JOURNAL OF PROTEOMICS 91 (2013) 210–225

Available online at www.sciencedirect.com

SciVerse ScienceDirect

www.elsevier.com/locate/jprot

Global proteomic analysis in trypanosomes reveals unique proteins and conserved cellular processes impacted by arginine methylation

Kaylen Lott[a], Jun Li[b], John C. Fisk[a], Hao Wang[b], John M. Aletta[c], Jun Qu[b], Laurie K. Read[a,*]

[a]Department of Microbiology Immunology, School of Medicine and Biomedical Sciences, University at Buffalo, State University of New York, Buffalo, NY, USA
[b]Department of Pharmaceutical Sciences, University at Buffalo, Buffalo, NY, USA
[c]CH3 BioSystems, New York State Center of Excellence in Bioinformatics and Life Sciences, Buffalo, NY, USA

ARTICLE INFO

Article history:
Received 25 April 2013
Accepted 7 July 2013

Keywords:
Arginine methylation
Trypanosomes
Mass spectrometry
RNA processing

ABSTRACT

Arginine methylation is a common posttranslational modification with reported functions in transcription, RNA processing and translation, and DNA repair. Trypanosomes encode five protein arginine methyltransferases, suggesting that arginine methylation exerts widespread impacts on the biology of these organisms. Here, we performed a global proteomic analysis of *Trypanosoma brucei* to identify arginine methylated proteins and their sites of modification. Using an approach entailing two-dimensional chromatographic separation and alternating electron transfer dissociation and collision induced dissociation, we identified 1332 methylarginines in 676 proteins. The resulting data set represents the largest compilation of arginine methylated proteins in any organism to date. Functional classification revealed numerous arginine methylated proteins involved in flagellar function, RNA metabolism, DNA replication and repair, and intracellular protein trafficking. Thus, arginine methylation has the potential to impact aspects of *T. brucei* gene expression, cell biology, and pathogenesis. Interestingly, pathways with known methylated proteins in higher eukaryotes were identified in this study, but often different components of the pathway were methylated in trypanosomes. Methylarginines were often identified in glycine rich contexts, although exceptions to this rule were detected. Collectively, these data inform on a multitude of aspects of trypanosome biology and serve as a guide for the identification of homologous arginine methylated proteins in higher eukaryotes.

Biological significance
T. brucei is a protozoan parasite that causes lethal African sleeping sickness in humans and nagana in livestock, thereby imposing a significant medical and economic burden on sub-Saharan Africa. The parasite encounters very different environments as it cycles between mammalian and insect hosts, and must exert cellular responses to these varying milieus. One mechanism by which all cells respond to changing environments is through posttranslational modification of proteins. Arginine methylation is one such modification that can dramatically impact protein–protein and protein–nucleic acid interactions and subcellular localization of proteins. To define the breadth of arginine methylation in

* *Corresponding author*. Tel.: +1 7168295427; fax: +1 7168293534.
E-mail address: lread@buffalo.edu (L.K. Read).

http://dx.doi.org/10.1016/j.jprot.2013.07.010

trypanosomes and identify target proteins, we performed a global proteomic analysis of arginine methylated proteins in insect stage *T. brucei*. We identified 1332 methylarginines in 676 proteins, generating the largest compilation of methylarginine containing proteins in any organism to date. Numerous arginine methylated proteins function in RNA and DNA related processes, suggesting this modification can impact *T. brucei* genome integrity and gene regulation at numerous points. Other processes that appear to be strongly influenced by arginine methylation are intracellular protein trafficking, signaling, protein folding and degradation, and flagellar function. The widespread nature of arginine methylation in trypanosomes highlights its potential to greatly affect parasite biology and pathogenesis.

1. Introduction

Trypanosoma brucei, the etiologic agent of African sleeping sickness, undergoes a complex life cycle involving both mammalian and insect hosts. During its life cycle, the parasite must respond to numerous signals as it differentiates into multiple developmental forms, each with distinctive patterns of gene expression and cell biology [1]. Amazingly, trypanosomes achieve this task in the absence of transcriptional regulation [2,3]. Instead, regulation of RNA stability and possibly alternative pre-mRNA *trans*-splicing shapes the set of translatable mRNAs, while differential translation efficiency and protein stability affects the quantities of proteins generated from the transcript pool. Another factor that can profoundly alter protein function and, thus, gene expression and cell biology is posttranslational modification of proteins. For example, histone lysine methylation plays important roles in replication control in trypanosomes [4] and in the telomeric silencing and variant surface glycoprotein switching crucial to their evasion of host immune defenses [5,6]. A phosphorylation cascade required for differentiation from bloodstream stumpy to procyclic forms (PFs) has been identified in *T. brucei* [7], and cytokinesis in slender bloodstream forms (BFs) is regulated by a family of kinases [8]. Nearly 500 phosphoproteins were identified using a global proteomic approach in bloodstream form *T. brucei* [9]. However, much remains to be discovered regarding the types and scope of posttranslational modifications and their roles in trypanosome biology.

Arginine methylation is a common posttranslational modification with impacts on diverse cellular functions including transcriptional regulation, pre-mRNA splicing, DNA repair, and signal transduction [10]. Three major methylated arginine species exist: ω-N^G,N^G-asymmetric dimethylarginine (ADMA) in which two methyl groups are added to the same terminal nitrogen; ω-N^G,N'^G-symmetric dimethylarginine (SDMA) in which two methyl groups are added to adjacent terminal nitrogens of a given arginine residue; and ω-N^G-monomethylarginine (MMA) bearing only one methyl group. ADMA and SDMA are catalyzed by Type I and Type II protein arginine methyltransferases (PRMTs), respectively. MMA is generated as an intermediate by Type I and Type II PRMTs and as a terminal product by Type III PRMTs [10,11]. Functional studies in several systems have revealed arginine methylation as a significant regulator of interactome dynamics, with both positive and negative impacts on protein–protein and, less often, protein–nucleic acid associations [10,11]. While methylation does not affect the charge of the modified arginine, it increases its mass. Interestingly, although it has been widely thought that methylation decreases hydrogen bonding potential, recent evidence suggests that carbon atoms are also capable of forming hydrogen bonds [12]. Thus, arginine methylation has the potential to either enhance or reduce hydrogen-bonding interactions and thereby impact protein–protein and protein–nucleic acid interactions both positively and negatively.

Despite the reported widespread effects of arginine methylation, only a limited number of methylarginine containing proteins and specific methylarginine residues have been identified in any organism. A pioneering study by Boisvert, et al. [13] identified over 200 proteins immunoprecipitated by anti-methylarginine antibodies in HeLa cells; however, arginine methylated proteins could not be distinguished from proteins in complex with methylated proteins and sites of methylation were not determined. More recently Uhlmann et al. [14] used SILAC to identify 241 methylarginine sites in 131 proteins, thereby expanding the known repertoire of this modified protein class. We recently reported the first comprehensive study of arginine methylated proteins in mitochondria, mapping sites of methylarginine within 167 proteins from the mitochondria of PF *T. brucei* [15]. Mitochondrial arginine methylated proteins span diverse functional classes including metabolism, RNA processing, and kDNA replication. *T. brucei* possesses five PRMTs, including two Type I enzymes, one Type II enzyme, one Type III enzyme, and one enzyme that remains uncharacterized [16-20]. Knockdown of the Type I enzyme, TbPRMT1, affects mitochondrial gene expression, apparently through numerous effector proteins [21,22]. TbPRMT6, the other characterized Type I enzyme, is essential for normal growth of BF and PF and is implicated in cytokinesis [18]. To further probe potential roles of arginine methylation in the biology of trypanosomes, we used a global proteomic approach to identify arginine methylated proteins and map their sites of methylation in PF *T. brucei*. Using our previously described approach [15], which entails dual-enzyme proteolysis, efficient two-dimensional chromatographic separation, and an activation strategy alternating electron transfer dissociation (ETD) and collision induced dissociation (CID), we identified 1332 methylarginines in 676 proteins. Arginine methylated proteins are found in essentially all cellular compartments and in an array of biological pathways. For example, proteins involved in flagellar structure and function, RNA biology, protein trafficking, signaling, protein folding and degradation, and DNA repair were abundant in the set of arginine methylated proteins. A large percentage of arginine methylated proteins identified here were hypothetical proteins unique to kinetoplastids. In other cases, arginine methylation appeared to impact pathways in trypanosomes that are also affected by this modification in higher

eukaryotes, albeit through distinct components of a given pathway. Methylarginine was most often found in glycine rich contexts as described in other systems, although exceptions to this rule were identified. The present study greatly expands the known targets of arginine methylation in biology. It thus informs a multitude of aspects of trypanosome biology and serves as a guide for the identification of homologous arginine methylated proteins in higher eukaryotes.

2. Materials and methods

2.1. Antibodies

ASYM24 and SYM11 antibodies [13], with specificity for ADMA and SDMA, respectively, were purchased from EMD Millipore. mRG antibodies [23], which were raised against an RG peptide containing ADMA at every arginine residue, were purchased from CH3 Biosystems, Buffalo, NY. mRG antibodies specifically recognize ADMA relative to unmodified arginine, although their reactivity against SDMA or MMA is unknown. Anti-myc was purchased from ICL Labs. Antibodies against TbRND were described previously [24].

2.2. *T. brucei* cell culture

PF *T. brucei* strain 29–13 (Dr. George A.M. Cross, Rockefeller University), which contain integrated genes for the T7 RNA polymerase and the tet repressor, and its transgenic derivatives described below, were grown in SM media supplemented with 10% fetal bovine serum (FBS) as indicated previously [25]. BF single-marker *T. brucei* cells (Dr. George A.M. Cross, Rockefeller University) were cultured in HMI-9 medium supplemented with 10% FBS and 10% Serum Plus (SAFC) [26]. For creation of cells expressing Myc-His-TAP (MHT)-tagged proteins, genes encoding Tb927.5.4420, Tb927.4.2030, and Tb11.01.5680 were cloned into the pLew100 plasmid harboring a C-terminal Myc-His-TAP (MHT) tag and transfected into PF 29–13 *T. brucei*. Transformants were selected with 1.0 μg/ml of puromycin. Synthesis of MHT-tagged proteins was induced by addition of 2.5 μg/ml tetracycline to the growth medium for 4 days.

