (AlkB homologs) group (Kurowski et al., 2003) . After performing a search for proteins

containing this domain in *T. brucei*, five additional proteins were detected with a significant e-value (table 3.3.).

Table 3.3. *T. brucei* proteins with 2OG-Fe(II) oxygenase domain, putative *T. brucei* alkB homologs

TRITRYPDB ID	UNIPROT ID	PFAM E-VALUE	TRITRYPDB ANNOTATION	ATTRIBUTED NAME
Tb927.4.460	Q57US9	5.6e-54	Alkylated DNA repair protein	TbALKBH1
Tb927.5.980	Q57W43	4.0e-13	hypothetical , conserved	TbALKBH2
Tb927.7.1530	Q57WP7	1.4e-12	hypothetical , conserved	TbALKBH3
Tb927.10.6450	Q38AW1	1.9e-08	hypothetical , conserved	TbALKBH4
Tb927.11.10960	Q383I2	2.1e-08	hypothetical , conserved	TbALKBH5
Tb927.11.11390	Q383D9	6.8e-08	hypothetical , conserved	TbALKBH6

Considering the candidates on table 3.3., only Tb927.4.460 has a functional annotation as "Alkylated DNA repair protein (alkB homolog), putative". This protein was shown to partially complement AlkB knockout in *E. coli*, under DNA damage stress conditions, indicating that it could be a functional AlkB protein. However, evidence regarding the substrate and biochemical function in *T. brucei* has not been addressed (Simmons et al., 2012).

A MSA of these *T. brucei* proteins with AlkB and human ALKBH (1-8) revealed the conservation of critical residues on the candidate proteins (Figure 3.7.). These residues are the homologues to positions His131, Asp133, His187, Arg204 and Arg210 in *E. coli* AlkB. The presence of this domain with the conserved residues suggests these are *T. brucei* AlkB homologs (TbALKBH 1-6).

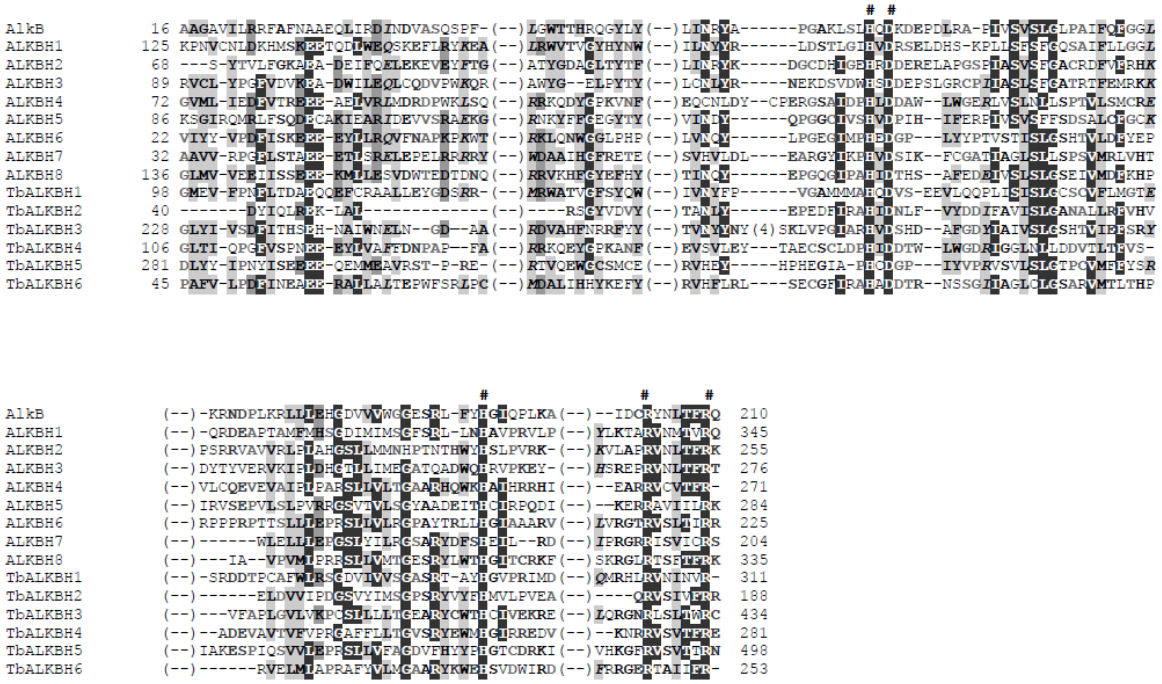


Figure 3.7. Multiple sequence alignment of sequences of the *E. coli* alkB, Human ALKBH and possible TbALKBH; conserved positions are marked with #; Numbers in parentheses are positions not represented. Numbers at the beginning and end represent the start and end positions of each sequence on the MSA. Columns with spaces inside parentheses denotes regions of the MSA not represented.

Similarly to ALKBH5 in *H. sapiens* where only one out of the eight ALKBH proteins is an RNA m^{6}A demethylase, it is possible that within that putative *T. brucei* AlkB homolog group are RNA m^{6}A demethylase enzymes. To test this possibility, the two best candidate genes were selected to generate the knockout (KO) cell lines: Tb927.5.980 and Tb927.4.460. Additionally in one candidate (TbALKBH1) was presented evidence that this is an AlkB protein (Simmons et al., 2012).

In summary, informatic analysis of *T. brucei* genome based on conserved sequence features lead me to identify one putative m^{6}A methyltransferase and five m^{6}A demethylases. In the next section, KO cell-lines were generated for the putative m^{6}A methyltransfrease gene (Tb927.7.6620) and for the two best candidate genes encoding for m^{6}A demethylases (Tb927.5.980 and Tb927.4.460).

3.5.3. Generation of knockout cell lines of putative enzymes

To test if the putative enzymes described above are indeed m^6A methyltransferase or demethylases, I decided to generate knockout parasite lines, so that I can genetically address the function of the genes. Generation of knockout cell lines consists in removing both alleles of the putative gene, usually by replacement with a drug resistance gene. Thus, my first task was to generate a construct that contained a drug selectable marker flanked by sequences that allow recombination in the genome and subsequent deletion of the targeted allele (Fig 3.8A). The untranslated regions (UTRs) of the drug resistance genes were either from the Aldolase or Actin genes. The endogenous UTRs of the targeted genes may have non-annotated regulatory elements that destabilize the transcripts and could, as a result, reduce the expression of the resistance genes. To avoid this possible effect, our constructs were designed to delete not only the targeted gene, but also their UTRs.

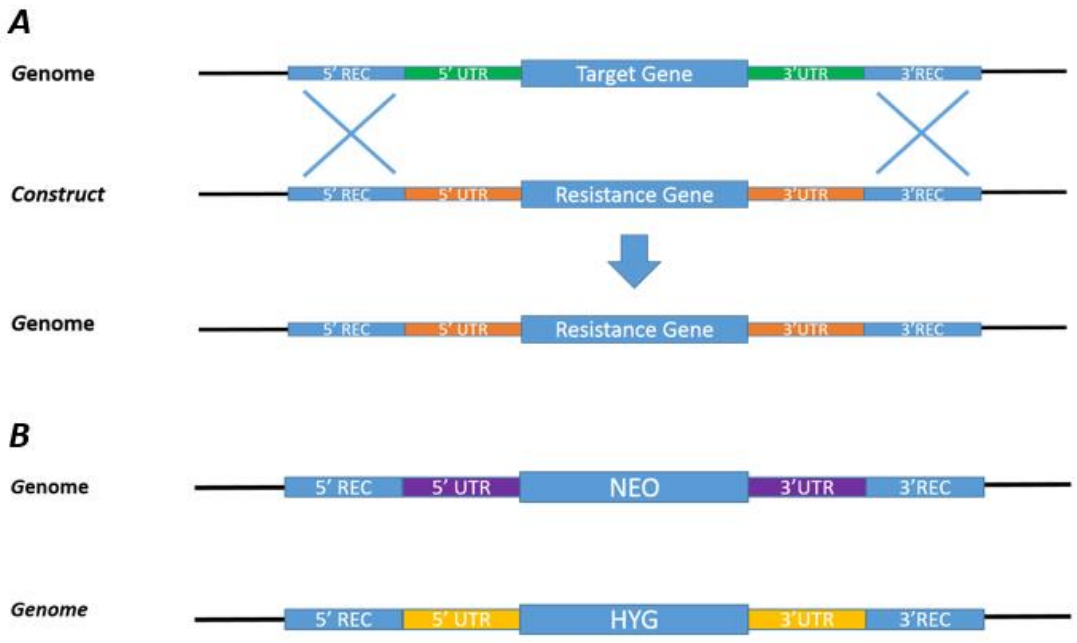


Figure 3.8. A- Strategy to replace the putative genes by the resistance gene via double recombination; B- Locus in the KO cell lines after the replacement of the two alleles; REC- Recombination region, UTR- untranslated region. Endogenous UTRs are in green, exogenous UTRs in orange; aldolase UTRs in purple and actin UTRs in yellow; NEO- G410; HYG- Hygromycin.

For each candidate gene, two vectors were made, each with a different selectable marker (NEO or HYG) (table 3.4.) to replace the two wild type alleles (figure 3.8B)

Table 3.4. Vectors designed to KO the candidate genes.

PUTATIVE FUNCTION	TARGET GENE	VECTOR	RESISTANCE GENE	CELL LINE	DESCRIPTION
RNA m^6A demethylase	Tb927.5.980	pIV1	NEO	IV1	KO of first allele
		pIV2	HYG	IV2	KO of second allele
RNA m^6A demethylase	Tb927.4.460	pIV3	NEO	IV3	KO of first allele
		pIV4	HYG	IV4	KO of second allele
RNA m^6A methyltransferase	Tb927.7.6620	pIV5	NEO	IV5	KO of first allele
		pIV6	HYG	IV6	KO of second allele

The fragments to generate the constructs were obtained by PCR amplification and are described in table 3.5. and Figure 3.9.

Table 3.5. Inserts to generate the pIVs vectors.

PCR PRODUCT	TARGET GENE	DESCRIPTION	VECTOR	EXPECTED SIZE (BP)
1	Tb927.5.980	5' recombination region to ligate with NEO	pIV1	230
2	Tb927.5.980	3' recombination region to ligate with NEO	pIV1	301
3	Tb927.5.980	5' recombination region to ligate with HYG	pIV2	230
4	Tb927.5.980	3' recombination region to ligate with HYG	pIV2	301
5	Tb927.4.460	5' recombination region to ligate with NEO	pIV3	361
6	Tb927.4.460	3' recombination region to ligate with NEO	pIV3	324
7	Tb927.4.460	5' recombination region to ligate with HYG	pIV4	361
8	Tb927.4.460	3' recombination region to ligate with HYG	pIV4	324
9	Tb927.7.6620	5' recombination region to ligate with NEO	pIV5	452
10	Tb927.7.6620	3' recombination region to ligate with NEO	pIV5	521
11	Tb927.7.6620	5' recombination region to ligate with HYG	pIV6	452
12	Tb927.7.6620	3' recombination region to ligate with HYG	pIV6	521
13	---	NEO gene	pIV(1,3,5)	1176
14	---	HYG gene	pIV(2,4,6)	1445

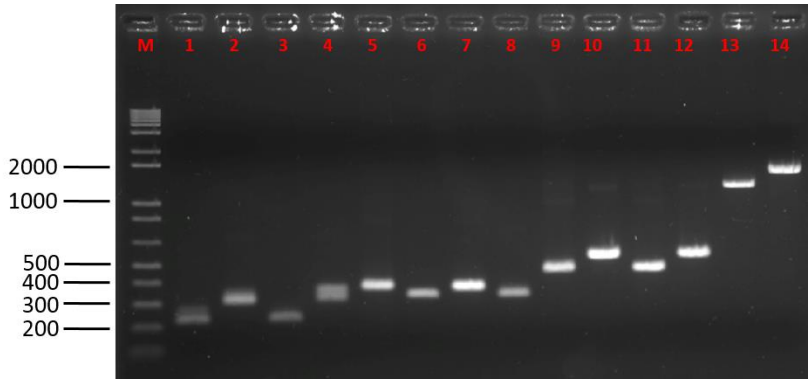


Figure 3.9. Agarose gel of the inserts amplifications products after gel extraction, lane M: marker 1 Kb ladder; lanes 1 to 14 correspond to PCR products in table 3.5. Ladder size is marked in bp.