2.3. Immunoprecipitation and western blot analysis

For analysis of arginine methylation patterns in different life cycle stages, *T. brucei* PF 29–13 and BF SM cells were lysed in SDS-PAGE sample buffer and resolved via 10% SDS-PAGE. 5×10^6 cell equivalents were loaded in each lane and resulting blots were probed with the anti-methylarginine antibodies ASYM24 (diluted 1:1000), mRG (diluted 1:1000), and SYM11 (diluted 1:1000). For western blot confirmation of arginine methylation in selected proteins, 1×10^{10} cells expressing MHT-tagged versions of Tb927.5.4420, Tb927.4.2030, Tb11.01.5680, and TbRND (Tb09.211.3670) [24] were lysed in 50 mM HEPES (pH 7.5), 150 mM KCl, 150 mM and 3 mM $MgCl_2$. Total cell lysates from each transgenic cell line were incubated with IgG Sepharose resin (GE Healthcare Life Sciences), and beads were washed in buffer containing 20 mM Tris (pH 7.7), 150 mM KCl, and 3 mM $MgCl_2$ before cleavage with AcTEV protease (Invitrogen). Five percent of each elution was separated by 10% SDS-PAGE, followed by immunoblot with the methyl-specific antibodies indicated above, anti-myc antibody (diluted 1:2000), or normal rabbit serum (diluted 1:500). For all western blots, HRP-conjugated anti-rabbit secondary antibody was used at a dilution of 1:5000 and detection was performed with the ECL kit (Pierce).

2.3.1. *Protein extraction, sample cleanup and digestion*

Sample preparation was carried out essentially as described in [15]. Briefly, mitochondrially enriched PF *T. brucei* extracts were on-pellet-digested using both trypsin and GluC. For GluC digestion, pellet dissolution was facilitated by adding the RapGest cleavable detergent to a final concentration of 0.1%. The digest was immediately fractionated with SCX chromatography.

2.3.2. *Two-dimensional chromatography coupled to CID/ETD analysis*

Offline 2-D chromatography was used to resolve the complex proteome prior to MS analysis. A Waters 2796 Bioseparations HPLC system (Milford, MA) was used for SCX fractionation. Separation was performed by injecting 600 μL of the tryptic or GluC peptides onto a Thermo Scientific BioBasic SCX column (4.6 × 250 mm, 5-μm particle size) at a flow rate of 1 ml/min. The optimized mobile phases for SCX column were A: 3 mM ammonium formate in 25% acetonitrile, pH 3.0 and B: 400 mM ammonium formate in 25% acetonitrile, pH 4.5. Fractions were collected into tubes every 2 min after the start of the gradient. A total of 40 fractions were collected for each sample, but the fractions with low-peptide contents (based on the absorbance at 280 nm) were combined, which resulted in 25 total fractions. All fractions were lyophilized and reconstituted in 2% acetonitrile containing 0.1% formic acid.

The Nano-RPLC system consisted of a Spark Endurance autosampler (Emmen, Holland) and an ultra-high pressure Eksigent (Dublin, CA) Nano-2D Ultra capillary/nano-LC system. A nano-LC/nanospray setup was used to obtain a comprehensive separation of the complex peptide mixture. Mobile phases A and B were 0.1% formic acid in 2% acetonitrile and 0.1% formic acid in 88% acetonitrile, respectively. Samples were loaded onto a large-ID trap (300 μm ID × 1 cm, packed with Zorbax 3-μm C18 material) with 1% B at a flow rate of 10 μL/min. A series of nanoflow gradients was used to back-flush the trapped samples onto the nano-LC column (75 μm ID × 75 cm, packed with Pepmap 3-μm C18 material). Two different gradient elution profiles were employed: (i) a linear increase from 3% to 8% B over 5 min; (ii) an increase from 8 to 27% B over 65 min; (iii) an increase from 27 to 45% B over 30 min; (iv) an increase from 45 to 98% B over 20 min; and (v) isocratic at 98% B for 20 min. A shallower gradient was used to resolve complex samples: (i) 3 to 8% B over 5 min; (ii) 8 to 24% B over 145 min; (iii) 24 to 38% B over 95 min; (iv) 38 to 63% B over 55 min; (v) 63 to 97% B in 35 min, and finally (vi) isocratic at 97% B for 20 min.

An LTQ/Orbitrap-ETD hybrid mass spectrometer (Thermo Fisher Scientific, San Jose, CA) operating under data-dependent product ion mode was used for protein identification. One scan cycle included an MS1 scan followed by either MS2 scans by alternating CID and ETD, to fragment the three most abundant precursors found in the MS1 spectrum. The target value for MS1

by Orbitrap was 8×10^6, under which the Orbitrap was finely tuned for mass accuracy and FT transmission. For CID, the activation time was 30 ms, the isolation width was 1.5 amu, the normalized activation energy was 35%, and the activation q was 0.25. As for ETD, a mixture of helium and nitrogen (25% helium and 75% nitrogen, purity > 99.995%) was used as the reaction gas. The ETD reaction time was set at 110 ms and the isolation width was 2 amu for the precursor and 10 amu for the fluoranthene anions; supplemental activation, which uses a short CID activation process to dissociate the charge-stripped precursors, was used to enhance the fragmentation efficiency for doubly charged precursors. The AGC value of fluoranthene anions was set at 5×10^5.

2.3.3. *Database search and protein/peptide identification*

Tandem mass spectra that were embedded in Thermo .raw files were extracted by Bioworks 3.3.1 SP1 QF31448 (Thermofisher Scientific) and dta files were generated using ZSA and Combolon filters. Charger.exe (Thermofisher Scientific) was used to preprocess the ETD data. The filtered dta files were searched on Sequest Cluster 3.3.1 with a 64-CPU license (Thermofisher Scientific), against the TriTryp database containing 11425 gene entries (ver 10-20-2010). The search parameters are as following: 15 ppm tolerance for precursor ion masses and 1.0 Da for fragment ion masses to process all CID and ETD data. The fasta database was indexed for the tryptic and GluC peptides with the assumption of fully enzymatic cleavage at both ends, and two and five missed cleavages were permitted respectively for trypsin and GluC. In terms of modifications, carbamidomethylation of cysteines was specified as a static modification, and methionine oxidation, di-methylation, and mono-methylation on arginine were specified as variable modifications.

Scaffold (version Scaffold_4.0.4, Proteome Software Inc., Portland, OR) was used to validate MS/MS based peptide and protein identifications. High Xcorr and deltaCN cutoffs were used to yield a peptide FDR of 0.5%. Protein identifications were accepted if they could be established by at least 2 unique peptides identified using the above criteria. Proteins that contained similar peptides and could not be differentiated based on MS/MS analysis alone were grouped to satisfy the principles of parsimony.

3. Results and discussion

3.1. *Arginine methylated proteins are regulated during the T. brucei life cycle*

We previously detected numerous arginine methylated proteins in whole cell extracts of PF *T. brucei* by immunoblot analysis with anti-methylarginine antibodies, and showed that a distinct set of arginine methylated proteins is enriched in mitochondria [15]. To further probe the global occurrence of arginine methylation in *T. brucei*, we asked whether production of arginine methylated proteins is regulated during the *T. brucei* life cycle by immunoblot of total PF and BF cell extracts. Antibodies ASYM24 and mRG both recognize ADMA, although within different contexts, and SYM11 detects SDMA [13,23,27]. As seen in Fig. 1, both PF and BF cells contain proteins that are recognized by the anti-methylarginine antibodies, with different patterns emerging for each antibody due to their specificities. Although we detect only 20–25 arginine methylated proteins, this is a gross underestimation of the total number of such proteins due to the limited binding properties of these antibodies [13,23,27]. Notably, life cycle stage specific patterns of expression of proteins harboring methylated arginines are apparent with each of the three antibodies. For example, 45 and 16 kDa asymmetrically dimethylated proteins detected by ASYM24 and mRG, respectively, appear to be PF specific. A 32 kDa symmetrically dimethylated protein detected by SYM11 is apparently present only in BF. We conclude that numerous arginine methylated proteins are present in both PF and BF *T. brucei*, some of which may have life cycle specific functions.

3.2. *Two-dimensional liquid chromatography (LC) coupled to CID/ETD mass spectrometry reveals abundant arginine methylation in T. brucei*

Using a highly sensitive method that couples two-dimensional LC to CID/ETD mass spectrometry [15], we analyzed partially purified extracts of PF *T. brucei* for peptides harboring methylarginine. We employed a stringent set of criteria for protein identification from ETD or CID spectra [15]. In addition to using the cutoff criteria, we manually inspected all ETD spectra to rule out false-positive identification and posttranslational modification localization. We previously reported the identification of 167 mitochondrially localized proteins from this mitochondrially enriched fraction [15]. However, due to the incomplete nature of the enrichment procedure, which is in part due to fragmentation and resealing of the single large mitochondrion, combined with the sensitive nature of our mass

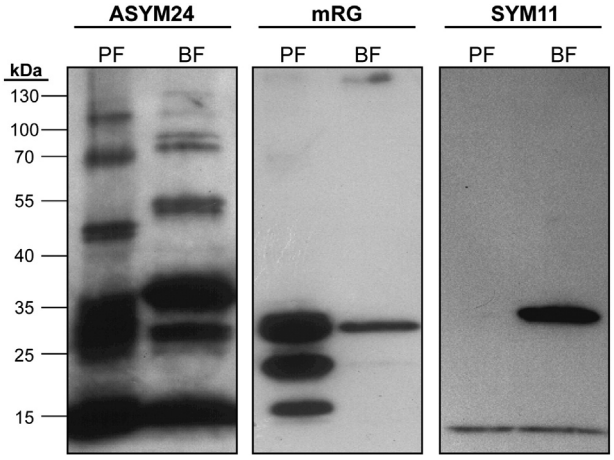

Fig. 1 – Different arginine methylated proteins are present in *T. brucei* procyclic and bloodstream form life stages. *T. brucei* procyclic (PF) and bloodstream form (BF) cells were lysed in SDS-PAGE sample buffer and resolved via 10% SDS-PAGE. 5×10^6 cell equivalents were loaded in each lane. Western blot analysis was carried out using three different anti-methylarginine antibodies: ASYM24 and mRG detect ADMA and SYM11 detects SDMA.

spectrometry analysis, we also identified a large number of proteins of non-mitochondrial origin. Proteins denoted here as "non-mitochondrial" include hypothetical proteins that did not meet our criteria for mitochondrial localization as well as those that are homologues of proteins with cytoplasmic or nuclear localization in other organisms and trypanosome proteins whose localization has been previously reported. While we cannot rule out the possibility that some hypothetical proteins not meeting our criteria for mitochondrial localization could in fact be mitochondrial, for simplicity we will refer to the set of proteins described above hereafter as "non-mitochondrial". Within this set of non-mitochondrial proteins, we identified 1332 methylarginines in 838 unique peptides representing 676 proteins (Table S1). Representative fragmentation spectra are shown in Fig. S1. This study represents by far the largest compilation of methylarginine-containing proteins in any organism to date. Indeed, together with the 167 mitochondrial arginine methylated proteins previously identified [15], the data reported here demonstrate that approximately 10% of the *T. brucei* proteome contains methylarginine [28]. This suggests that arginine methylation plays a key role in the biology of one of the earliest branching eukaryotes.