All purified DNA fragments showed the expected molecular masses. The backbone for my constructs, which contains the ampicillin resistance gene and the *E. coli* origin of replication, was obtain by digestion of pFAB2, a plasmid previously available in the laboratory. Upon transformation of the reaction products in *E. coli*, colonies were tested for the presence of the desired plasmids by digesting plasmid DNA with restriction enzymes (table 3.6., Figure 3.10.).

Table 3.6. Restriction digestions to confirm the cloned plasmids.

DIGESTION PRODUCT	VECTOR	RESTRICTION ENZYMES	EXPECTED PRODUCTS (BP)
1	pIV1	EcoRI + KpnI	3022 + 1501
2	pIV2	BamHI + SacI	2870 + 1922
3	pIV3	BsaI + KpnI	3166 + 1511
4	pIV4	EcoRI + NotI	3534 + 1368
5	pIV5	BamHI + KpnI	3570 + 1395
6	pIV6	EcoRI + NotI	3669 + 1565

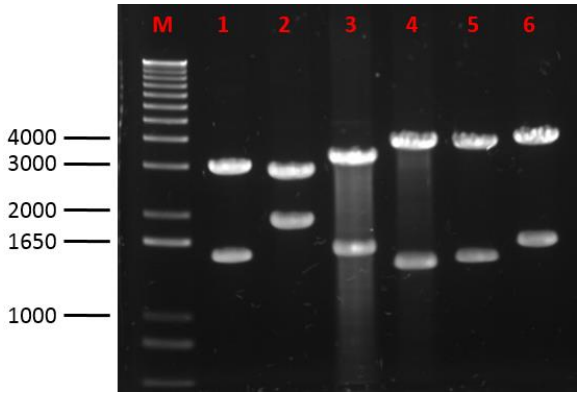


Figure 3.10. Agarose gel of the vectors digestion. Lane M: marker 1Kb ladder; lanes 1 to 6 correspond to digestion products of table 3.6. Ladder size is marked in bp.

The extracted plasmids exhibit the expected digestion pattern as shown in table 3.6. and Figure 3.10. Therefore, those should be the desired plasmids. After these preliminary digestion patterns, sequencing confirmed that all constructs have the expected sequences and are therefore ready for transfection in *T. brucei*.

The knockout mutants were generated in a pleomorphic strain of *T. brucei*: AnTat1.1E. A modified version of this strain, SMOx, contains the T7 polymerase and TET repressor. These elements are not critical to produce the KO cell line, but they may be useful in future studies to inducibly complementing the phenotype, for example. Therefore, I chose to transfect the KO constructs into SMOx.

SMOx was first transfected with constructs to replace the first allele (pIV1, pIV3 and pIV5 independently) and the clones were selected with G418. G418-resistant clones were genotyped to confirm the correct integration. Two amplifications were performed: one amplification in which the primers hybridize upstream of the 5' recombination region and inside the construct and another amplification in which the primers hybridize inside the construct and downstream of the 3' recombination regions (primers table annexes). I observed that on average 4/5 clones showed the correct integration. The clones that exhibit the correct integration correspond to the cell lines IV1, IV3 and IV5. These clones were subsequently used to transfect the second construct (pIV2, pIV4 and pIV6) to replace the second allele. Hygromycin-resistant clones were generated and called IV2, IV4 and IV6. Two clones were obtained for each KO cell line in which the genotyping results confirmed to be the desired cell line (tables 3.7., 3.8. and 3.9., Figures 3.11., 3.12. and 3.13.). Genotyping results with the two amplifications described above, for each resistance gene, reveal that in the three knockout cell lines resistance genes recombined in the right locus, replacing the gene of interest.

Table 3.7. Amplifications to genotype resistance genes integration in Tb927.5.980 locus.

SUBSTRATE	INTEGRATION	EXPECTED SIZE (BP)	PCR PRODUCT
AnTat	5' NEO	---	1
IV2	5' NEO	1406	2
AnTat	3' NEO	---	3
IV2	3' NEO	839	4
AnTat	5' HYG	---	5
IV2	5' HYG	1614	6
AnTat	3' HYG	---	7
IV2	3' HYG	759	8

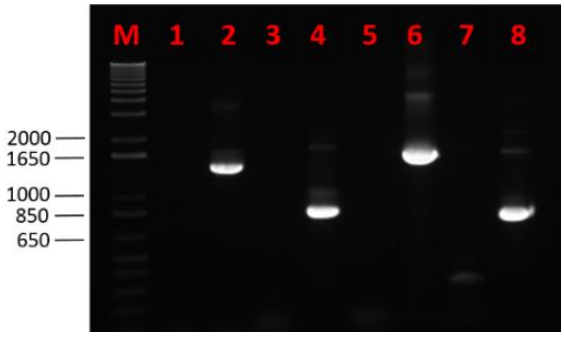


Figure 3.11. Agarose gel of resistance genes integration regions amplification. Lanes: M-marker 1 Kb ladder, lanes 1 to 8 correspond to PCR products in table 3.7. Ladder size is marker in bp.

Table 3.8. Amplifications to genotype resistance genes integration in Tb927.4.460 locus.

SUBSTRATE	INTEGRATION	EXPECTED SIZE (BP)	PCR PRODUCT
AnTat	5' NEO	---	1
IV4	5' NEO	1347	2
AnTat	3' NEO	---	3
IV4	3' NEO	841	4
AnTat	5' HYG	---	5
IV4	5' HYG	1555	6
AnTat	3' HYG	---	7
IV4	3' HYG	763	8

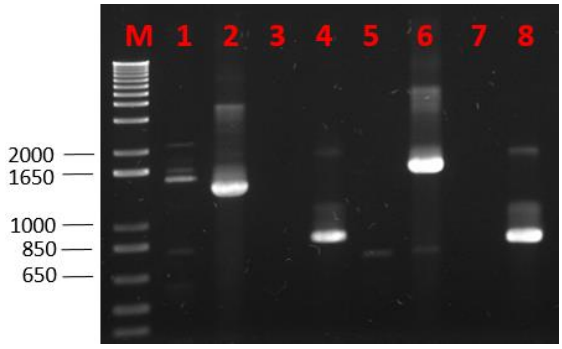


Figure 3.12. Agarose gel of resistance genes integration regions amplification. Lanes: M-marker 1 Kb ladder, lanes 1 to 8 correspond to PCR products in table 3.8. Ladder size is marker in bp.

Table 3.9. Amplifications to genotype resistance genes integration Tb927.7.6620 locus.

SUBSTRATE	INTEGRATION	EXPECTED SIZE (BP)	PCR PRODUCT
AnTat	5' NEO	---	1
IV6	5' NEO	1807	2
AnTat	3' NEO	---	3
IV6	3' NEO	1095	4
AnTat	5' HYG	---	5
IV6	5' HYG	2015	6
AnTat	5' HYG	---	7
IV6	5' HYG	1015	8

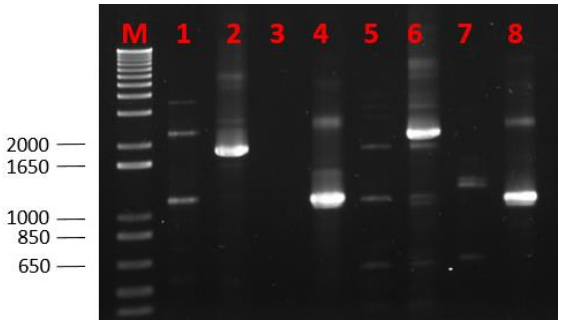


Figure 3.13 Agarose gel of resistance genes integration regions amplification. Lanes: M-marker 1 Kb ladder, lanes 1 to 8 correspond to PCR products in table 3.9. Ladder size is marker in bp.

Besides amplification of recombination regions, the locus was amplified (the three loci), the different size of each allele (wild type, G418 and Hygromycin) allows to distinguish each other (with the exception of Tb927.5.980 and G418, in which the difference in size is not distinguishable in agarose gel) (Table 3.10. and Figure 3.14.). In all three loci, amplification in AnTat 1.1 SMOx generates the expected molecular size product. After removal of the first allele, of the three candidates, through replacement with G418 gene (intermediate cell lines only with one allele removed, IV1, IV3

and IV5), amplification of the locus reveals an additional band that correspond to G418. In the double knockout cells, the amplification reveals the presence of the two alleles (G418 and hygromycin) instead of wild type allele.

Table 3.10. Amplifications of the three candidate genes locus

LOCUS	ALLELE	EXPECTED SIZE (BP)	PCR PRODUCT
Tb927.5.980	WT	2001	1
	NEO	2085	2
	HYG	2354	3
Tb927.4.460	WT	2853	4
	NEO	2030	5
	HYG	2299	6
Tb927.7.6620	WT	4052	7
	NEO	2742	8
	HYG	3011	9

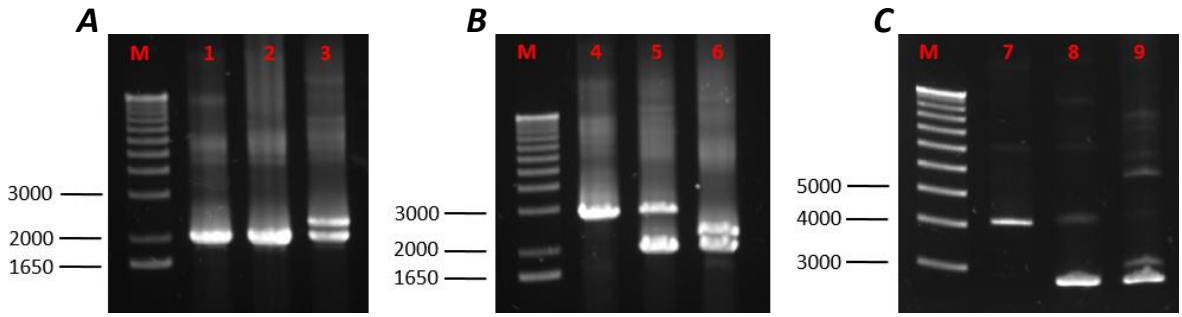


Figure 3.14. Agarose gel showing genotyping of single and double-knockout cell lines of the three putative gene loci A) Tb927.5.980 B) Tb927.4.460 and C) Tb927.7.6620 locus. Lanes: M- marker 1 Kb ladder, lanes 1 to 9 correspond to PCR products in table 3.10. Ladder size is marker in bp

Amplification of integration regions reveals that the resistance genes integrated in the expected locus, in the three selected candidates. Accordingly, amplification of the locus revealed products with sizes corresponding to the two resistance genes. These data demonstrate that the clones obtained are knockouts of the target genes.

Obtaining knockout cells indicates that the genes (three candidates) are not essential. In knockout of Tb927.5.980 gene, putative demethylase TbALKBH2, a reduction on growth rate was observed (figure 3.15A): relative to parental cell-line, which has a doubling time of 7:05hr, two clones

showed doubling times of 9:19hr and 8:51hr. In the other knockout cell line of Tb927.4.460 gene (putative demethylase TbALKBH1), a reduction in growth rate was also observed (figure 3.15B), the doubling time of which is longer in the two clones (10:13hr for clone1 and 9:17hr for clone 2). The knockout cell line of the putative methyltransferase gene, Tb927.7.6620, showed only a minor effect in the growth rate in the two clones (7:36 h for clone 1 and 7:09 h for clone 2) (figure 3.15C).

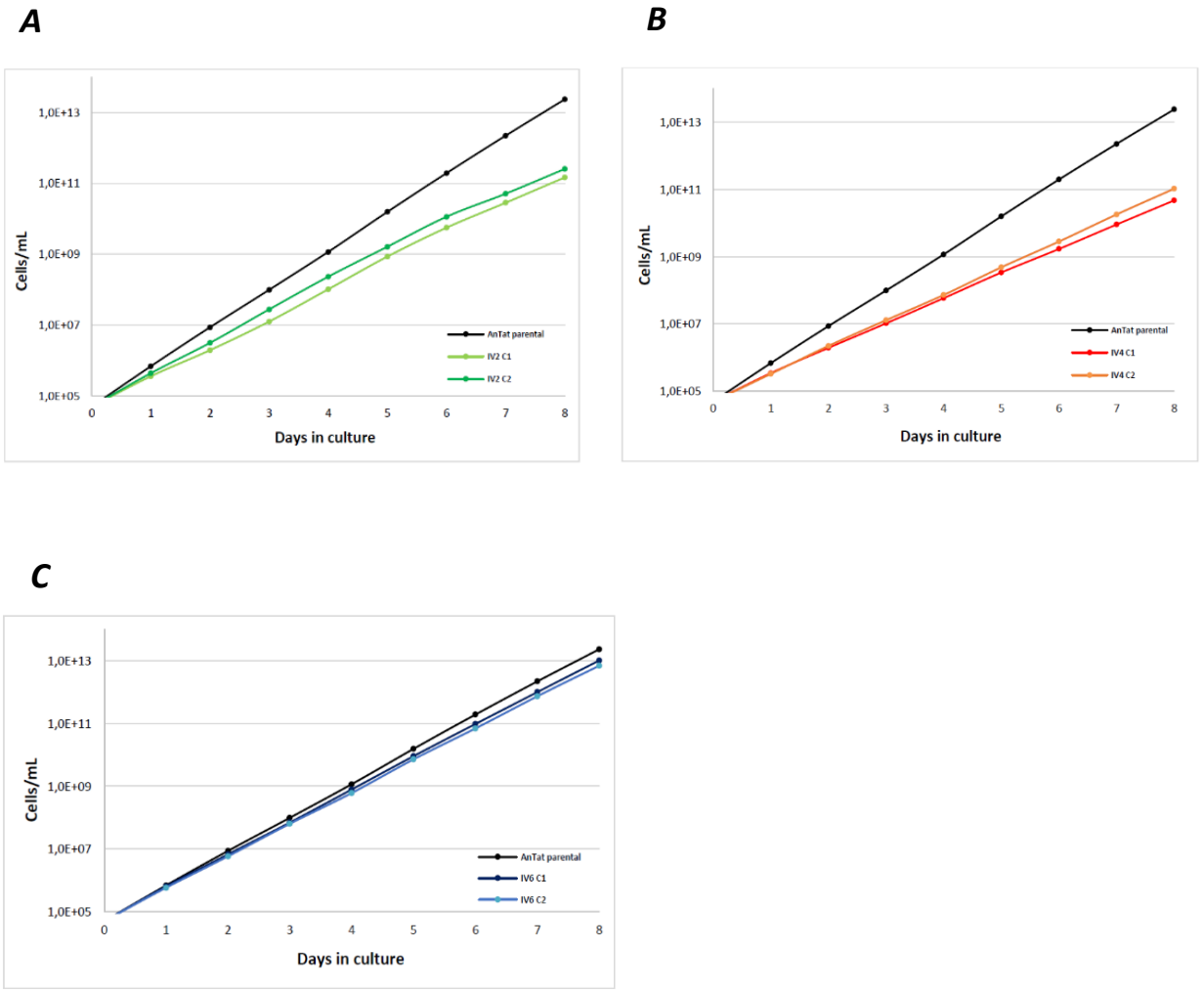


Figure 3.15. Growth curve of knockout cell lines during 8 days of culture. A) IV2 cell line (Tb927.5.980 putative RNA m^{6}A demethylase) B) IV4 cell line (Tb927.4.460 putative RNA m^{6}A demethylase) C) IV6 cell line (Tb927.7.6620 putative RNA m^{6}A methyltransferase). Result from one single experiment with the two obtained clones of each candidate.

3.5.4. Measurement of m^6A levels in knockout cell lines

To test if the three selected candidate proteins are involved in m^6A modification, I measured the levels of m^6A on the KO cell-lines. I expected the lack of the methyltransferase to result in a reduction of the m^6A levels, while the lack of the demethylase should result in an increase of m^6A levels. The levels of m^6A were measured in total RNA from the KO cell lines (figure 3.16.). The levels in one demethylase candidate (Tb927.5.980, IV2 cell line) cell line seems slightly higher than parental cell line (AnTat 1.1 SMOx), however the number of biological replicates does not allow to establish statistical support of this increase. A small increase in m^6A levels (9%) have also been observed in ALKBH5 mutants (Zheng et al., 2013a) . The KO of the other demethylase candidate (Tb927.4.460, IV4 cell line) does not change the levels of m^6A relative to the parental cell line. This result might suggest that one candidate (Tb927.5.980) could be a RNA m^6A demethylase, whilst the second candidate (Tb927.4.460) does not.

In contrast to what I expected, m^6A levels did not decrease in the KO cell-line of the putative methyltransferase gene (Tb927.7.6620, IV6 cell line), therefore this candidate does not seem to be an RNA m^6A methyltransferase.

The same analysis was done in procyclic forms, a different stage of *T. brucei* life cycle. For this, I differentiated one clone of each of three knockout cell-lines using the protocol described earlier. In procyclic, I did not detect any alteration in the m^6A levels between any of the cell-lines.

I obtained no evidence to prove that the three putative genes are indeed m^6A methyltransferases or demethylases. As discussed later in this thesis, there are several scenarios that could justify that the levels m^6A remain unchanged, even if the enzymes are truly involved in m^6A metabolism.

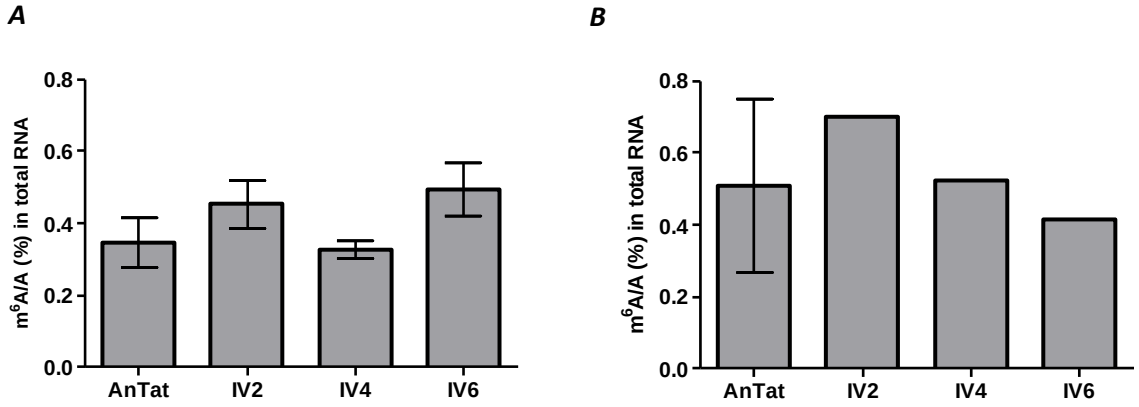


Figure 3.16. Levels of m^{6}A in total RNA in the generated KO cell lines, A) in bloodstream form and B) procyclic form. IV2 and IV4 are the knockout of the two RNA m^{6}A demethylase candidates (Tb927.5.980 and Tb927.4.460) and IV6 a knockout of RNA m^{6}A methyltransferase candidate (Tb927.7.6620). In the three case clone 1 was used. AnTat measurements result from 3 independent experiments. Knockout bloodstream forms measurements result from two independent experiments and knockout procyclic measurements from one single experiment.

Overall, my study suggests that one of the three putative genes (Tb927.5.980) might be an m^{6}A demethylase, while the other two might not have a role in the m^{6}A modification process. However, due to the small sample size and the possible small changes associated with the knockout of published orthologs, other assays would be necessary to further validate my observations.

4. Discussion

4.1. N6-methyladenosine (m^6A) in RNA of T. brucei

In several eukaryotes m^6A in RNA has been associated with post-transcriptional gene regulation, through diverse molecular mechanisms and possibly affecting mRNA metabolism in diverse steps (Fu et al., 2014; Meyer and Jaffrey, 2014) . *T. brucei* gene expression regulation occurs mainly at post-transcriptional level (Clayton, 2002). Therefore in this thesis I proposed that m^6A in RNA could be a novel mechanism of gene regulation in this parasite.

In this study, I identified m^6A modification for the first time in *T. brucei* RNA. I detected this modification in the two different life-cycle stages: bloodstream slender form and procyclic form. Levels of m^6A in RNA relative to the total number of adenines is around 0,3% in bloodstream form and around 0,5% in procyclic form. These levels are in the same order of magnitude as in other eukaryotes: 0,1%-0,4% in mRNA in human cells (Perry and Kelley, 1975), 0,25% in *S. cerevisiae* (Bodi et al., 2010) and 1,5% in *A. thaliana* (Zhong et al., 2008). Although the levels between these two life-cycle stages are very similar, it does not mean that the targets are the same, neither the stoichiometry within the targets. It is plausible that in different life stages, the methylation targets are different, and/or in the same transcript population, the stoichiometry varies for developmental regulation purposes.

T. brucei is an eukaryote that exhibits features that are distinct from canonical eukaryotes, including polycistronic gene organization (Berriman et al., 2005), RNA maturation by trans-splicing (Liang et al., 2003) and lack of mRNA transcriptional control (Clayton, 2002). These distinctive features could be explained due to the fact that *T. brucei* speciation occurred early in the eukaryotic lineage (Douzery et al., 2004). Nevertheless, it is possible that some m^6A molecular mechanisms and functions observed in other eukaryotes are conserved in *T. brucei*. For example, RNA Recognition Motifs (RRM) are present in several *T. brucei* proteins (Clayton, 2013) and they have been associated to the stabilization of mRNAs (Wang et al., 2014b). Expression of human HuR protein (an RRM-containing protein) in *T.brucei* leads to stabilization of some mRNAs, through binding to ARE (AU-rich element) containing transcripts (Quijada et al., 2002). These observations suggest that possibly, this mechanism of mRNA stabilization through HuR binding could be conserved in *T. brucei.*

In the eukaryotic lineage, m^6A modification is found in rRNA, snRNA and in mRNA (Cantara et al., 2011). In *T. brucei*, I detected m^6A levels in total RNA, which means that all methylated transcripts were quantified. With our assay, it was not possible to discriminate in which types of RNA this modification is present. An immunoblot after RNA separation is a complementary approach,

which would tell me the size of the molecules that are methylated. Definitive identification of the targets would require more sophisticated methods involving immunoprecipitation and RNA sequencing (meRIP-seq). The possible presence of m^6A in the mRNA of *T. brucei* is more promising in the context of the hypothesis that m^6A is involved in *T. brucei* post-transcriptional gene regulation. Nevertheless, if this modification is only present on other types of RNAs it may also be interesting, since the function of this modification in other RNAs is unknown.

To understand the function of this modification, it is important not only to discriminate which types of RNAs contain m^6A, but also to localize the methylated nucleotides in the transcripts. This localization could indicate possible functions and reveal putative methylation motifs, which could be empirically tested after, for example through mutational analysis. Also, with the methylated transcriptome revealed, it will be possible, for example, to select a set of methylated transcripts and a set of non-methylated transcripts to compare different properties, like the kinetics of decay using transcription inhibitors, knockout cell lines (of the enzymes involved in the methylation process) or methylation inhibitors. This strategy may indicate if this modification is important for the stability of mRNA transcripts, as it has been previously shown for transcripts that regulate self-renewal ability in mouse stem cells (Wang et al., 2014b) .

Gene ontology analysis of m^6A targets may indicate the biochemical pathways and biological process potentially regulated by RNA m^6A methylation. Comparing methylated transcriptome in different life-cycle stages could help us understand if the modification is dynamic, regulated and which biological processes are more regulated by m^6A in RNA.