The most comprehensive analysis of arginine methylation prior to this study is that of Uhlmann et al. [14], who identified 131 arginine methylated proteins, including the sites of methylation in human T cells using SILAC labeling. Although some cellular pathways that are impacted by arginine methylation in mammals were also similarly implicated in our study (see below), few homologous methylarginine-containing proteins are evident between the two studies, likely due in part to the evolutionary divergence of trypanosomes. Nevertheless, many of the proteins identified here as containing methylarginine do have homologues in mammals, and may yet be discovered to be methylated in other organisms. Thus, in addition to informing trypanosome biology, our large expansion of known arginine methylated proteins may also be useful as a prediction tool for methylation substrates in higher eukaryotes.

3.3. *Co-immunoprecipitation validates mass spectrometry findings*

To validate our mass spectrometry results, we performed co-immunoprecipitations on a subset of the identified methylarginine-containing proteins. We generated three transgenic cell lines, each expressing a Myc-His-TAP (MHT)-tagged version of a protein that was shown by mass spectrometry to contain methylarginine residues. Tagged proteins were precipitated using IgG resin, the Protein A moieties were removed by Tev protease cleavage, and the resulting proteins were analyzed by anti-methylarginine immunoblot. A positive signal using anti-myc antibody ensured that each protein was present in its specific eluate (Fig. 2). Each of the three proteins was recognized by a subset of methyl-specific antibodies. Tb927.5.4420, annotated as a putative nucleolar RNA helicase II, contained five sites of methylation, including MMA, ADMA, and DMA whose type could not be resolved in our mass spectrometry analysis. This protein was detected by ASYM24, thereby confirming the presence of ADMA. While the mRG antibody also recognizes

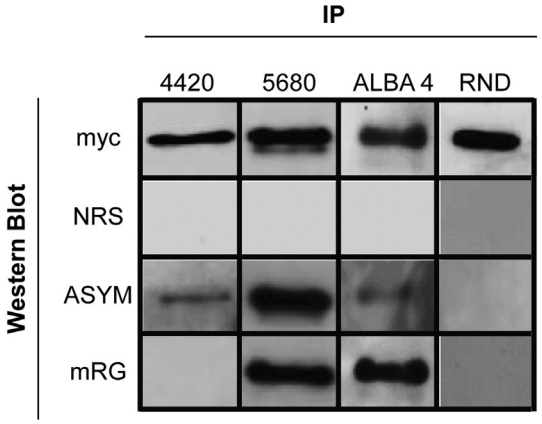

Fig. 2 – Immunoblot confirmation of methylarginine-containing proteins. Proteins from procyclic form cells expressing Myc-His-TAP tagged Tb927.5.4420 (4420), Tb11.01.5680 (5680), TbALBA 4 (ALBA 4), or TbRND (RND) were immunoprecipitated using IgG resin. The TEV eluates were then subjected to western blot analysis with anti-myc antibodies to detect the tagged protein or ASYM24 (ASYM) and mRG antibodies to detect ADMA. Normal rabbit serum (NRS) and TbRND were used as negative controls. IP, immunoprecipitation.

ADMA, it was raised against a peptide containing only RG repeats [23]. Two of three ADMA residues in Tb927.5.4420 reside within a glycine-rich region; however, the absence of extensive RG repeats presumably precludes detection by the mRG antibody. Tb11.01.5680, a hypothetical protein with two RNA recognition motifs (RRMs), was identified by mass spectrometry as containing both MMA and uncharacterized DMA. Methylated within an RGRG context, Tb11.01.5680 was efficiently recognized by both ASYM24 and mRG antibodies, indicating the presence of ADMA in Tb11.01.5680. Similarly, numerous sites of unclassified DMA and MMA were identified in the RNA binding protein, ALBA4 [29,30], within a region harboring RG and RGG repeats, and this protein was also recognized by both ASYM24 and mRG, indicating asymmetric dimethylation of arginine within this protein. Control reactions included immunoblots of presumed arginine methylated proteins with normal rabbit serum, which failed to generate signal as expected. Additionally, TbRND, a protein that apparently lacks methylarginine based on 70% mass spectrometry coverage [24], failed to react with any of the anti-methylargine antibodies. Together, these data biochemically validate a subset of our mass spectrometry results and reveal the presence of ADMA in two proteins for which mass spectrometry was not able to classify the type of methylarginine.

3.4. *Functional classification of arginine methylated proteins*

Our mass spectrometry data indicate that over 10% of the proteins encoded in the *T. brucei* genome are modified by arginine methylation [28]. To begin to discern the relevance of arginine methylation in this parasite, we utilized the fully annotated genome (TriTrypDB) to identify correlations between

arginine methylation and subcellular localization and/or biological function. Initially, we classified arginine methylated proteins based on their subcellular localizations using the GO terms designated on TriTrypDB. As seen in Fig. 3A, we were able to predict the localizations of only 24% of arginine methylated proteins due to the large number of hypothetical proteins. Of those proteins with known or predicted subcellular localizations, the highest percentage was nuclear, representing 25% of classifiable proteins (41 proteins total) (Fig. 3B). Other compartments that were well represented in the arginine methylome were integral membrane proteins, cytoskeleton associated proteins, and cytoplasmic proteins, which comprised 18%, 14%, and 12% of the classifiable proteins, respectively. In total, we identified 13 different cellular compartments predicted to contain arginine methylated proteins, and we previously reported over 150 arginine methylproteins localized to mitochondria [15]. Thus, protein arginine methylation is widespread in PF *T. brucei*.

The diverse localizations of arginine methylproteins suggest that this modification is likely to affect a wide range of cellular functions. To define the functional classes of arginine methylproteins, we again mined the *T. brucei* annotated genome database for all identified methylated proteins. To provide a standard classification, we first sorted proteins into 13 different categories based on their combined biological process and molecular function GO terms. Similar to the localization GO terms, we were able to predict functional classes for only 34% of the identified arginine methylated proteins due to the large number of hypothetical proteins (Fig. 4A). Of the annotated proteins, the most highly represented class of proteins is those associated with the cytoskeleton and involved in locomotion (Fig. 4B). There are 37 proteins within this category; however, this is primarily due to the presence of kinesins and dyneins, of which there are 14 and 9, respectively. While several proteins involved in flagellar function are arginine methylated in *Chlamydomonas* [31], this is the first study to identify arginine methylation of dyneins and kinesins. In light of the reported functions of flagella in *T. brucei*, arginine methylation of flagellar proteins has the potential to impact pathogenesis through effects on host cell attachment, motility, morphogenesis, cell division and immune evasion [32,33].

3.5. *Arginine methylation and RNA processing*

Consistent with the fact that many known PRMT substrates in other organisms are RNA binding proteins [10,13,14,34], proteins involved in RNA metabolism and translation constituted a large fraction of classifiable proteins in the *T. brucei* arginine methylome (Fig. 4B). Because of the prominent roles of RNA processing, turnover, and translation in trypanosomatid gene regulation along with the abundance of hypothetical proteins in our data set, we further investigated the potential impact of arginine methylation on proteins known or predicted to be involved in RNA biology by manual inspection of the arginine methylome. We identified a total of 65 proteins that either bear motifs indicative of RNA binding are homologues of proteins that function in RNA metabolism or translation in other organisms have been reported to have such functions in *T. brucei*. As shown in Table 1, this includes 10 putative RNA helicases, a class of proteins that functions at every step of RNA metabolism in higher eukaryotes [35]. Two of these, TbDED1-1 and TbDED1-2, were recently shown to have partially redundant functions in translation [36], while MTR4 reportedly functions in rRNA processing [37]. Also identified as containing methylarginine were 10 RRM containing proteins. One of these, U2AF65, is an mRNA splicing and stability factor [38,39], and p34/37 are homologous proteins involved in ribosome biogenesis [40–42]. Among non-RRM proteins, we identified DCL2, a Dicer-like protein that is a component of the nuclear RNAi pathway [43], the Not1 component of the major deadenylase complex [44,45], and the ALBA3/4 proteins that appear to regulate parasite development through modulation of mRNA stability and/or translation [29,30]. Also present in the arginine methylome were additional proteins with known or predicted roles in RNAi, pre-mRNA splicing, mRNA 3′ end formation, RNA

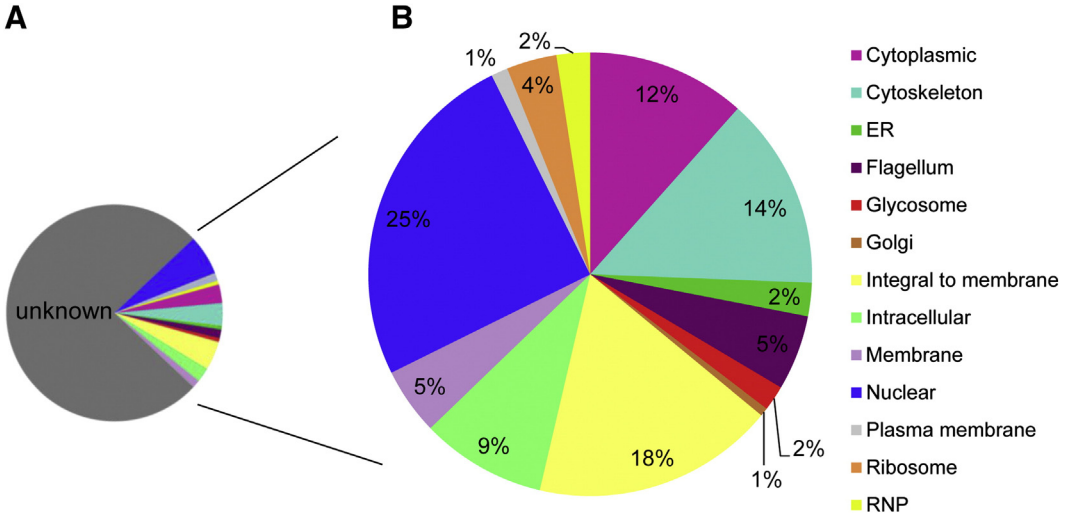

Fig. 3 – Arginine methylated proteins classified according to subcellular localization. (A) Predicted subcellular localization of arginine methylated proteins identified in this study according to their GO localization terms listed on TriTrypDB. (B) Classification of the subset of arginine methylated proteins with known or predicted localizations.