Spliced Leader (SL) cap modifications , namely 7-methylguanosine, 2-O-ribose methylations, N6,N6,2'-O-trimethyladenosine, 3,2'-O-dimethyluridine and pseudouridine (Bangs et al., 1992; Zamudio et al., 2009) and the presence of 5-methylcytosine in tRNAs (Militello et al., 2014), were, until now, the only RNA modifications known in *T. brucei*. This finding, the presence of m^6A in *T. brucei* RNA, adds a new modification to the epitranscriptome.

4.2. N6-methyladenosine (m^6A) in DNA of T. brucei

One of the results of this thesis is the identification for the first time of N6-methyladenosine in DNA of *T. brucei*. This novel DNA modification was detected in both life-cycle stages. In every 10.000 adenines, 12 are N6-methylated in bloodstream. In procyclic the frequency is about the double. DNA modifications are epigenetic marks. 5-methylcytosine is one of the best characterized in several

eukaryotes and it is well established its role in transcriptional gene regulation (Jones, 2012). In *T. brucei*, two DNA modifications have been previously described: 5-methylcytosine (Militello et al., 2008) and base J (Gommers-ampt et al., 1993). The function of 5-methylcytosine and base J in *T. brucei* have not been well established, however it seems that base J is involved in regulation of transcription termination (Reynolds et al., 2014). In DNA, m^6A is found in bacteria, archaea, protists and fungi, where besides being part of modification-restriction systems, it is an epigenetic mark involved in diverse processes, including chromosome replication, DNA repair, transposition and transcription of specific genes (Wion and Casadesús, 2006). Some of these functions could be conserved in *T. brucei*. The levels of 5-methylcytosine in *T. brucei* are lower than in canonical eukaryotes (around 0,01% in *T. brucei* and around 0, 25% in human tissues, varying from tissue to tissue (Ehrlich et al., 1982) , therefore it is possible that m^6A in DNA compensates some functions usually associated to m^5C. One major function of m^5C is the involvement in transcription repression, *in T. brucei* poll II transcription is mainly constitutive and therefore it is not surprising to have reduced levels of m^5C. Nevertheless, a large part of the genome encodes VSGs in silent state, m^5C and m^6A could be important to repress this loci.

Because promoters for Pol II in *T. brucei* lack recognisable sequence elements, a putative function of m^6A in DNA could be to mark specific genomic loci like, such as transcription start sites and transcription termination sites. Chromatin alterations observed in these genomic regions (Siegel et al., 2009) could be promoted by the presence of m^6A in DNA in these genomic regions. This hypothesis could be addressed by mapping the modification in the genome, which could be achieved by chromatin immunoprecipitation followed by high throughput sequencing.

4.3. Levels of RNA m^6A during differentiation

During *T. brucei* differentiation, several changes occur in gene expression (Queiroz et al., 2009). If m^6A is involved in gene regulation during differentiation it is possible that the levels of this modification are regulated during this process. For example, in *S. cerevisiae* this modification only appears in RNA during meiosis (Clancy et al., 2002).

In vivo bloodstream slender cells differentiate into non-dividing stumpy cells, leading to a cell-cycle arrest in G1 phase, in the mammalian host (Fenn and Matthews, 2007). These cells, upon transfer to the tsetse in its blood meal, continue the development to the procyclic form while the bloodstream slender forms that are uptaken by the tsetse eventually die (Dyer et al., 2013). *In vitro* differentiation starts with induction with cis-aconitate and temperature reduction to 27°C of a bloodstream slender form culture. In the first 12 hours of *in vitro* differentiation, cells are arrested in

G1 phase, do not grow and express genes characteristic to stumpy forms (Queiroz et al., 2009). Transcriptomic analysis has shown that during *in vitro* differentiation, after 12 hours several genes important for division, including flagella and ribosomal proteins, glycolysis and other energy metabolism pathways reach transcript levels comparable to the observed levels in procyclics. Between 12 hours and 48 hours, the stumpy-like form restart growing and fully differentiate into procyclic forms (Queiroz et al., 2009).

In this work, I assayed differentiation *in vitro*. The levels of m^6A in RNA remained constant (around 0,3%) in the first 12 hours, but at 24 hours the levels of m^6A had doubled to around 0,7%. Levels remained high till 72hr. Therefore, it seems that m^6A levels increase when differentiation from stumpy-like form to procyclic is triggered. This correlation, between the m^6A levels increase and this transition suggests that m^6A may have a role in differentiation. Extrapolating to differentiation *in vivo*, it is possible that m^6A plays an important role in the transition from stumpy to procyclic, unlike the transition from bloodstream slender to stumpy. To address this possibility, the measurement of the levels of m^6A in stumpy forms will be important. Another possible assay is to measure m^6A levels in a differentiation assay that starts with stumpy cells (rather than bloodstream slender forms). Stumpies can be obtained from an infection in mouse. If the hypothesis is correct, the increase in m^6A levels in a differentiation started with stumpy cells would be observed without an initial lag phase of 12hr.

During differentiation several transcripts are regulated with the same pattern, forming post-transcriptional regulons (Queiroz et al., 2009). It is possible that the presence of m^6A defines some of these co-regulated clusters. With the information of the methylation targets obtained by meRIP-seq, it should be possible to compare the methylation status in the transcripts of different regulons. This co-regulation could be tested using methylation inhibitors.

4.4. m^6A RNA modification is sensitive to cell density

Cell density is an important factor in *T. brucei* life-cycle because high cell density triggers the differentiation of bloodstream slender forms to stumpy forms (Reuner et al., 1997). I observed a gradual increase in m^6A levels from 0,4% to 1,5% when cells were cultivated at higher densities (0,5 $\times 10^6$ cells/mL to 3,5$\times$ 10^6 cells/mL). Is m^6A higher because parasites are responding to cell density stress or because they initiated differentiation to stumpies? A quorum sensing mechanism is responsible for differentiation from bloodstream slender forms to stumpies *in vivo*, and this process can be reproduced *in vitro* with agar plates (Reuner et al., 1997). However, in liquid medium, keeping cultures at high densities does not promote proper differentiation to stumpies but to a poorly

characterized stumpy-like status (Queiroz et al., 2009). Our cell-density assay was performed *in vitro*, which means that I did not have stumpies in culture, but I probably had stumpy-like forms. In our differentiation assay (also *in vitro*) I observed constant levels of m^6A in the first 12hr, indicating that stumpy-like forms have the same levels of m^6A as bloodstream slender forms. Therefore, the increase in m^6A observed upon increasing cell-density is likely due to the stress of being at high concentration, rather than a differentiation process to stumpy-like cells. Stress could result from the accumulation of metabolites and the reduction of available nutrients in the culture medium. Response to stress conditions usually involves alterations in gene expression (de Nadal et al., 2011). Thus, m^6A regulation may be involved in *T. brucei* response to stress, in this case high cell density.

Increase of m^6A levels may be involved in the regulation of specific stress response genes. For example, response to heat shock in *T. brucei* involves the regulation of heat shock genes, through the binding of an RNA binding protein that stabilizes the transcripts (Droll et al., 2013). If m^6A destabilizes transcripts, the increase in m^6A levels could result in down-regulation of genes usually expressed in normal conditions. An alternative effect is an increased transcript stability, promoting an up-regulation, for example, of genes involved in stress response. Comparing the up-regulated and down-regulated genes in high stress conditions with the methylation status could reveal a relationship between m^6A and regulation of stress response genes.

4.5. Putative RNA m^6A methyltransferase

So far, the only proteins identified as responsible for the catalysis of m^6A formation in RNA have a domain from MT-A70 family (Bujnicki et al., 2002; Liu et al., 2014). In the *T. brucei* genome, no annotated gene codifies for a protein with such a domain. The enzyme (or enzymes) that catalyse(s) the formation of m^6A in *T. brucei* RNA may be from a different family. On the other hand, a possible stronger divergent evolution (more substitutions in the sequence) in trypanosomes, could difficult the detection of MT-A70 homologues. The candidate I selected belongs to a protein family N6-adenineMlase (PF10237) annotated as "Probable N6-adenine methyltransferase", a distinct family from MT-A70 family. The hypothesis that this protein could be an RNA m^6A is due to the presence of a conserved characteristic motif of m^6A methyltransferases (called motif IV) (Bujnicki et al., 2002; Malone et al., 1995). Although the function of this domain family is not known, according to Pfam it is only found in the eukaryote lineage.

In higher eukaryotes the presence of m^6A in DNA has never been detected (Wion and Casadesús, 2006), although this domain (N6-adenineMlase, PF10237) is distributed in several higher eukaryotes including *Homo sapiens, Mus musculus, Drosophila melanogaster* and *Arabidopsis*

thaliana. In these organisms, m^6A has been identified in RNA, therefore if this family is an m^6A methyltransferase, RNA is a more plausible substrate than DNA. In this study, the putative function of m^6A methyltransferase for RNA of the gene Tb927.7.6620 (that has a N6-adenineMlase, PF10237 domain) was tested, through the measurement of m^6A levels in a KO cell line for this candidate. If this candidate was indeed an RNA m^6A methyltransferase, a decrease in the m^6A RNA methylation levels was expected in KO cell-lines. The levels of m^6A in the KO cells were measured in both bloodstream and procyclic forms. In bloodstream forms, instead of a decrease in RNA m^6A levels, a small increase was observed relative to the parental cell line. However, this may not be an actual increase due the small sample size and variability in measurements. In procyclic forms, the levels in the KO and parental cell lines are equivalent. However, the measurements are very variable. Therefore, if the decrease in procyclic forms was just slight, it might not be observed due to variability in the measurements. For example in mouse cells, after knockdown of m^6A methyltransferase (MT-A70 and Mettl14), a decrease was observed from around 0,15% (m^6A/A) to around 0,05% (m^6A/A), corresponding to a reduction to 1/3 of the normal levels (Wang et al., 2014b). Moreover, in another study, after MT-A70 knockdown in human cells the decrease in m^6A levels was around 30%, that correspond to a decrease from 0,42% (m^6A/A) to 0,30 % (m^6A/A) (Liu et al., 2014). Because our analysis of the KO cell-line was inconclusive, in the future, mass spectroscopy and *in vitro* biochemical assays may allow us to conclude if the putative RNA m^6A methyltransferase is real. Another approach is to test if this KO can differentiate properly and if levels of m^6A are identical to wild-type or not. I should also measure m^6A levels in this KO cell-line upon stress of high-density and test if levels also increase as observed in wild-type.

4.6. Putative RNA m^6A demethylases

In other eukaryotes, the RNA demethylases identified so far are FTO (Jia et al., 2011) and ALKBH5 (Zheng et al., 2013a). In *T. brucei* genome there are six genes that encode for proteins with a domain from the same family as ALKBH5, the 2OG-Fe(II) oxygenase superfamily. Proteins that contains this domain catalyse diverse reactions, usually the oxidation of an organic compost and this domain is present, for example in EGL-9 and Prolyl 3-hydroxylase-1 (Aravind and Koonin, 2001). EGL-9 response to low oxygen conditions (hipoxia) through hydroxylation of hypoxia-inducible factor (HIF) (Shao et al., 2009). Prolyl 3-hydroxylase-1 (P3H1) converts proline to 3-hydroxyproline in type I collagen (Hudson and Eyre, 2014). Multiple sequence alignment of *T. brucei* proteins with human ALKBH (1-8) and *E. coli* AlkB revealed that the most conserved amino acids are also present in *T. brucei* proteins. Due to the sequence similarity and conservation of the critical amino acids, it is possible that these proteins could be functional AlkB proteins, and possibly one (or more than one)

could be an m^6A RNA demethylase. Thus, I termed these six putative proteins as *T. brucei* AlkB homologues (TbALKBH (1-6)).

In this study, two out of the six identified *T. brucei* ALKBH proteins (Tb927.4.460 called TbALKBH1 and Tb927.5.980 called TbALKBH2) were tested for a putative function of RNA m^6A demethylase, by measuring RNA m^6A levels in KO cell lines. If one of the two proteins is an RNA m^6A demethylase, I expected an increase in the methylation levels. The levels were measured in both life-cycle stages, bloodstream and procyclic form. RNA m^6A demethylase candidate knockout, Tb927.4.460 (TbALKBH1), showed identical levels of m^6A in RNA between parental and KO cell-lines in both life-cycle stages. Therefore, it seems that this TbALKBH1 is not an RNA m^6A demethylase. In the bloodstream form of Tb927.5.980 (TbALKBH2) knockout, a slight increase in the m^6A levels in RNA was observed from 0, 34% to 0, 45%, relative to total adenosines. Therefore this protein might be an m^6A RNA demethylase. In procyclic form, the m^6A levels were identical between parental and KO cell-lines, suggesting that there is either a second demethylase or that the change in m^6A RNA levels is more subtle in procyclic forms.

Two relevant factors need to be considered when analysing the putative RNA m^6A demethylase activities in the knockout parasites. First, the fold increase in the levels of m^6A in RNA expected. For example, Zheng and colleagues only found a "modest" increase (from around 0,50% to 0,55% (m^6A/A)) in m^6A levels in cultured cells after knockdown of the human RNA m^6A demethylase that belongs to the same superfamily of *T. brucei* candidates (Zheng et al., 2013a). Second, the high variability observed in measurements with this method. Together, the high variability combined with possible slight increase in the m^6A levels can make difficult to evaluate the demethylase activity of the candidates. The small sample size does not allow a statistical analysis neither the identification of outliers, that may be increasing the variability observed. Measurement of m^6A levels in KO cell lines with mass spectroscopy and with a bigger sample size may give more robust evidence to determine if the candidates are true RNA m^6A demethylases. Biochemical assays *in vitro* (for example, incubation of purified candidates with m^6A methylated RNA substrate and analyse the products) and overexpression (observe if the levels in RNA decrease) can generate supplementary evidence. Together, the data so far suggests that one candidate could be an RNA m^6A demethylase, which is mainly active in bloodstream form (Tb927.5.980).

.

5. Conclusion

N6-methyladenosine is an RNA modification distributed in the transcriptome of eukaryotes and was associated to post-transcriptional gene regulation. This regulation seems to occur through diverse molecular mechanisms, affecting diverse steps in the RNA processing, including transport and stability (Fu et al., 2014; Meyer and Jaffrey, 2014). In *T. Brucei*, gene expression regulation occurs mainly at post-transcriptional level (Clayton, 2002). My hypothesis is that this modification exists and is a mechanism of gene expression regulation in *T. brucei*.

I detected the presence of m^6A in *T. brucei* RNA and DNA, in both bloodstream and procyclic life stages. The levels are regulated in two different biological conditions. First, I observed that during differentiation from bloodstream to procyclic form the levels increased twelve hours after inducing differentiation. This time window correlates with the transition between the arrested stumpy-like cells, in the first hours of differentiation, to growing procyclic cells. This correlation suggests that this modification could have a role in this transition from stumpy-like to procyclic forms. Second, I also observed that an increase in cell density also lead to an increase in m^6A levels in RNA, suggesting a role of this RNA modification in response to high density stress.

In the genome of *T. brucei* is encoded a protein that, due to the presence of a characteristic conserved motif, is a plausible m^6A methyltransferase. However the generation of a knockout cell line followed by measurement of m^6A levels in RNA did not allow to demonstrate this putative function. Six proteins are encoded in *T. brucei* genome, forming a putative group of AlkB homologues group, named TbALKBH (1-6). From this group, two were selected (TbALKBH1 and TbALKBH2) to test the putative function of RNA m^6A demethylase through knockout cell lines generation. Only in TbALKBH2 a small increase in m^6A levels was observed, indicating that this candidate could be an RNA m^6A demethylase. Further investigation is required to confirm this observation in these putative enzymes.

In this thesis I identified a novel modification in RNA. My initial hypothesis that m^6A is important for post-transcriptional gene regulation in *T. brucei* was not confirmed, but it was also not disproved. The detection of m^6A in RNA of two life-cycle stages and the fact that it is dynamic, are however strong indications that this hypothesis is plausible. The evidence presented in this thesis adds a new modification in *T. brucei* epitranscriptome, with unknown function in this parasite and raises diverse possibilities, including the possibility of gene regulation through RNA m^6A modification.

6. References

Agarwala, S.D., Blitzblau, H.G., Hochwagen, A., and Fink, G.R. (2012). RNA methylation by the MIS complex regulates a cell fate decision in yeast. PLoS Genet. *8*, e1002732.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. *215*, 403–410.

Altschul, S.F., Madden, T.L., Schäffer, a a, Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. *25*, 3389–3402.

Apweiler, R., Bairoch, A., Wu, C.H., Barker, W.C., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H., Lopez, R., Magrane, M., et al. (2004). UniProt: the Universal Protein knowledgebase. Nucleic Acids Res. *32*, D115–D119.

Aravind, L., and Koonin, E. V (2001). The DNA-repair protein AlkB, EGL-9, and leprecan define new families of 2-oxoglutarate- and iron-dependent dioxygenases. Genome Biol. *2*, RESEARCH0007.

Aslett, M., Aurrecoechea, C., Berriman, M., Brestelli, J., Brunk, B.P., Carrington, M., Depledge, D.P., Fischer, S., Gajria, B., Gao, X., et al. (2010). TriTrypDB: a functional genomic resource for the Trypanosomatidae. Nucleic Acids Res. *38*, D457–D4662.

Balagopal, V., and Parker, R. (2009). Polysomes, P bodies and stress granules: states and fates of eukaryotic mRNAs. Curr. Opin. Cell Biol. *21*, 403–408.

Bangs, J.D., Crainll, P.F., Hashizumell, T., Mccloskeyllii, J.A., and Boothroyd, J.C. (1992). Mass spectrometry of mRNA cap 4 from trypanosomatids reveals two novel nucleosides. *267*, 9805–9815.

Bannister, A.J., and Kouzarides, T. (2011). Regulation of chromatin by histone modifications. Cell Res. *21*, 381–395.

Berriman, M., Ghedin, E., Hertz-Fowler, C., Blandin, G., Renauld, H., Bartholomeu, D.C., Lennard, N.J., Caler, E., Hamlin, N.E., Haas, B., et al. (2005). The genome of the African trypanosome Trypanosoma brucei. Science *309*, 416–422.

Bhattacharyya, R.P., Reményi, A., Yeh, B.J., and Lim, W. a (2006). Domains, motifs, and scaffolds: the role of modular interactions in the evolution and wiring of cell signaling circuits. Annu. Rev. Biochem. *75*, 655–680.

Bird, A. (2002). DNA methylation patterns and epigenetic memory. Genes Dev. *16*, 6–21.

Bodi, Z., Button, J.D., Grierson, D., and Fray, R.G. (2010). Yeast targets for mRNA methylation. Nucleic Acids Res. *38*, 5327–5335.

Bokar, J., Shambaugh, E., Polayes, D., Gregory, M., and Fritz, R. (1997). Purification and cDNA cloning of the AdoMet-binding subunit of the human mRNA (N6-adenosine)-methyltransferase. RNA *3*, 1233–1247.

Bokar, J.A., Rath-shambaugh, M.E., Ludwiczak, R., Narayann, P., and Rottmanll, F. (1994). Characterization and Partial Purification of mRNA NG-Adenosine Methyltransferase from HeLa Cell Nuclei. J. Biol. Chem. *269*, 17697–17704.

Borst, P., and Sabatini, R. (2008). Base J: discovery, biosynthesis, and possible functions. Annu. Rev. Microbiol. *62*, 235–251.

Bravard, A., Vial, G., Chauvin, M.-A., Rouillé, Y., Bailleul, B., Vidal, H., and Rieusset, J. (2014). FTO contributes to hepatic metabolism regulation through regulation of leptin action and STAT3 signalling in liver. Cell Commun. Signal. *12*, 4.

Brun, R., Blum, J., Chappuis, F., and Burri, C. (2010). Human African trypanosomiasis. Lancet *375*, 148–159.

Bujnicki, J.M., Feder, M., Radlinska, M., and Blumenthal, R.M. (2002). Structure prediction and phylogenetic analysis of a functionally diverse family of proteins homologous to the MT-A70 subunit of the human mRNA:m(6)A methyltransferase. J. Mol. Evol. *55*, 431–444.

Burkard, G., Fragoso, C.M., and Roditi, I. (2007). Highly efficient stable transformation of bloodstream forms of Trypanosoma brucei. Mol. Biochem. Parasitol. *153*, 220–223.

Butter, F., Bucerius, F., Michel, M., Cicova, Z., Mann, M., and Janzen, C.J. (2013). Comparative proteomics of two life cycle stages of stable isotope-labeled Trypanosoma brucei reveals novel components of the parasite's host adaptation machinery. Mol. Cell. Proteomics *12*, 172–179.

Cantara, W. a, Crain, P.F., Rozenski, J., McCloskey, J. a, Harris, K. a, Zhang, X., Vendeix, F. a P., Fabris, D., and Agris, P.F. (2011). The RNA Modification Database, RNAMDB: 2011 update. Nucleic Acids Res. *39*, D195–201.

Carroll SM, Narayan P, Rottman FM (1990). N6-Methyladenosine Residues in an Intron-Specific Region of Prolactin Pre-mRNA. Mol. Cell. Biol. *10*, 4456–4465.

Cassola, A. (2011). RNA Granules Living a Post-transcriptional Life: the Trypanosomes' Case. *5*, 108–117.

Church, C., Moir, L., McMurray, F., Girard, C., Banks, G.T., Teboul, L., Wells, S., Brüning, J.C., Nolan, P.M., Ashcroft, F.M., et al. (2010). Overexpression of Fto leads to increased food intake and results in obesity. Nat. Genet. *42*, 1086–1092.

Clancy, M.J., Shambaugh, M.E., Timpte, C.S., and Bokar, J. a (2002). Induction of sporulation in Saccharomyces cerevisiae leads to the formation of N6-methyladenosine in mRNA: a potential mechanism for the activity of the IME4 gene. Nucleic Acids Res. *30*, 4509–4518.

Clapier, C.R., and Cairns, B.R. (2009). The biology of chromatin remodeling complexes. Annu. Rev. Biochem. *78*, 273–304.

Clayton, C. (2013). The Regulation of Trypanosome Gene Expression by RNA-Binding Proteins. PLoS Pathog. *9*, 9–12.

Clayton, C.E. (2002). Life without transcriptional control? From fly to man and back again. EMBO *21*, 1881–1888.

Clayton, C.E. (2014). Networks of gene expression regulation in Trypanosoma brucei. Mol. Biochem. Parasitol. *195*, 96–106.

Cole, C., Barber, J.D., and Barton, G.J. (2008). The Jpred 3 secondary structure prediction server. Nucleic Acids Res. *36*, W197–W201.

Cowling, V.H. (2010). Regulation of mRNA cap methylation. Biochem. J. *425*, 295–302.

Cross, G. a M., Kim, H.-S., and Wickstead, B. (2014). Capturing the variant surface glycoprotein repertoire (the VSGnome) of Trypanosoma brucei Lister 427. Mol. Biochem. Parasitol. *195*, 59–73.

Csepany, T., Lin, a, Baldick, C.J., and Beemon, K. (1990). Sequence specificity of mRNA N6-adenosine methyltransferase. J. Biol. Chem. *265*, 20117–20122.

Czichos J, Nonnengaesser C, O.P. (1986). Trypanosoma brucei : cis-Aconitate and Temperature Reduction as Triggers of Synchronous Transformation of Bloodstream to Procyclic Trypomastigotes in Vitro. Exp. Parasitol. *291*, 283–291.

Darnellt, J.E. (1975). Methylated, Blocked 5' Termini in HeLa Cell mRNA. *72*, 1904–1908.