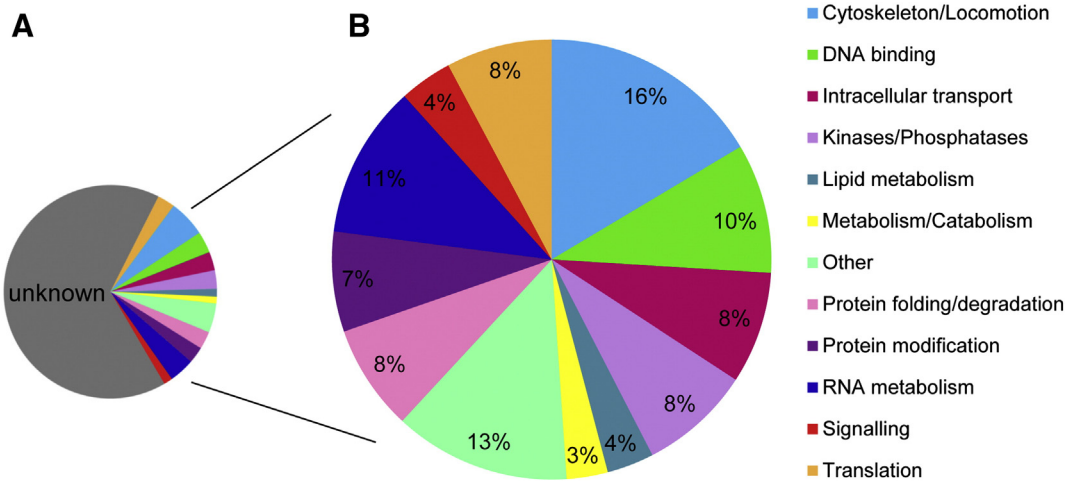

Fig. 4 – Arginine methylated proteins classified according to biological function. (A) Predicted biological functions of arginine methylated proteins identified in this study according to their GO terms listed on TriTrypDB. (B) Classification of the subset of arginine methylated proteins with known or predicted functions.

decay, translation, ribosome biogenesis, and tRNA modification. Finally, several ribosomal proteins were also identified as being substrates for methylation, which in higher eukaryotes affects ribosome assembly [46,47]. Collectively, these data indicate that arginine methylation has the potential to impact essentially all steps of posttranscriptional gene regulation in trypanosomes.

For those proteins containing RRMs, interactions with RNA typically occur through aromatic residues within the RNP1 and RNP2 domains located on conserved β-sheets [48]. Having identified 10 RRM containing arginine methylated proteins, we next asked whether the positions of methylarginine relative to conserved domains could provide insight into functions of methylarginine in this evolutionarily conserved class of RNA binding proteins. We thus mapped the positions of methylated arginines detected by mass spectrometry relative to bioinformatically defined RRM domains (Superfamily SSF54928 in InterProScan). As shown in Fig. 5, methylarginines are typically located outside of RRM domains, often near the N and C termini of the proteins or centrally positioned between two RRM domains. The exception to this generalization is Tb927.10.7030, in which methylarginines are located within RRM domains, thereby potentially impacting protein-RNA binding. The positioning of methylarginine within the majority of these proteins suggests that methylation does not frequently impact protein-RNA interactions but more likely affects protein–protein interactions or subcellular localization, as reported for several methylarginine containing RNA binding proteins in higher eukaryotes [10].

In a few cases, we identified methylarginine in proteins whose homologues are arginine methylated in higher eukaryotes. For example, the Vasa family of RNA helicases, of which *T. brucei* arginine methylated proteins TbDed1-1 and TbDed1-2 are members, are symmetrically and asymmetrically dimethylated in *Drosophila*, *Xenopus*, and mouse, although the functional significance of these modifications is unclear [49]. The ribosomal protein Rps2 is arginine methylated in yeast and humans, albeit by different PRMTs and possibly with different impacts in the two systems [50–52]. The Mtr4 RNA helicase was recently shown to interact with the major yeast Type I PRMT in vivo, suggesting it might be methylated like its *T. brucei* homologue [53]. These findings suggest that some functions of arginine methylation exhibit extraordinary evolutionary conservation. In contrast, certain mechanistic roles of arginine methylation in RNA biology that are well understood in higher eukaryotes are clearly absent in trypanosomes. Most strikingly, symmetric dimethylation of arginines in the RG-rich tails of Sm proteins in Sm proteins B/B′, D1, and D3 and subsequent binding of methylarginines by the Tudor domain containing protein, SMN, is essential for spliceosome assembly in mammals [54]. Trypanosome Sm proteins lack RG-rich tails and were not identified in our methylome. Consistent with that finding, *T. brucei* SMN lacks the Tudor domain that recognizes SDMA [55]. However, the absence of Sm protein methylation does not imply that *trans* splicing, or the relatively rare *cis* splicing, in trypanosomes is unaffected by protein arginine methylation. We identified several splicing factors in our methylome including U2AF65, SF3B subunit 1, and Gemin2. In particular, Gemin2 is a key regulator of snRNP assembly in trypanosomes [56]. Thus, arginine methylation may dramatically impact spliceosome assembly in trypanosomes as it does in mammals, albeit through a distinct mechanism.

3.6. *Arginine methylation and DNA associated processes*

Another biological process GO term that constituted a large percentage of classifiable *T. brucei* arginine methylated proteins was DNA binding proteins (Fig. 4B). Based on GO terms and annotations, proteins involved in transcription and DNA repair were the most common, with 8 and 7 identified as containing methylarginine, respectively. Interestingly, the mammalian enzymes PRMT1, PRMT6, and PRMT7 have all been implicated in DNA repair responses [10]. The most well understood mechanism is the arginine methylation-dependent recruitment

Table 1 – Methylated proteins involved in RNA metabolism.

TriTryp DB number	Name/domains/functions	Type	# of methylargs
Tb09.211.3510	DED1-2	Helicase	4
Tb09.211.4430	ATP-dependent DEXDc RNA helicase, putative	Helicase	2
Tb927.10.14550	DED1-1	Helicase	3
Tb927.10.9780	ATP-dependent DEAD/H RNA helicase, putative	Helicase	1
Tb927.10.7740	MTR4	Helicase	1
Tb927.10.7280	Pre-mRNA splicing factor ATP-dependent RNA helicase	Helicase	2
Tb11.01.3930	RNA helicase, putative	Helicase	1
Tb927.5.4420	Nucleolar RNA helicase II (RNA helicase Gu), putative	Helicase	4
Tb927.6.740	ATP-dependent DEAH-box RNA helicase, putative	Helicase	2
Tb927.8.3020	RNA helicase, DEAD-box type, Q motif	Helicase	3
Tb927.10.7030		RRM domain	1
Tb927.10.3500	RBSR4/U2AF65	RRM domain	3
Tb11.01.5680		RRM domain	3
Tb11.02.4110		RRM domain	1
Tb11.02.5210		RRM domain	2
Tb927.3.1030	RPB28	RRM domain	1
Tb927.3.3670	TRRM3	RRM domain	1
Tb927.8.710	DRBD17	RRM domain	3
Tb927.8.990	RBP33	RRM domain	1
Tb11.01.5570/5590	RNA-binding protein (p34/37)	RRM domain	1
Tb09.211.4670	HAD-like, SM-domain	ZC3H ZNF	1
Tb927.10.12330	ZC3H34	ZC3H ZNF	1
Tb927.3.5250	ZC3H8	ZC3H ZNF	2
Tb11.01.1270	Nucleic acid binding protein, putative	CCHC ZnF	1
Tb11.01.2580	Puf11	Puf domain	1
Tb11.02.4570	Puf10	Puf domain	2
Tb09.160.3730	Glutaminyl-tRNA synthetase, putative	RNA modifying	2
Tb09.211.4350	Ribonuclease, PARN3	RNA modifying	2
Tb927.10.14750	Fibrillarin, putative	RNA modifying	12
Tb927.10.8360	Ribonuclease, PARN2	RNA modifying	2
Tb927.10.7500	Fibrillarin (NOP1)	RNA modifying	4
Tb10.v4.0040	mRNA capping methyltransferase, putative	RNA modifying	2
Tb11.01.5780	Tudor SN, RNA interference	RNA modifying	1
Tb11.46.0008	Arginyl-tRNA synthetase, putative	RNA modifying	4
Tb11.55.0013	Cysteine desulfurase, putative	RNA modifying	1
Tb927.3.1230	DCL2, dicer-like protein	RNA modifying	3
Tb927.3.3160	Poly(A) polymerase (PAP)	RNA modifying	2
Tb927.8.7570	tRNA-dihydrouridine synthase 1, putative	RNA modifying	3
Tb927.10.15210	CBP30	Other	2
Tb927.10.9020	Eukaryotic initiation factor 3, gamma subunit	Other	4
Tb927.10.5640	Splicing, Gemin 2	Other	2
Tb927.10.4570	Elongation factor 2	Other	2
Tb927.10.1510	NOT1	Other	2
Tb11.01.3420	Eukaryotic translation initiation factor, putative	Other	1
Tb11.01.3690	Splicing factor 3B subunit 1, putative	Other	1
Tb11.01.5535	U6 snRNA-associated Sm-like protein LSm4p	Other	2
Tb11.02.4950	Predicted translation machinery-associated	Other	1
Tb927.2.3620	GTP-binding elongation factor Tu family	Other	2
Tb927.2.4950	Zinc-binding domain of translation initiation factor 2 beta	Other	4
Tb927.2.5240	PRP19, pre-mRNA splicing	Other	2
Tb927.4.2030	TbALBA4	Other	7
Tb927.4.2040	TbALBA3	Other	1
Tb927.4.470	snoRNP protein gar1, putative	Other	16
Tb927.5.2570	Translation initiation factor, putative|(EIF3B)	Other	2
Tb927.7.1620	Cleavage and polyadenylation specificity factor subunit 5	Other	1
Tb927.7.270	Ribosome biogenesis protein, putative	Other	2
Tb927.7.3280	Translation initiation factor IF-2, putative	Other	1
Tb927.7.4220	Dip2/Utp12 Family, rRNA processing	Other	2
Tb927.7.970	NMD3, rRNA processing	Other	1
Tb927.8.6880	Translation initiation factor IF-2, putative	Other	3
Tb927.8.750	Nucleolar RNA-binding protein, putative	Other	2
Tb927.8.760	Nopp44/46, nucleolar RNA-binding protein	Other	10
Tb927.10.190	40S ribosomal protein S6, putative	Ribosomal	1
Tb927.10.14710	40S ribosomal protein S2, putative	Ribosomal	1
Tb11.02.4350	40S ribosomal protein S21, putative	Ribosomal	1