DeGrasse, J. a, DuBois, K.N., Devos, D., Siegel, T.N., Sali, A., Field, M.C., Rout, M.P., and Chait, B.T. (2009). Evidence for a shared nuclear pore complex architecture that is conserved from the last common eukaryotic ancestor. Mol. Cell. Proteomics *8*, 2119–2130.

Desrosiers, R., Friderici, K., and Rottman, F. (1974). Identification of methylated nucleosides in messenger RNA from Novikoff hepatoma cells. Proc. Natl. Acad. Sci. U. S. A. *71*, 3971–3975.

Dina, C., Meyre, D., Gallina, S., Durand, E., Körner, A., Jacobson, P., Carlsson, L.M.S., Kiess, W., Vatin, V., Lecoeur, C., et al. (2007). Variation in FTO contributes to childhood obesity and severe adult obesity. Nat. Genet. *39*, 724–726.

Dominissini, D., Moshitch-Moshkovitz, S., Schwartz, S., Salmon-Divon, M., Ungar, L., Osenberg, S., Cesarkas, K., Jacob-Hirsch, J., Amariglio, N., Kupiec, M., et al. (2012). Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. Nature *485*, 201–206.

Douzery, E.J.P., Snell, E. a, Bapteste, E., Delsuc, F., and Philippe, H. (2004). The timing of eukaryotic evolution: does a relaxed molecular clock reconcile proteins and fossils? Proc. Natl. Acad. Sci. U. S. A. *101*, 15386–15391.

Droll, D., Minia, I., Fadda, A., Singh, A., Stewart, M., Queiroz, R., and Clayton, C. (2013). Post-transcriptional regulation of the trypanosome heat shock response by a zinc finger protein. PLoS Pathog. *9*, e1003286.

Dyer, N. a, Rose, C., Ejeh, N.O., and Acosta-Serrano, A. (2013). Flying tryps: survival and maturation of trypanosomes in tsetse flies. Trends Parasitol. *29*, 188–196.

Eddy, S.R. (2011). Accelerated Profile HMM Searches. PLoS Comput. Biol. *7*, e1002195.

Ehrlich, M., Gama-sosa, M.A., Huang, L., Midgett, R.M., Kenneth, C., Mccune, R.A., and Gehrke, C. (1982). Amount and distribution of 5-methylcytosine in human DNA from different types of tissues of cells. *10*, 11–14.

El-Sayed, N.M., Hegde, P., Quackenbush, J., Melville, S.E., and Donelson, J.E. (2000). The African trypanosome genome. Int. J. Parasitol. *30*, 329–345.

Ersfeld, K. (2011). Nuclear architecture, genome and chromatin organisation in Trypanosoma brucei. Res. Microbiol. *162*, 626–636.

Eulalio, A., Behm-Ansmant, I., and Izaurralde, E. (2007). P bodies: at the crossroads of post-transcriptional pathways. Nat. Rev. Mol. Cell Biol. *8*, 9–22.

Falnes, P.Ø., and Rognes, T. (2003). DNA repair by bacterial AlkB proteins. Res. Microbiol. *154*, 531–538.

Färber, V., Erben, E., Sharma, S., Stoecklin, G., and Clayton, C. (2013). Trypanosome CNOT10 is essential for the integrity of the NOT deadenylase complex and for degradation of many mRNAs. Nucleic Acids Res. *41*, 1211–1222.

Fedorova, E., and Zink, D. (2008). Nuclear architecture and gene regulation. Biochim. Biophys. Acta *1783*, 2174–2184.

Fenn, K., and Matthews, K.R. (2007). The cell biology of Trypanosoma brucei differentiation. Curr. Opin. Microbiol. *10*, 539–546.

Figueiredo, L.M., Cross, G. a M., and Janzen, C.J. (2009). Epigenetic regulation in African trypanosomes: a new kid on the block. Nat. Rev. Microbiol. *7*, 504–513.

Finn, R.D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Heger, A., Hetherington, K., Holm, L., Mistry, J., et al. (2014). Pfam: the protein families database. Nucleic Acids Res. *42*, D222–30.

Franco, J.R., Simarro, P.P., Diarra, A., and Jannin, J.G. (2014). Epidemiology of human African trypanosomiasis. Clin. Epidemiol. *6*, 257–275.

Fu, Y., Jia, G., Pang, X., Wang, R.N., Wang, X., Li, C.J., Smemo, S., Dai, Q., Bailey, K. a, Nobrega, M. a, et al. (2013). FTO-mediated formation of N6-hydroxymethyladenosine and N6-formyladenosine in mammalian RNA. Nat. Commun. *4*, 1798.

Fu, Y., Dominissini, D., Rechavi, G., and He, C. (2014). Gene expression regulation mediated through reversible m^6A RNA methylation. Nat. Rev. Genet. *15*, 293–306.

Fustin, J.-M., Doi, M., Yamaguchi, Y., Hida, H., Nishimura, S., Yoshida, M., Isagawa, T., Morioka, M.S., Kakeya, H., Manabe, I., et al. (2013). RNA-methylation-dependent RNA processing controls the speed of the circadian clock. Cell *155*, 793–806.

Garneau, N.L., Wilusz, J., and Wilusz, C.J. (2007). The highways and byways of mRNA decay. Nat. Rev. Mol. Cell Biol. *8*, 113–126.

Ge, J., and Yu, Y.-T. (2013). RNA pseudouridylation: new insights into an old modification. Trends Biochem. Sci. *38*, 210–218.

Gebauer, F., and Hentze, M.W. (2004). Molecular mechanisms of translational control. Nat. Rev. Mol. Cell Biol. *5*, 827–835.

Ghildiyal, M., and Zamore, P.D. (2009). Small silencing RNAs: an expanding universe. Nat. Rev. Genet. *10*, 94–108.

Glisovic, T., Bachorik, J.L., Yong, J., and Dreyfuss, G. (2008). RNA-binding proteins and post-transcriptional gene regulation. FEBS Lett. *582*, 1977–1986.

Goedecke, K., Pignot, M., Goody, R.S., Scheidig, a J., and Weinhold, E. (2001). Structure of the N6-adenine DNA methyltransferase M.TaqI in complex with DNA and a cofactor analog. Nat. Struct. Biol. *8*, 121–125.

Goldberg, A.D., Allis, C.D., and Bernstein, E. (2007). Epigenetics: a landscape takes shape. Cell *128*, 635–638.

Gommers-ampt, J.H., Leeuwen, F. Van, and Vliegenthart, J.F.G. (1993). A Novel Modified Base Present in the DNA of the Parasitic Protozoan T . brucei. *75*, 1129–1136.

Goodstadt, L., and Ponting, C.P. (2001). CHROMA: consensus-based colouring of multiple alignments for publication. Bioinformatics *17*, 845–846.

Gunasekera, K., Wüthrich, D., Braga-Lagache, S., Heller, M., and Ochsenreiter, T. (2012). Proteome remodelling during development from blood to insect-form Trypanosoma brucei quantified by SILAC and mass spectrometry. BMC Genomics *13*, 556.

Han, Z., Niu, T., Chang, J., Lei, X., Zhao, M., Wang, Q., Cheng, W., Wang, J., Feng, Y., and Chai, J. (2010). Crystal structure of the FTO protein reveals basis for its substrate specificity. Nature *464*, 1205–1209.

He, C. (2010). RNA epigenetics? Nat. Chem. Biol. *6*, 863–865.

He, L., and Hannon, G.J. (2004). MicroRNAs: small RNAs with a big role in gene regulation. Nat. Rev. Genet. *5*, 522–531.

Helm, M., and Alfonzo, J.D. (2014). Posttranscriptional RNA Modifications: playing metabolic games in a cell's chemical Legoland. Chem. Biol. *21*, 174–185.

Hess, M.E., Hess, S., Meyer, K.D., Verhagen, L. a W., Koch, L., Brönneke, H.S., Dietrich, M.O., Jordan, S.D., Saletore, Y., Elemento, O., et al. (2013). The fat mass and obesity associated gene (Fto) regulates activity of the dopaminergic midbrain circuitry. Nat. Neurosci. *16*, 1042–1048.

Hirumi, H.H.; K. (1989). Continuous Cultivation of Trypanosoma brucei Blood Stream Forms in a Medium Containing a Low Concentration of Serum Protein without Feeder Cell Layers. *75*, 985–989.

Hongay, C.F., and Orr-Weaver, T.L. (2011). Drosophila Inducer of MEiosis 4 (IME4) is required for Notch signaling during oogenesis. Proc. Natl. Acad. Sci. U. S. A. *108*, 14855–14860.

Hudson, D.M., and Eyre, D.R. (2014). Collagen prolyl3-hydroxylation: a major role for a minor post-translational modification? *54*, 245–251.

Hulo, N., Bairoch, A., Bulliard, V., Cerutti, L., De Castro, E., Langendijk-Genevaux, P.S., Pagni, M., and Sigrist, C.J. a (2006). The PROSITE database. Nucleic Acids Res. *34*, D227–D230.

Iyer, L.M., Anantharaman, V., Wolf, M.Y., and Aravind, L. (2008). Comparative genomics of transcription factors and chromatin proteins in parasitic protists and other eukaryotes. Int. J. Parasitol. *38*, 1–31.

Jaenisch, R., and Bird, A. (2003). Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. Nat. Genet. *33 Suppl*, 245–254.

Jakovcevski, M., and Akbarian, S. (2012). Epigenetic mechanisms in neurological disease. Nat. Med. *18*, 1194–1204.

Jia, G., Fu, Y., Zhao, X., Dai, Q., Zheng, G., Yang, Y., Yi, C., Lindahl, T., Pan, T., Yang, Y.-G., et al. (2011). N6-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. Nat. Chem. Biol. *7*, 885–887.

Jia, G., Fu, Y., and He, C. (2013). Reversible RNA adenosine methylation in biological regulation. Trends Genet. *29*, 108–115.

Jones, P. a (2012). Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat. Rev. Genet. *13*, 484–492.

Juven-Gershon, T., and Kadonaga, J.T. (2010). Regulation of gene expression via the core promoter and the basal transcriptional machinery. Dev. Biol. *339*, 225–229.

Kaikkonen, M.U., Lam, M.T.Y., and Glass, C.K. (2011). Non-coding RNAs as regulators of gene expression and epigenetics. Cardiovasc. Res. *90*, 430–440.

Kennedy, P.G.E. (2004). Human African trypanosomiasis of the CNS : current issues and challenges. *113*.

Köhler, A., and Hurt, E. (2007). Exporting RNA from the nucleus to the cytoplasm. Nat. Rev. Mol. Cell Biol. *8*, 761–773.

Kolev, N.G., Franklin, J.B., Carmi, S., Shi, H., Michaeli, S., and Tschudi, C. (2010). The transcriptome of the human pathogen Trypanosoma brucei at single-nucleotide resolution. PLoS Pathog. *6*, e1001090.

Kolev, N.G., Ramey-Butler, K., Cross, G. a M., Ullu, E., and Tschudi, C. (2012). Developmental progression to infectivity in Trypanosoma brucei triggered by an RNA-binding protein. Science *338*, 1352–1353.

Kolev, N.G., Ullu, E., and Tschudi, C. (2014). The emerging role of RNA-binding proteins in the life cycle of Trypanosoma brucei. Cell. Microbiol. *16*, 482–489.

Kolovos, P., Knoch, T. a, Grosveld, F.G., Cook, P.R., and Papantonis, A. (2012). Enhancers and silencers: an integrated and simple model for their function. Epigenetics Chromatin *5*, 1.

Kouzarides, T. (2007). Chromatin modifications and their function. Cell *128*, 693–705.

Kung, J.T.Y., Colognori, D., and Lee, J.T. (2013). Long noncoding RNAs: past, present, and future. Genetics *193*, 651–669.

Kurowski, M. a, Bhagwat, A.S., Papaj, G., and Bujnicki, J.M. (2003). Phylogenomic identification of five new human homologs of the DNA repair enzyme AlkB. BMC Genomics *4*, 48.

Lemon, B. (2000). Orchestrated response: a symphony of transcription factors for gene control. Genes Dev. *14*, 2551–2569.

Li, S., and Mason, C.E. (2014). The Pivotal Regulatory Landscape of RNA Modifications. Annu. Rev. Genomics Hum. Genet. *15*, 127–150.

Liang, X., Haritan, A., and Uliel, S. (2003). trans and cis Splicing in Trypanosomatids: Mechanism, Factors, and Regulation. Eukaryot. Cell *2*, 830–840.

Liu, J., Yue, Y., Han, D., Wang, X., Fu, Y., Zhang, L., Jia, G., Yu, M., Lu, Z., Deng, X., et al. (2014). A METTL3-METTL14 complex mediates mammalian nuclear RNA N6-adenosine methylation. Nat. Chem. Biol. *10*, 93–95.

Van Luenen, H.G. a M., Farris, C., Jan, S., Genest, P.-A., Tripathi, P., Velds, A., Kerkhoven, R.M., Nieuwland, M., Haydock, A., Ramasamy, G., et al. (2012). Glucosylated hydroxymethyluracil, DNA base J, prevents transcriptional readthrough in Leishmania. Cell *150*, 909–921.

Lunde, B.M., Moore, C., and Varani, G. (2007). RNA-binding proteins: modular design for efficient function. Nat. Rev. Mol. Cell Biol. *8*, 479–490.

Machnicka, M. a, Milanowska, K., Osman Oglou, O., Purta, E., Kurkowska, M., Olchowik, A., Januszewski, W., Kalinowski, S., Dunin-Horkawicz, S., Rother, K.M., et al. (2013). MODOMICS: a database of RNA modification pathways--2013 update. Nucleic Acids Res. *41*, D262–D267.

Madera, M., and Gough, J. (2002). A comparison of profile hidden Markov model procedures for remote homology detection. Nucleic Acids Res. *30*, 4321–4328.

Mallick, B., Ghosh, Z., and Chakrabarti, J. (2008). MicroRNA switches in Trypanosoma brucei. Biochem. Biophys. Res. Commun. *372*, 459–463.

Malone, T., Blumenthal, R.M., Cheng, X., and Structural, W.M.K. (1995). Structure-guided Analysis Reveals Nine Sequence Motifs Conserved among DNA Amino-methyl- transferases , and Suggests a Catalytic Mechanism for these Enzymes. 618–632.

Manful, T., Fadda, A., and Clayton, C. (2011). The role of the 5'-3' exoribonuclease XRNA in transcriptome-wide mRNA degradation. RNA *17*, 2039–2047.

Mao, Y., Najafabadi, H.S., and Salavati, R. (2009). Genome-wide computational identification of functional RNA elements in Trypanosoma brucei. BMC Genomics *10*, 355.

Matthews, K.R. (2005). The developmental cell biology of Trypanosoma brucei. J. Cell Sci. *118*, 283–290.

Matthews, K.R., Ellis, J.R., and Paterou, A. (2004). Molecular regulation of the life cycle of African trypanosomes. Trends Parasitol. *20*, 40–47.

McGuffin, L.J., Bryson, K., and Jones, D.T. (2000). The PSIPRED protein structure prediction server. Bioinformatics *16*, 404–405.

Mercer, T.R., and Mattick, J.S. (2013). Structure and function of long noncoding RNAs in epigenetic regulation. Nat. Struct. Mol. Biol. *20*, 300–307.

Merrow, M., Spoelstra, K., and Roenneberg, T. (2005). The circadian cycle: daily rhythms from behaviour to genes. EMBO Rep. *6*, 930–935.

Meyer, K.D., and Jaffrey, S.R. (2014). The dynamic epitranscriptome: N6-methyladenosine and gene expression control. Nat. Rev. Mol. Cell Biol. *15*, 313–326.

Meyer, K.D., Saletore, Y., Zumbo, P., Elemento, O., Mason, C.E., and Jaffrey, S.R. (2012). Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. Cell *149*, 1635–1646.

Michael G. Rossmann, D.M.& K.W.O. (1974). Chemical and biological evolution of a nucleotide-binding protein. Nature *250*, 194–199.

Mignone, F., Gissi, C., Liuni, S., and Pesole, G. (2002). Untranslated regions of mRNAs. Genome Biol. *3*, REVIEWS0004.

Militello, K.T., Wang, P., Jayakar, S.K., Pietrasik, R.L., Dupont, C.D., Dodd, K., King, A.M., and Valenti, P.R. (2008). African trypanosomes contain 5-methylcytosine in nuclear DNA. Eukaryot. Cell *7*, 2012–2016.

Militello, K.T., Chen, L.M., Ackerman, S.E., Mandarano, A.H., and Valentine, E.L. (2014). A map of 5-methylcytosine residues in Trypanosoma brucei tRNA revealed by sodium bisulfite sequencing. Mol. Biochem. Parasitol. *193*, 122–126.

Moore, A.D., Björklund, A.K., Ekman, D., Bornberg-Bauer, E., and Elofsson, A. (2008). Arrangements in the modular evolution of proteins. Trends Biochem. Sci. *33*, 444–451.

Motorin, Y., Lyko, F., and Helm, M. (2010). 5-methylcytosine in RNA: detection, enzymatic formation and biological functions. Nucleic Acids Res. *38*, 1415–1430.

De Nadal, E., Ammerer, G., and Posas, F. (2011). Controlling gene expression in response to stress. Nat. Rev. Genet. *12*, 833–845.

Narayan, P., Ludwiczak, R.L., Goodwin, E.C., and Rottman, F.M. (1994). Context effects on N6-adenosine methylation sites in prolactin mRNA. Nucleic Acids Res. *22*, 419–426.

Ngô, H., Tschudi, C., Gull, K., and Ullu, E. (1998). Double-stranded RNA induces mRNA degradation in Trypanosoma brucei. Proc. Natl. Acad. Sci. U. S. A. *95*, 14687–14692.

Nilsson, D., Gunasekera, K., Mani, J., Osteras, M., Farinelli, L., Baerlocher, L., Roditi, I., and Ochsenreiter, T. (2010). Spliced leader trapping reveals widespread alternative splicing patterns in the highly dynamic transcriptome of Trypanosoma brucei. PLoS Pathog. *6*, e1001037.

Ong, C.-T., and Corces, V.G. (2011). Enhancer function: new insights into the regulation of tissue-specific gene expression. Nat. Rev. Genet. *12*, 283–293.

Palenchar, J.B., and Bellofatto, V. (2006). Gene transcription in trypanosomes. Mol. Biochem. Parasitol. *146*, 135–141.

Patrick, K.L., Shi, H., Kolev, N.G., Ersfeld, K., Tschudi, C., and Ullu, E. (2009). Distinct and overlapping roles for two Dicer-like proteins in the RNA interference pathways of the ancient eukaryote Trypanosoma brucei. Proc. Natl. Acad. Sci. U. S. A. *106*, 17933–17938.

Peacock, L., Ferris, V., Sharma, R., Sunter, J., Bailey, M., Carrington, M., and Gibson, W. (2011). Identification of the meiotic life cycle stage of Trypanosoma brucei in the tsetse fly. Proc. Natl. Acad. Sci. U. S. A. *108*, 3671–3676.

Perry, F.P., and Kelley, D.E. (1974). Existence of Methylated Mensenger RNA in Mouse L Cells. Cell *I*, 37–42.

Perry, R.P., and Kelley, D.E. (1975). The Methylated Constituents of L Cell Messenger RNA : Evidence for an Unusual Cluster at the 5 ' Terminus. Cell *4*, 387–394.

Ping, X.-L., Sun, B.-F., Wang, L., Xiao, W., Yang, X., Wang, W.-J., Adhikari, S., Shi, Y., Lv, Y., Chen, Y.-S., et al. (2014). Mammalian WTAP is a regulatory subunit of the RNA N6-methyladenosine methyltransferase. Cell Res. *24*, 177–189.

Queiroz, R., Benz, C., Fellenberg, K., Hoheisel, J.D., and Clayton, C. (2009). Transcriptome analysis of differentiating trypanosomes reveals the existence of multiple post-transcriptional regulons. *19*, 1–19.

Quijada, L., Guerra-Giraldez, C., Drozdz, M., Hartmann, C., Irmer, H., Ben-Dov, C., Cristodero, M., Ding, M., and Clayton, C. (2002). Expression of the human RNA-binding protein HuR in Trypanosoma brucei increases the abundance of mRNAs containing AU-rich regulatory elements. Nucleic Acids Res. *30*, 4414–4424.

Ratel, D., Ravanat, J.-L., Berger, F., and Wion, D. (2006). N6-methyladenine: the other methylated base of DNA. Bioessays *28*, 309–315.

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

Reynolds, D., Cliffe, L., Forstner, K.U., Hon, C.-C., Siegel, T.N., and Sabatini, R. (2014). Regulation of transcription termination by glucosylated hydroxymethyluracil, base J, in Leishmania major and Trypanosoma brucei. Nucleic Acids Res. 1–13.

Rodríguez-Navarro, S., and Hurt, E. (2011). Linking gene regulation to mRNA production and export. Curr. Opin. Cell Biol. *23*, 302–309.

Rojas, M. V, and Galanti, N. (1990). DNA methylation in Trypanosoma cruzi. FEBS Lett. *263*, 113–116.

Rudenko, G. (2010). Epigenetics and transcriptional control in African trypanosomes. Essays Biochem. *48*, 201–219.

Saha, A., Wittmeyer, J., and Cairns, B.R. (2006). Chromatin remodelling: the industrial revolution of DNA around histones. Nat. Rev. Mol. Cell Biol. *7*, 437–447.

Saletore, Y., Meyer, K., Korlach, J., Vilfan, I.D., Jaffrey, S., Mason, C.E., and Jaff, S. (2012). The birth of the Epitranscriptome: deciphering the function of RNA modifi cations. Genome Biol. *13*, 1–11.

Schimanski, B., Nguyen, T.N., Günzl, A., and Gu, A. (2005). Characterization of a Multisubunit Transcription Factor Complex Essential for Spliced-Leader RNA Gene Transcription in Trypanosoma brucei. Mol. Cell. Biol. *25*, 7303–7313.

Schneider, R., and Grosschedl, R. (2007). Dynamics and interplay of nuclear architecture, genome organization, and gene expression. Genes Dev. *21*, 3027–3043.

Schwartz, S., Agarwala, S.D., Mumbach, M.R., Jovanovic, M., Mertins, P., Shishkin, A., Tabach, Y., Mikkelsen, T.S., Satija, R., Ruvkun, G., et al. (2013). High-resolution mapping reveals a conserved, widespread, dynamic mRNA methylation program in yeast meiosis. Cell *155*, 1409–1421.

Schwartz, S., Mumbach, M.R., Jovanovic, M., Wang, T., Maciag, K., Bushkin, G.G., Mertins, P., Ter-Ovanesyan, D., Habib, N., Cacchiarelli, D., et al. (2014). Perturbation of m6A Writers Reveals Two Distinct Classes of mRNA Methylation at Internal and 5' Sites. Cell Rep. *8*, 284–296.

Shao, Z., Zhang, Y., and Powell-Coffman, J.A. (2009). Two distinct roles for EGL-9 in the regulation of HIF-1-mediated gene expression in Caenorhabditis elegans. Genetics *183*, 821–829.

Shi, H., Djikeng, A., Tschudi, C., and Ullu, E. (2004). Argonaute Protein in the Early Divergent Eukaryote Trypanosoma brucei : Control of Small Interfering RNA Accumulation and Retroposon Transcript Abundance. *24*, 420–427.

Siegel, T.N., Hekstra, D.R., Kemp, L.E., Figueiredo, L.M., Lowell, J.E., Fenyo, D., Wang, X., Dewell, S., and Cross, G. a M. (2009). Four histone variants mark the boundaries of polycistronic transcription units in Trypanosoma brucei. Genes Dev. *23*, 1063–1076.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. *7*, 539.