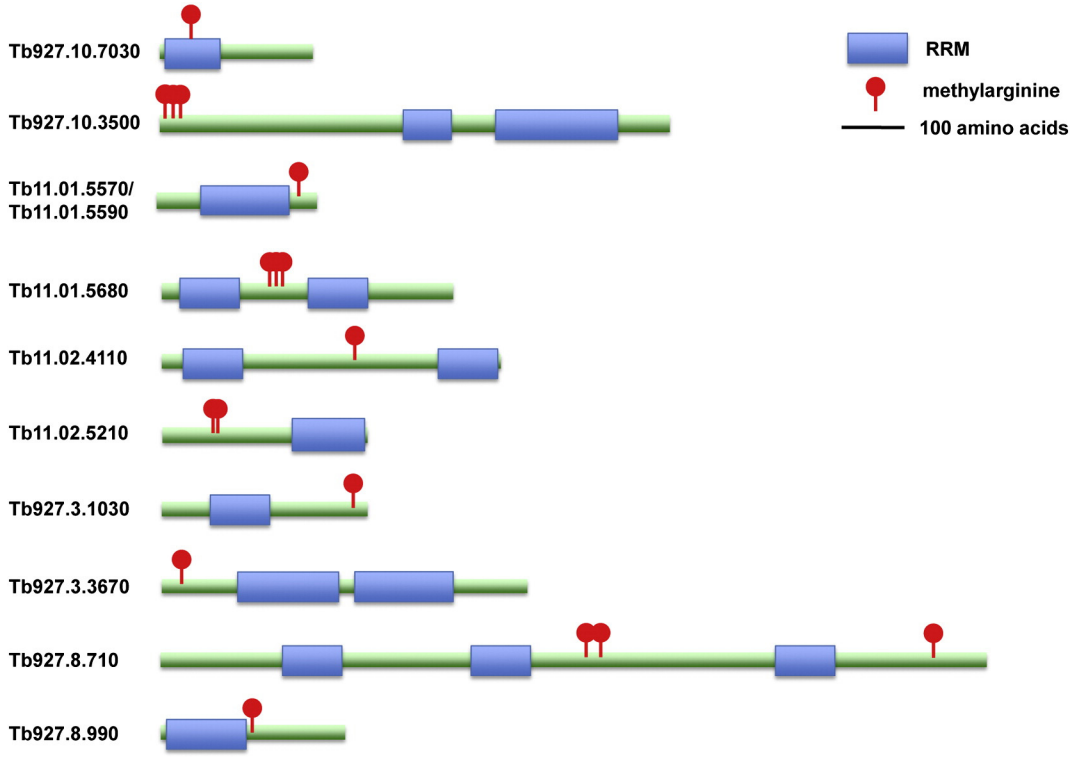

Fig. 5 – Relative positions of methylarginines and RNA recognition motifs (RRMs). Schematic representation of the ten methylarginine-containing proteins that harbor RRM domains. Blue boxes represent RRM domains as defined by InterProScan SSF54928. Red circles indicate positions of methylarginines.

of the MRN complex to double stranded breaks through modification of an RG-rich region in its MRE11 subunit [57,58]. Additionally, PRMT1 and PRMT6 play distinct roles in modulation of DNA polymerase β action in base excision repair [59,60]. In light of these reports, we wanted to further assess the potential role of arginine methylation in DNA replication and repair in *T. brucei* by manually curating our list to identify arginine methylated proteins likely to be involved in these processes based on their homology to proteins with such functions in other organisms (Table 2). With regard to DNA replication, we identified two subunits of the essential eukaryotic replicative DNA helicase, TbMCM6 and TbMCM7 [61,62]. Methylarginine was also identified in TbMCM8, which has been linked to types of homologous recombination in other systems [61,63]. Notably, TbMCM6-8 were also recently identified in complex with the TbMCM-BP protein required for repression of life cycle specific, RNA polymerase I transcribed genes [64]. Regarding DNA repair, we identified the DNA excision repair protein, TbSnf2, as an arginine methylated

Table 2 – Methylated proteins with a potential role in DNA replication and repair.

TriTryp DB number	Name/domains/functions	# of methylargs
Tb09.211.3310	Replication factor C, subunit 3, putative	2
Tb927.6.3890	Replication factor C, subunit 2, putative	2
Tb927.10.10410	Minichromosome maintenance (MCM) complex subunit 8, putative	1
Tb11.01.3510	Minichromosome maintenance (MCM) complex subunit 6, putative	2
Tb11.01.7810	Minichromosome maintenance (MCM) complex subunit 7, putative	1
Tb11.01.3390	DNA topoisomerase II, putative	2
Tb09.211.3040	Rad-4 like DNA-repair protein, putative	1
Tb11.01.0340	RAD50 DNA repair-like protein	1
Tb11.01.6300	Phosphatidylinositol 3-related kinase TbATR, putative	1
Tb11.02.3400	SNF2 DNA repair protein, putative	3
Tb927.2.2260	Phosphatidylinositol kinase domain protein TbATM, putative	6
Tb927.3.3090	RTEL1-like DNA repair helicase, putative	1
Tb927.7.4080	SNF2 family helicase-like DNA excision repair protein, putative	4
Tb927.8.5510	Apurinic/apyrimidinic endonuclease, putative	1
Tb927.8.5710	Recombination initiation protein NBS1, putative	2

protein. In yeast, Snf2 is associated in vivo with and methylated in vitro by the major Type I PRMT, suggesting that modulation of Snf2 function by arginine methylation may be evolutionarily conserved [53]. We detected methylarginine residues in the *T. brucei* homologues of the PI3K-like kinases, ATM and ATR, which are master regulators of the DNA damage response and repair pathways in other eukaryotes [65]. In addition, the DNA mismatch repair protein, PMS1, harbors ADMA in its highly conserved ATPase domain, suggesting a role for methylation in the global mismatch repair pathway. Regarding the MRN complex described above, the trypanosome MRE11 protein lacks the RG repeats whose methylation is critical in human double stranded break repair, and MRE11 was not identified in our methylome. However, we did detect arginine methylation of the other two MRN complex subunits, RAD50 and NBS1 (Table 2 and Table S1), which have not been reported as methylated in other systems. Thus, similar to pre-mRNA splicing, our data suggest that trypanosomes use arginine methylation to modulate similar pathways as in higher eukaryotes, although the specific proteins involved differ.

3.7. Arginine methylation and intracellular protein trafficking

Also identified as substrates for arginine methylation were proteins involved in diverse aspects of intracellular trafficking as depicted in Fig. 6. This includes TbRab1 (Tb927.8.890) and TbRab6 (Tb927.2.2130). TbRab1 is involved in transport between ER and Golgi [66], while the putative TbRab6 is thought to be involved in intra-Golgi transport [67]. Homologues of three different vacuolar sorting proteins (VPS), VPS45 (Tb927.10.6780), VPS33 (Tb927.3.2280), and VPS51 (Tb927.10.2130) also contain methylarginine. While these proteins have yet to be studied in *T. brucei*, higher eukaryote homologues of VPS45 and VPS33 function in vesicle transport to the endosome and vacuole, respectively, while VPS51 facilitates the tethering of vesicles from endosomes to the trans Golgi network during retrograde transport [68–71]. Tb09.211.4240 is a member of the sorting nexin family, most similar to human sorting nexin 2 (SNX2). A different member of this family, sorting nexin 3, was reported to contain methylarginine in a recent study of the human T cell methylome [14], and human sorting nexin 17 was identified in anti-ADMA immunoprecipitates from HeLa cells [13]. Also arginine methylated in *T. brucei* is the delta subunit of the adaptor protein-3 (AP-3) complex (Tb927.5.3610), which is essential for acidocalicisome biogenesis, growth, and virulence in *T. brucei* [72]. We identified two sites of ADMA in the putative Sec10 protein (Tb927.8.2580) which, in higher eukaryotes, functions as a part of the larger exocyst complex that tethers secretory vesicles at the plasma membrane [73]. Finally, we identified methylarginine on two proteins that function at all stages of transport. RabGAP (Tb11.02.3320) enhances Rab's inherent GTPase activity to produce the inactive, RabGDP form of the protein. The putative N-ethylmaleimide sensitive factor (NSF; Tb927.1.1560), which dissociates SNARE complexes for multiple rounds of vesicle fusion, was shown to contain two sites of methylation; one of which could be resolved as ADMA. It is striking that protein trafficking was discovered as a pathway in which arginine methylation plays an important role in both this study (9 proteins) and a recent arginine methylome analysis of human T cells (6 proteins) [14]. Again, while trypanosomes and humans share an impact of arginine methylation on a common pathway, different players were identified in the two studies. The potential impacts of arginine methylation on protein trafficking revealed by these data have implications for numerous aspects of *T. brucei* biology and pathogenesis including variant surface glycoprotein-mediated immune evasion [74].

3.8. Amino acid sequences surrounding methylarginines

The results presented here constitute by far the most extensive, unbiased data set gathered to date regarding sites of arginine methylation in any organism. Thus, we analyzed our data to gain insight into the modes of PRMT recognition by querying the amino acid contexts in which methylation of arginine residues occurs. In many cases we were able to differentiate between ADMA and SDMA, allowing us to address whether the sites surrounding these two methylation products differ from each other and from MMA-containing sites. Arginine methylation is frequently reported to occur within the contexts of RG repeats and RXR motifs, but there are some key examples where this dogma does not hold true [10,75]. In this study, we identified a total of 1332 methylarginines, of which 649 were MMA and 683 were DMA. Of the identified DMA residues, 165 could be resolved as ADMA and 75 as SDMA. To determine the amino acid composition around specific classes of methylarginines, we analyzed the six N-terminal and C-terminal residues of our MMA, ADMA, and SDMA data sets using WebLogo (Fig. 7). We note that the MMA analysis is somewhat confounded by the fact that MMA can either represent an intermediate product of DMA synthesis or a final PRMT7-catalyzed product. Fig. 7 clearly shows that glycine is the preferred amino acid C-terminal to all classes of methylarginine as well as being highly represented both N- and C-terminal of MMA, ADMA, and SDMA. Nevertheless, differences in the motifs surrounding each type of methyarginine were observed. Most strikingly, ADMA displayed the strongest preference for a C-terminal glycine, while regions surrounding SDMA were less enriched in glycine than were sites surrounding other classes of methylarginine. Regions N-terminal to both SDMA and ADMA displayed a preference for small hydrophobic amino acids such as leucine, alanine, and valine. However, regions C-terminal to SDMA sites differed from those C-terminal to ADMA sites in that charged polar amino acids such as arginine and glutamic acid were preferred over neutral polar amino acids downstream of the methylated arginine. Therefore, although DMA is most often followed by a glycine, SDMA typically exhibits fewer glycine residues surrounding it and displays a stronger preference for both upstream and downstream glutamic acid residues than does ADMA. Regarding the amino acids surrounding MMA, we found that glycine is slightly less prevalent at the +1 position compared to ADMA and SDMA. Additionally, regions N-terminal to MMA exhibited the strongest overall preference for glutamic acid. The distinct pattern of amino acids surrounding MMA residues suggests that a substantial proportion of these modified amino acids are terminal products of PRMT7 rather than intermediates in ADMA or SDMA synthesis.

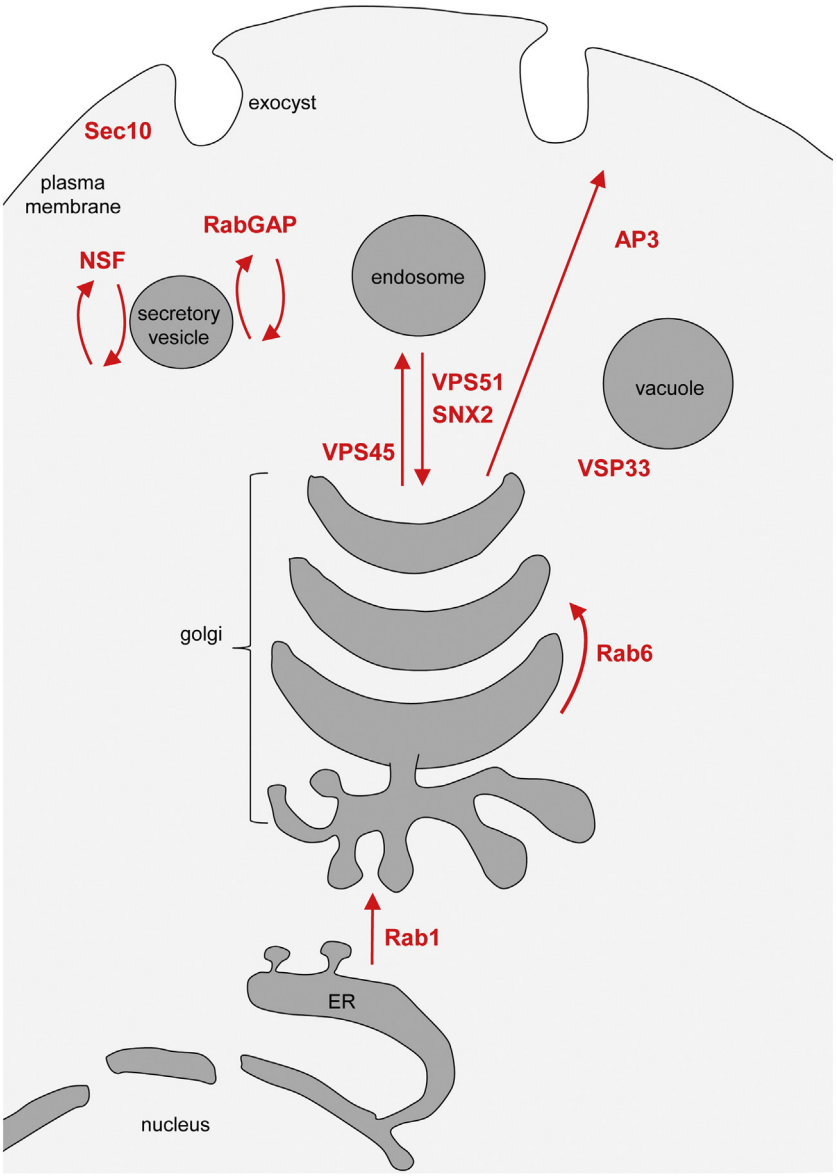

Fig. 6 – Schematic diagram of identified arginine methylated proteins involved in intracellular trafficking. Organelles shown in dark gray; methylated proteins shown in red. Opposing arrows denote function of NSF and RabGAP in vesicle recycling. See text for details.

We recently described amino acid motifs surrounding ADMA and SDMA in mitochondrial proteins [15], and next wondered whether motifs surrounding a given type of DMA differ between the mitochondrial and non-mitochondrial sets of proteins. Examination of the two populations showed that the regions surrounding ADMA are quite similar. The major exception is that the non-mitochondrial proteins display a much stronger preference for serine downstream of the ADMA compared to the mitochondrial ADMA-containing proteins, whose downstream regions are relatively devoid of serines and enriched in charged amino acids. With regard to SDMA-containing proteins, we observed two qualitative differences between the mitochondrial and non-mitochondrial protein sets. First, the mitochondrial proteins are more highly enriched in arginines upstream of the SDMA residue compared to the non-mitochondrial SDMA-containing proteins, while the non-mitochondrial proteins contain more small amino acids such as alanine, leucine, and valine. Second, the non-mitochondrial SDMA-containing proteins have a higher frequency of glycine at the +1 position compared to the mitochondrial SDMA-containing proteins. Differences between the motifs surrounding mitochondrial and non-mitochondrial DMA residues raise the interesting question as to the identities of mitochondrial PRMTs. One possibility is that the same PRMTs are active in mitochondrial and non-mitochondrial compartments, but that distinct PRMT-associated proteins differentially modulate substrate specificity. We also cannot rule out that the PRMTs have somewhat flexible substrate requirements, and it is the substrate pools that differ in different compartments. Additionally, we have so far not detected any of characterized *T. brucei* PRMTs in mitochondria

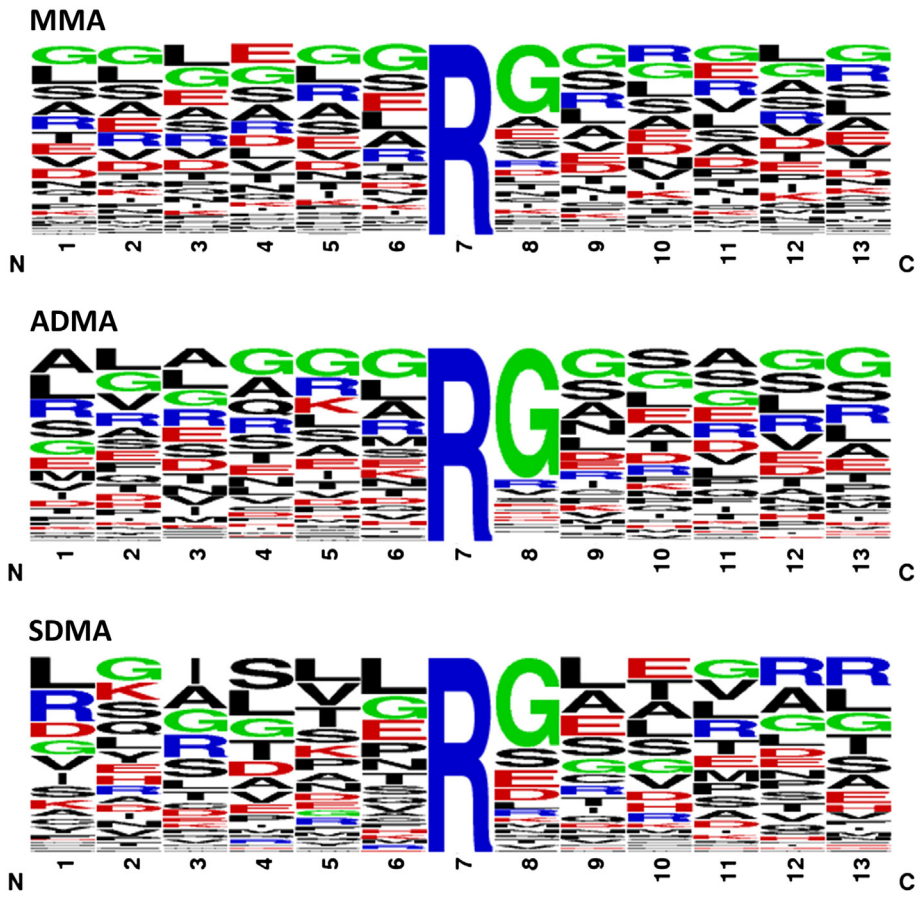

Fig. 7 – Amino acid motifs surrounding MMA, ADMA and SDMA. Weblogo3 (http://weblogo.threeplusone.com/) was used to analyze the six amino acids N-terminal and C-terminal of MMA, ADMA and SDMA. MMA (top), ADMA (middle) and SDMA (bottom) had an n of 651, 165 and 75, respectively. Arginines are shown in blue, glycines are in green, and charged amino acids other than arginine are labeled in red.

by western blot, suggesting that mitochondria might harbor novel PRMTs.

3.9. *Co-occurrence of arginine methylation and phosphorylation*

Different types of posttranslational modifications frequently influence one another and can act either synergistically or antagonistically. Arginine methylation reportedly affects lysine methylation, acetylation, and phosphorylation [10,76]. For example, antagonism between arginine methylation and nearby serine phosphorylation of yeast Npl3 regulates the protein's nucleocytoplasmic transport [77], and a similar antagonism between arginine methylation and serine phosphorylation on the RNA Polymerase II C-terminal domain impacts transcription of specific gene classes in mammals [78]. While there have been no genome-wide studies addressing lysine methylation or acetylation in *T. brucei*, a proteomic study by Nett, et al. [9] identified phosphorylation sites in bloodstream form *T. brucei*. We analyzed these data to determine whether arginine methylated proteins identified here have been reported to harbor a phosphorylation modification as well. Of the 676 methylarginine containing proteins, we found that 121 (18%) reportedly contain phosphoamino acids. Reasoning that posttranslational modifications in close proximity would have a higher probability of influencing one another, we next determined if the reported phosphorylated residue was within 10 amino acids of a methylarginine identified in the present study. Using these criteria, we identified 13 proteins with a phosphorylated residue in proximity to methylarginine. For example, the putative kinase Tb11.01.5880 harbors six phosphoserines, all of which lie outside of its predicted catalytic domain. We identified three methylarginines surrounding two of these phosphoserines, two lying one and six residues upstream and one lying six residues downstream of the phosphoserines. Thus, it is possible that these five residues may be regulating one another, similar to arginine methylation-phosphorylation cross talk reported in other systems [76–78]. It is important to note that although posttranslational modifications may not be near each other in primary sequence, they may be in close vicinity within the tertiary structure of the protein. Therefore, these numbers are likely a vast underestimate of proteins containing methylarginine in close proximity to other modified residues. It will be of great interest to determine the biological effects of cross talk between arginine methylation and other posttranslational modifications as new data sets profiling other modifications become available.

3.10. *Evolutionary conservation of arginine methylation sites*

T. brucei is an early branching eukaryote of the order Kinetoplastida, which includes additional human pathogens such as *Trypanosoma cruzi* and *Leishmania* spp. We next asked whether specific sites of arginine methylation is select proteins are conserved between kinetoplastid species, thereby possibly suggesting a conserved function for specific modified arginines in the biology of these organisms. To this end, we aligned a subset of *T. brucei* arginine methylated proteins with their homologues from *T. cruzi* and *Leishmania major*. A random pool of arginine methylated proteins encoded on chromosome 9 and accounting for over 10% of the total methylarginine-containing proteins was analyzed. Of the 97 methylarginine sites in 79 proteins that were examined, 21 sites in 19 proteins were conserved in both *T. cruzi* and *L. major*. After accounting for 9 of these proteins that had no homologue in *L. major*, these numbers translate to 23.4% of methylarginines being conserved between all three kinetoplastid species. We next performed additional comparison between the two *Trypanosoma* species. Here, we found that 53 of the 97 methylarginine sites (54.6%) identified in *T. brucei* harbored an arginine residue at the conserved position in *T. cruzi*. Reasoning that methylarginines might be more conserved in certain structural contexts than others, we next used the Jpred 3 secondary structure prediction program to determine the spatial context of the methylated arginine residues in our subset of proteins. Of the 97 sites analyzed, 55 (56.7%) methylarginines were predicted to be in an unstructured region, 37 (38.1%) were predicted to reside in an alpha helix, while only 5 (5.2%) were predicted to be in a beta sheet. When we correlated predicted structure with the conservation of methylarginines between *T. brucei* and *T. cruzi*, we found that the extent of conservation was similar in all structural contexts. Collectively, these analyses show that methylarginines are somewhat conserved between kinetoplastid species and suggest that approximately one-quarter of *L. major* and half of *T. cruzi* arginine methylated proteins may be subject to similar modes of methylation-dependent regulation. In addition, beta sheets typically exhibit low levels of arginine methylation, while alpha helices and unstructured regions of proteins both serve as good PRMT substrates.