Simmons, J.M., Koslowsky, D.J., and Hausinger, R.P. (2012). Characterization of a Trypanosoma brucei Alkb homolog capable of repairing alkylated DNA. Exp. Parasitol. *131*, 92–100.

Steverding, D. (2008). The history of African trypanosomiasis. Parasit. Vectors *1*.

Strambio-De-Castillia, C., Niepel, M., and Rout, M.P. (2010). The nuclear pore complex: bridging nuclear transport and gene regulation. Nat. Rev. Mol. Cell Biol. *11*, 490–501.

Taylor, J.E., and Rudenko, G. (2006). Switching trypanosome coats: what's in the wardrobe? Trends Genet. *22*, 614–620.

Vasquez, J.-J., Hon, C.-C., Vanselow, J.T., Schlosser, A., and Siegel, T.N. (2014). Comparative ribosome profiling reveals extensive translational complexity in different Trypanosoma brucei life cycle stages. Nucleic Acids Res. *42*, 3623–3637.

Vassella, E., and Boshart, M. (1996). High molecular mass agarose matrix supports growth of bloodstream forms of pleomorphic Trypanosoma brucei strains in axenic culture. Mol. Biochem. Parasitol. *82*, 91–105.

Veitch, N.J., Johnson, P.C.D., Trivedi, U., Terry, S., Wildridge, D., and MacLeod, A. (2010). Digital gene expression analysis of two life cycle stages of the human-infective parasite, Trypanosoma brucei gambiense reveals differentially expressed clusters of co-regulated genes. BMC Genomics *11*, 124.

Wang, H., Dong, S., Xu, H., Qian, J., and Yang, J. (2012). Genetic variants in FTO associated with metabolic syndrome: a meta- and gene-based analysis. Mol. Biol. Rep. *39*, 5691–5698.

Wang, X., Lu, Z., Gomez, A., Hon, G.C., Yue, Y., Han, D., Fu, Y., Parisien, M., Dai, Q., Jia, G., et al. (2014a). N6-methyladenosine-dependent regulation of messenger RNA stability. Nature *505*, 117–120.

Wang, Y., Li, Y., Toth, J.I., Petroski, M.D., Zhang, Z., and Zhao, J.C. (2014b). N6-methyladenosine modification destabilizes developmental regulators in embryonic stem cells. Nat. Cell Biol. *16*, 191–198.

Waterhouse, A.M., Procter, J.B., Martin, D.M. a, Clamp, M., and Barton, G.J. (2009). Jalview Version 2--a multiple sequence alignment editor and analysis workbench. Bioinformatics *25*, 1189–1191.

WHO (2012). Research priorities for Chagas disease, human African trypanosomiasis and leishmaniasis. World Health Organ. Tech. Rep. Ser.

Wilusz, C.J., and Wilusz, J. (2004). Bringing the role of mRNA decay in the control of gene expression into focus. Trends Genet. *20*, 491–497.

Wilusz, J.E., Sunwoo, H., and Spector, D.L. (2009). Long noncoding RNAs: functional surprises from the RNA world. Genes Dev. *23*, 1494–1504.

Wion, D., and Casadesús, J. (2006). N6-methyl-adenine: an epigenetic signal for DNA-protein interactions. Nat. Rev. Microbiol. *4*, 183–192.

Yun, M., Wu, J., Workman, J.L., and Li, B. (2011). Readers of histone modifications. Cell Res. *21*, 564–578.

Zamudio, J.R., Mittra, B., Chattopadhyay, A., Wohlschlegel, J. a, Sturm, N.R., and Campbell, D. a (2009). Trypanosoma brucei spliced leader RNA maturation by the cap 1 2'-O-ribose methyltransferase and SLA1 H/ACA snoRNA pseudouridine synthase complex. Mol. Cell. Biol. *29*, 1202–1211.

Zheng, G., Dahl, J.A., Niu, Y., Fedorcsak, P., Huang, C.-M., Li, C.J., Vågbø, C.B., Shi, Y., Wang, W.-L., Song, S.-H., et al. (2013a). ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility. Mol. Cell *49*, 18–29.

Zheng, L., Wen, Y., Yang, J., Liao, J., Shao, P., and Xu, H.U.I. (2013b). Comparative transcriptome analysis of small noncoding RNAs in different stages of Trypanosoma brucei Trypanosoma brucei. RNA *19*, 1–13.

Zhong, S., Li, H., Bodi, Z., Button, J., Vespa, L., Herzog, M., and Fray, R.G. (2008). MTA is an Arabidopsis messenger RNA adenosine methylase and interacts with a homolog of a sex-specific splicing factor. Plant Cell *20*, 1278–1288.

7. Annexes

Supplementary Table 1. Primers to amplify the inserts cloned in pIV vectors. In the third column, integration means that is the regions that flank the target genes to promote the integration in the genome. The 15 bp identical are in bold, and indicates the 15 bp identical to the vector backbone, G418 gene or hygromycin gene, to clone the vectors through the In-fusion system.

Target gene	Sequence (5'→3')	Product / 15 bp identical	Fw/Rv
Tb927.5.980	**TATAGGGCGAATTGG**GTACCACCTAAAGTGGATCCAGTGC	5' integration/Plasmid	Fw
Tb927.5.980	**CGCTACACAGCTTGA**AGAGCAGAGGTAGCACAACC	5' integration/G418	Rv
Tb927.5.980	**AACGGAAGAGTGAAA**CTTGTTCCCCTCCCGTAGG	3' integration/G418	Fw
Tb927.5.980	**ACCGCGGTGGCGGCC**GCACTGTTATAGTTCCTGGCGCC	3' integration/Plasmid	Rv
Tb927.5.980	**AGACCTTGCTGTGCC**AGAGCAGAGGTAGCACAACC	5' integration/Hygro	Rv
Tb927.5.980	**TTATCTATGCAGTAT**CTTGTTCCCCTCCCGTAGG	3' integration/Hygro	Fw
Tb927.4.460	**TATAGGGCGAATTGG**GTACCGCTGATGGGTGTCTTTTAACC	5' integration/Plasmid	Fw
Tb927.4.460	**CGCTACACAGCTTGA**AATGAGAGATGCAGAAAGTGG	5' integration/G418	Rv
Tb927.4.460	**AACGGAAGAGTGAAA**CGGTTTGTCGTTTGGTCTTG	3' integration/G418	Fw
Tb927.4.460	**ACCGCGGTGGCGGCC**GCTCAGTGCAAAATCGAAAGCG	3' integration/Plasmid	Rv
Tb927.4.460	**AGACCTTGCTGTGCC**AATGAGAGATGCAGAAAGTGG	5' integration/Hygro	Rv
Tb927.4.460	**TTATCTATGCAGTAT**CGGTTTGTCGTTTGGTCTTG	3' integration/Hygro	Fw
Tb927.7.6620	**TATAGGGCGAATTGG**GTACCTGCACTGCAGAGACGAAGG	5' integration/Plasmid	Fw
Tb927.7.6620	**CGCTACACAGCTTGA**GACAATGTGAAGTTGCGAAGG	5' integration/G418	Rv
Tb927.7.6620	**AACGGAAGAGTGAAA**GAAGAGGAGCGTTGCTGTGG	3' integration/G418	Fw
Tb927.7.6620	**ACCGCGGTGGCGGCC**GCCCACCTTCGGCTAGTGTTGG	3' integration/Plasmid	Rv
Tb927.7.6620	**AGACCTTGCTGTGCC**GACAATGTGAAGTTGCGAAGG	5' integration/Hygro	Rv
Tb927.7.6620	**TTATCTATGCAGTAT**GAAGAGGAGCGTTGCTGTGG	3' integration/Hygro	Fw
G418	TCAAGCTGTGTAGCGCACG	Gene	Fw
G418	TTTCACTCTTCCGTTGCACC	Gene	Rv
Hygromycin	GGCACAGCAAGGTCTTCTG	Gene	Fw
Hygromycin	ATACTGCATAGATAACAAACGC	Gene	Rv

Supplementary Table 2. Primers that hybridize upstream and downstream of the target locus, outside of the integration regions.

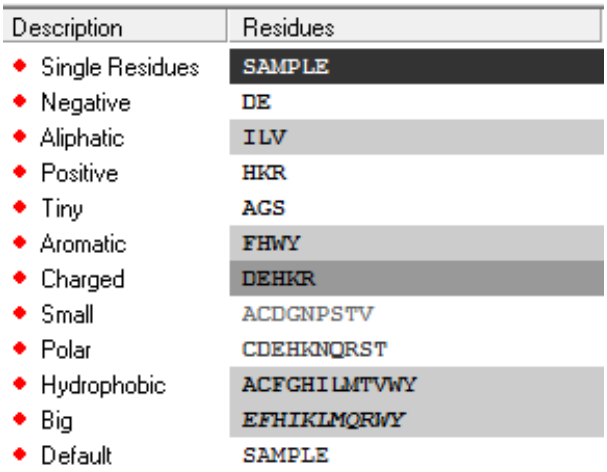
Target Gene	Primers	
	Foward	Reverse
Tb927.5.980	GTTCGGAAAAGGAAGGATGC	AGGATTGGTATCCCCACTGC
Tb927.4.460	GCAGAGGAGGAGACGGAGG	CGGTTCCGTTGTGCAGTCC
Tb927.7.6620	CCAAGTAAGCGGTTAGGAGG	CTCCGCCGCTTTAATTCC

Supplementary Table 3. Primers to amplify the integration regions in knockout cell lines.

Substrate	Integration	Foward	Reverse	Expected Size (bp)
IV2	5' NEO	GTTCGGAAAAGGAAGGATGC	ACCGTAAAGCACGAGGAAGC	1406
IV2	3' NEO	CTTGCCGAATATCATGGTGGA	AGGATTGGTATCCCCACTGC	839
IV2	5' HYG	GTTCGGAAAAGGAAGGATGC	GTCGGTTTCCACTATCGGCG	1614
IV2	3' HYG	GCCGATAGTGGAAACCGAC	AGGATTGGTATCCCCACTGC	759
IV4	5' NEO	GCAGAGGAGGAGACGGAGG	ACCGTAAAGCACGAGGAAGC	1347
IV4	3' NEO	CTTGCCGAATATCATGGTGGA	CGGTTCCGTTGTGCAGTCC	841
IV4	5' HYG	GCAGAGGAGGAGACGGAGG	GTCGGTTTCCACTATCGGCG	1555
IV4	3' HYG	GCCGATAGTGGAAACCGAC	CGGTTCCGTTGTGCAGTCC	763
IV6	5' NEO	CCAAGTAAGCGGTTAGGAGG	ACCGTAAAGCACGAGGAAGC	1807
IV6	3' NEO	CTTGCCGAATATCATGGTGGA	CTCCGCCGCTTTAATTCC	1095
IV6	5' HYG	CCAAGTAAGCGGTTAGGAGG	GTCGGTTTCCACTATCGGCG	2015
IV6	3' HYG	GCCGATAGTGGAAACCGAC	CTCCGCCGCTTTAATTCC	1015

Supplementary Table 4. Primers to sequence the inserts cloned in pIV vectors.# primer used only in pIV5 sequencing.

Plasmid	Foward	Reverse
pIV1,pIV3 and pIV5	TAATACGACTCACTATAGGG	AATTAACCCTCACTAAAGGG
	CTTGCCGAATATCATGGTGGA	---
	ATCGCCTTCTATCGCCTTC	---
	ATGGATTGCACGCAGGTTC	---
	AACGGAAGAGTGAAAGAAGAGGAGCGTTGCTGTGG[#]	---
pIV2,pIV4 and pIV6	TAATACGACTCACTATAGGG	AATTAACCCTCACTAAAGGG
	CCTGACCTATTGCATCTCCC	ATACTGCATAGATAACAAACGC
	GGCACAGCAAGGTCTTCTG	---



Supplementary figure 1. Correspondence of amino acid properties and colour in multiple sequence alignments. Colouring scheme is from Chroma program with black and white default settings.