4. Conclusion

Using a sensitive, proteome-wide mass spectrometry approach, we identified 676 arginine methylated proteins in PF *T. brucei*, greatly expanding the set of known methylarginine-containing proteins in biology. Proteins harboring methylarginine account for approximately 10% of the *T. brucei* proteome, arise from numerous subcellular compartments, and function in diverse cellular pathways. Hence, our data indicate that arginine methylation arose very early in evolution as a widespread mechanism for modulating protein function. We identified only a handful of common arginine methylated proteins between trypanosomes and higher eukaryotes, in part due to the highly diverged nature of trypanosomes as well as the limited datasets available. However, with regard to specific functions such as spliceosome assembly and double-stranded DNA break repair, our data suggest that homologous pathways may be impacted by arginine methylation, although the target proteins that undergo methylation may differ between evolutionarily divergent organisms.

We identified protein trafficking as one pathway in which methylarg proteins function at numerous steps, reminiscent of recent findings in human T cells [14]. Interestingly, this class of arginine methylated proteins was not enriched in previous arginine methylome studies that entailed immunoprecipitation with anti-methylarginine antibodies, possibly due to the specificity of available antibodies for methylarginines in glycine-rich contexts [13]. Overwhelmingly, the motifs surrounding methylarginines in trafficking proteins were non-canonical, with less than 10% of the methylarginines preceeded by a glycine and less than 30% followed by one. This suggests that continued use of unbiased approaches such as the one used here is likely to reveal new functions for arginine methylation. We also identified several methylarginine-containing proteins associated with DNA transcription, replication, and repair. However, histones, whose methylation on arginine residues is an important aspect of the histone code in higher eukaryotes, were not identified in our study, likely due to the method of sample preparation. Previous analysis of histone modifications in trypanosomes also failed to identify methylarginine, although technical limitations may have been at play here as well [79,80]. Thus, targeted studies aimed at detecting methylarginine in *T. brucei* histones are underway in our laboratory. A substantial percentage of arginine methylated proteins identified here function in RNA biology. Given the pivotal role of posttranscriptional gene regulation in kinetoplastids, our data suggest that arginine methylation affects gene expression at almost every step. We showed that proteins that undergo arginine methylation differ between life cycle stages of *T. brucei*, and it is likely that arginine methylation will impact differentiation in trypanosomes as it does in mammals [81]. An important future question regards the proteins that read arginine methylmarks in trypanosomes. In humans, Tudor domain proteins are the sole known methylarginine readers [81]. However, there appears to be a paucity of Tudor domain proteins in *T. brucei*, suggesting novel proteins that read arginine methyl marks await discovery in this system. Finally, we expect that studies illuminating the molecular functions of specific arginine methylmarks in *T. brucei* biology will be rapidly forthcoming due to the large number of genetic techniques applicable to this organism.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jprot.2013.07.010.

Acknowledgments

We thank Drs. Michelle Ammerman and Sara Zimmer for critical reading of the manuscript and Shreya Mukhopadhyay for technical assistance. This work was supported by NIH grant RO1AI060260 to LKR, NIH grant U54HD071594 and AHA award 12SDG9450036 to JQ, and NIH postdoctoral fellowship F32AI100350 to KL.

REFERENCES

[1] Matthews KR. The developmental cell biology of *Trypanosoma brucei*. J Cell Sci 2005;118(Pt 2):283–90.
[2] Kramer S. Developmental regulation of gene expression in the absence of transcriptional control: the case of kinetoplastids. Mol Biochem Parasitol 2012;181(2):61–72.
[3] Clayton C, Shapira M. Post-transcriptional regulation of gene expression in trypanosomes and leishmanias. Mol Biochem Parasitol 2007;156(2):93–101.
[4] Gassen A, Brechtefeld D, Schandry N, Arteaga-Salas JM, Israel L, Imhof A, et al. DOT1A-dependent H3K76 methylation is required for replication regulation in *Trypanosoma brucei*. Nucleic Acids Res 2012;40(20):10302–11.
[5] Kawahara T, Siegel TN, Ingram AK, Alsford S, Cross GA, Horn D. Two essential MYST-family proteins display distinct roles in histone H4K10 acetylation and telomeric silencing in trypanosomes. Mol Microbiol 2008;69(4):1054–68.
[6] Figueiredo LM, Janzen CJ, Cross GA. A histone methyltransferase modulates antigenic variation in African trypanosomes. PLoS Biol 2008;6(7):e161.
[7] Szoor B, Ruberto I, Burchmore R, Matthews KR. A novel phosphatase cascade regulates differentiation in *Trypanosoma brucei* via a glycosomal signaling pathway. Genes Dev 2010;24(12):1306–16.
[8] Ma J, Benz C, Grimaldi R, Stockdale C, Wyatt P, Frearson J, et al. Nuclear DBF-2-related kinases are essential regulators of cytokinesis in bloodstream stage *Trypanosoma brucei*. J Biol Chem 2010;285(20):15356–68.
[9] Nett IR, Martin DM, Miranda-Saavedra D, Lamont D, Barber JD, Mehlert A, et al. The phosphoproteome of bloodstream form *Trypanosoma brucei*, causative agent of African sleeping sickness. Mol Cell Proteomics 2009;8(7):1527–38.
[10] Bedford MT, Clarke SG. Protein arginine methylation in mammals: who, what, and why. Mol Cell 2009;33(1):1–13.
[11] Bedford MT. Arginine methylation at a glance. J Cell Sci 2007;120(Pt 24):4243–6.
[12] Horowitz S, Trievel RC. Carbon-oxygen hydrogen bonding in biological structure and function. J Biol Chem 2012;287(50):41576–82.
[13] Boisvert FM, Cote J, Boulanger MC, Richard S. A proteomic analysis of arginine-methylated protein complexes. Mol Cell Proteomics 2003;2(12):1319–30.
[14] Uhlmann T, Geoghegan VL, Thomas B, Ridlova G, Trudgian DC, Acuto O. A method for large-scale identification of protein arginine methylation. Mol Cell Proteomics 2012;11(11):1489–99.
[15] Fisk JC, Li J, Wang H, Aletta JM, Qu J, Read LK. Proteomic analysis reveals diverse classes of arginine methylproteins in mitochondria of trypanosomes. Mol Cell Proteomics 2012;12(2):302–11.
[16] Pelletier M, Pasternack DA, Read LK. In vitro and in vivo analysis of the major type I protein arginine methyltransferase from *Trypanosoma brucei*. Mol Biochem Parasitol 2005;144(2):206–17.
[17] Pasternack DA, Sayegh J, Clarke S, Read LK. Evolutionarily divergent type II protein arginine methyltransferase in *Trypanosoma brucei*. Eukaryot Cell 2007;6(9):1665–81.
[18] Fisk JC, Zurita-Lopez C, Sayegh J, Tomasello DL, Clarke SG, Read LK. TbPRMT6 is a type I protein arginine methyltransferase that contributes to cytokinesis in *Trypanosoma brucei*. Eukaryot Cell 2010;9(6):866–77.
[19] Fisk JC, Sayegh J, Zurita-Lopez C, Menon S, Presnyak V, Clarke SG, et al. A type III protein arginine methyltransferase from the protozoan parasite *Trypanosoma brucei*. J Biol Chem 2009;284(17):11590–600.
[20] Fisk JC, Read LK. Protein arginine methylation in parasitic protozoa. Eukaryot Cell 2011;10(8):1013–22.
[21] Goulah CC, Pelletier M, Read LK. Arginine methylation regulates mitochondrial gene expression in *Trypanosoma brucei* through multiple effector proteins. RNA 2006;12(8):1545–55.
[22] Goulah CC, Read LK. Differential effects of arginine methylation on RBP16 mRNA binding, guide RNA (gRNA) binding, and gRNA-containing ribonucleoprotein complex (gRNP) formation. J Biol Chem 2007;282(10):7181–90.
[23] Duan P, Xu Y, Birkaya B, Myers J, Pelletier M, Read LK, et al. Generation of polyclonal antiserum for the detection of methylarginine proteins. J Immunol Methods 2007;320(1–2):132–42.
[24] Zimmer SL, McEvoy SM, Li J, Qu J, Read LK. A novel member of the RNase D exoribonuclease family functions in mitochondrial guide RNA metabolism in *Trypanosoma brucei*. J Biol Chem 2011;286(12):10329–40.
[25] Wirtz E, Leal S, Ochatt C, Cross GA. A tightly regulated inducible expression system for conditional gene knock-outs and dominant-negative genetics in *Trypanosoma brucei*. Mol Biochem Parasitol 1999;99(1):89–101.
[26] Hirumi H, Hirumi K. Continuous cultivation of *Trypanosoma brucei* blood stream forms in a medium containing a low concentration of serum protein without feeder cell layers. J Parasitol 1989;75(6):985–9.
[27] Vezzalini M, Aletta JM, Beghelli S, Moratti E, Della Peruta M, Mafficini A, et al. Immunohistochemical detection of arginine methylated proteins (MeRP) in archival tissues. Histopathology 2010;57(5):725–33.
[28] Berriman M, Ghedin E, Hertz-Fowler C, Blandin G, Renauld H, Bartholomeu DC, et al. The genome of the African trypanosome *Trypanosoma brucei*. Science 2005;309(5733):416–22.
[29] Mani J, Guttinger A, Schimanski B, Heller M, Acosta-Serrano A, Pescher P, et al. Alba-domain proteins of *Trypanosoma brucei* are cytoplasmic RNA-binding proteins that interact with the translation machinery. PLoS One 2011;6(7):e22463.
[30] Subota I, Rotureau B, Blisnick T, Ngwabyt S, Durand-Dubief M, Engstler M, et al. ALBA proteins are stage regulated during trypanosome development in the tsetse fly and participate in differentiation. Mol Biol Cell 2011;22(22):4205–19.
[31] Sloboda RD, Howard L. Protein methylation in full length *Chlamydomonas* flagella. Cell Motil Cytoskeleton 2009;66(8):650–60.
[32] Ralston KS, Hill KL. The flagellum of *Trypanosoma brucei*: new tricks from an old dog. Int J Parasitol 2008;38(8–9):869–84.
[33] Griffiths S, Portman N, Taylor PR, Gordon S, Ginger ML, Gull K. RNA interference mutant induction in vivo demonstrates the essential nature of trypanosome flagellar function during mammalian infection. Eukaryot Cell 2007;6(7):1248–50.
[34] Blackwell E, Ceman S. Arginine methylation of RNA-binding proteins regulates cell function and differentiation. Mol Reprod Dev 2012;79(3):163–75.
[35] Linder P, Jankowsky E. From unwinding to clamping - the DEAD box RNA helicase family. Nat Rev Mol Cell Biol 2011;12(8):505–16.
[36] Zinoviev A, Akum Y, Yahav T, Shapira M. Gene duplication in trypanosomatids - two DED1 paralogs are functionally redundant and differentially expressed during the life cycle. Mol Biochem Parasitol 2012;185(2):127–36.
[37] Cristodero M, Clayton CE. Trypanosome MTR4 is involved in rRNA processing. Nucleic Acids Res 2007;35(20):7023–30.
[38] Gunzl A. The pre-mRNA splicing machinery of trypanosomes: complex or simplified? Eukaryot Cell 2010;9(8):1159–70.

[39] Gupta SK, Carmi S, Waldman Ben-Asher H, Tkacz ID, Naboishchikov I, Michaeli S. Basal splicing factors regulate the stability of mature mRNAs in trypanosomes. J Biol Chem 2013;288(7):4991–5006.
[40] Prohaska K, Williams N. Assembly of the *Trypanosoma brucei* 60S ribosomal subunit nuclear export complex requires trypanosome-specific proteins P34 and P37. Eukaryot Cell 2009;8(1):77–87.
[41] Hellman KM, Ciganda M, Brown SV, Li J, Ruyechan W, Williams N. Two trypanosome-specific proteins are essential factors for 5S rRNA abundance and ribosomal assembly in *Trypanosoma brucei*. Eukaryot Cell 2007;6(10):1766–72.
[42] Ciganda M, Prohaska K, Hellman K, Williams N. A novel association between two trypanosome-specific factors and the conserved L5-5S rRNA complex. PLoS One 2012;7(7): e41398.
[43] Patrick KL, Shi H, Kolev NG, Ersfeld K, Tschudi C, Ullu E. 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 2009;106(42): 17933–8.
[44] Farber V, Erben E, Sharma S, Stoecklin G, Clayton C. Trypanosome CNOT10 is essential for the integrity of the NOT deadenylase complex and for degradation of many mRNAs. Nucleic Acids Res 2013;41(2):1211–22.
[45] Schwede A, Ellis L, Luther J, Carrington M, Stoecklin G, Clayton C. A role for Caf1 in mRNA deadenylation and decay in trypanosomes and human cells. Nucleic Acids Res 2008;36(10):3374–88.
[46] Ren J, Wang Y, Liang Y, Zhang Y, Bao S, Xu Z. Methylation of ribosomal protein S10 by protein-arginine methyltransferase 5 regulates ribosome biogenesis. J Biol Chem 2010;285(17): 12695–705.
[47] Shin HS, Jang CY, Kim HD, Kim TS, Kim S, Kim J. Arginine methylation of ribosomal protein S3 affects ribosome assembly. Biochem Biophys Res Commun 2009;385(2): 273–8.
[48] Maris C, Dominguez C, Allain FH. The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. FEBS J 2005;272(9):2118–31.
[49] Kirino Y, Vourekas A, Kim N, de Lima Alves F, Rappsilber J, Klein PS, et al. Arginine methylation of vasa protein is conserved across phyla. J Biol Chem 2010;285(11):8148–54.
[50] Young BD, Weiss DI, Zurita-Lopez CI, Webb KJ, Clarke SG, McBride AE. Identification of methylated proteins in the yeast small ribosomal subunit: a role for SPOUT methyltransferases in protein arginine methylation. Biochemistry 2012;51(25):5091–104.
[51] Lipson RS, Webb KJ, Clarke SG. Rmt1 catalyzes zinc-finger independent arginine methylation of ribosomal protein Rps2 in Saccharomyces cerevisiae. Biochem Biophys Res Commun 2010;391(4):1658–62.
[52] Swiercz R, Person MD, Bedford MT. Ribosomal protein S2 is a substrate for mammalian PRMT3 (protein arginine methyltransferase 3). Biochem J 2005;386(Pt 1):85–91.
[53] Jackson CA, Yadav N, Min S, Li J, Milliman EJ, Qu J, et al. Proteomic analysis of interactors for yeast protein arginine methyltransferase Hmt1 reveals novel substrate and insights into additional biological roles. Proteomics 2012;12(22): 3304–14.
[54] Friesen WJ, Massenet S, Paushkin S, Wyce A, Dreyfuss G. SMN, the product of the spinal muscular atrophy gene, binds preferentially to dimethylarginine-containing protein targets. Mol Cell 2001;7(5):1111–7.
[55] Palfi Z, Jae N, Preusser C, Kaminska KH, Bujnicki JM, Lee JH, et al. SMN-assisted assembly of snRNP-specific Sm cores in trypanosomes. Genes Dev 2009;23(14):1650–64.
[56] Tkacz ID, Gupta SK, Volkov V, Romano M, Haham T, Tulinski P, et al. Analysis of spliceosomal proteins in Trypanosomatids reveals novel functions in mRNA processing. J Biol Chem 2010;285(36):27982–99.
[57] Boisvert FM, Hendzel MJ, Masson JY, Richard S. Methylation of MRE11 regulates its nuclear compartmentalization. Cell Cycle 2005;4(7):981–9.
[58] Boisvert FM, Dery U, Masson JY, Richard S. Arginine methylation of MRE11 by PRMT1 is required for DNA damage checkpoint control. Genes Dev 2005;19(6):671–6.
[59] El-Andaloussi N, Valovka T, Toueille M, Steinacher R, Focke F, Gehrig P, et al. Arginine methylation regulates DNA polymerase beta. Mol Cell 2006;22(1):51–62.
[60] El-Andaloussi N, Valovka T, Toueille M, Hassa PO, Gehrig P, Covic M, et al. Methylation of DNA polymerase beta by protein arginine methyltransferase 1 regulates its binding to proliferating cell nuclear antigen. FASEB J 2007;21(1):26–34.
[61] Tye BK. MCM proteins in DNA replication. Annu Rev Biochem 1999;68:649–86.
[62] Dang HQ, Li Z. The Cdc45.Mcm2-7.GINS protein complex in trypanosomes regulates DNA replication and interacts with two Orc1-like proteins in the origin recognition complex. J Biol Chem 2011;286(37):32424–35.
[63] Nishimura K, Ishiai M, Horikawa K, Fukagawa T, Takata M, Takisawa H, et al. Mcm8 and Mcm9 form a complex that functions in homologous recombination repair induced by DNA interstrand crosslinks. Mol Cell 2012;47(4):511–22.
[64] Kim HS, Park SH, Gunzl A, Cross GA. MCM-BP is required for repression of life-cycle specific genes transcribed by RNA polymerase I in the mammalian infectious form of *Trypanosoma brucei*. PLoS One 2013;8(2):e57001.
[65] Nojima H. Protein kinases that regulate chromosome stability and their downstream targets. Genome Dyn 2006;1:131–48.
[66] Ackers JP, Dhir V, Field MC. A bioinformatic analysis of the RAB genes of *Trypanosoma brucei*. Mol Biochem Parasitol 2005;141(1):89–97.
[67] Dhir V, Goulding D, Field MC. TbRAB1 and TbRAB2 mediate trafficking through the early secretory pathway of *Trypanosoma brucei*. Mol Biochem Parasitol 2004;137(2): 253–65.
[68] Stepp JD, Huang K, Lemmon SK. The yeast adaptor protein complex, AP-3, is essential for the efficient delivery of alkaline phosphatase by the alternate pathway to the vacuole. J Cell Biol 1997;139(7):1761–74.
[69] Conibear E, Stevens TH. Multiple sorting pathways between the late Golgi and the vacuole in yeast. Biochim Biophys Acta 1998;1404(1–2):211–30.
[70] Conibear E, Cleck JN, Stevens TH. Vps51p mediates the association of the GARP (Vps52/53/54) complex with the late Golgi t-SNARE Tlg1p. Mol Biol Cell 2003;14(4):1610–23.
[71] Perez-Victoria FJ, Schindler C, Magadan JG, Mardones GA, Delevoye C, Romao M, et al. Ang2/fat-free is a conserved subunit of the Golgi-associated retrograde protein complex. Mol Biol Cell 2010;21(19):3386–95.
[72] Huang G, Fang J, Sant'Anna C, Li ZH, Wellems DL, Rohloff P, et al. Adaptor protein-3 (AP-3) complex mediates the biogenesis of acidocalcisomes and is essential for growth and virulence of *Trypanosoma brucei*. J Biol Chem 2011;286(42): 36619–30.
[73] He B, Guo W. The exocyst complex in polarized exocytosis. Curr Opin Cell Biol 2009;21(4):537–42.
[74] Silverman JS, Bangs JD. Form and function in the trypanosomal secretory pathway. Curr Opin Microbiol 2012;15(4):463–8.
[75] Wooderchak WL, Zang T, Zhou ZS, Acuna M, Tahara SM, Hevel JM. Substrate profiling of PRMT1 reveals amino acid sequences that extend beyond the "RGG" paradigm. Biochemistry 2008;47(36):9456–66.
[76] Fischle W, Wang Y, Allis CD. Binary switches and modification cassettes in histone biology and beyond. Nature 2003;425(6957):475–9.

[77] Yun CY, Fu XD. Conserved SR protein kinase functions in nuclear import and its action is counteracted by arginine methylation in *Saccharomyces cerevisiae*. J Cell Biol 2000;150(4):707–18.
[78] Sims III RJ, Rojas LA, Beck D, Bonasio R, Schuller R, Drury III WJ, et al. The C-terminal domain of RNA polymerase II is modified by site-specific methylation. Science 2011;332(6025):99–103.
[79] Janzen CJ, Fernandez JP, Deng H, Diaz R, Hake SB, Cross GA. Unusual histone modifications in *Trypanosoma brucei*. FEBS Lett 2006;580(9):2306–10.
[80] Mandava V, Fernandez JP, Deng H, Janzen CJ, Hake SB, Cross GA. Histone modifications in *Trypanosoma brucei*. Mol Biochem Parasitol 2007;156(1):41–50.
[81] Yang Y, Bedford MT. Protein arginine methyltransferases and cancer. Nat Rev Cancer 2013;13(1):37–50.