Eukaryote-Conserved Methylarginine Is Absent in Diplomonads and Functionally Compensated in *Giardia*

Samantha J. Emery-Corbin,*[1,2] Joshua J. Hamey,[3] Brendan R.E. Ansell,[1,2] Balu Balan,[1,2,4] Swapnil Tichkule,[1,2] Andreas J. Stroehlein,[4] Crystal Cooper,[5] Bernie V. McInerney,[6] Soroor Hediyeh-Zadeh,[2,7] Daniel Vuong,[8] Andrew Crombie,[8] Ernest Lacey,[8,9] Melissa J. Davis,[2,7] Marc R. Wilkins,[3] Melanie Bahlo,[1,2] Staffan G. Svärd,[10] Robin B. Gasser,[4] and Aaron R. Jex[1,2,4]

[1]Population Health and Immunity Division, The Walter and Eliza Hall Institute of Medical Research, Melbourne, VIC, Australia

[2]Department of Medical Biology, The University of Melbourne, Parkville, VIC, Australia

[3]School of Biotechnology and Biomolecular Sciences, University of New South Wales, Sydney, NSW, Australia

[4]Department of Veterinary Biosciences, Melbourne Veterinary School, Faculty of Veterinary and Agricultural Sciences, The University of Melbourne, Parkville, VIC, Australia

[5]Central Analytical Research Facility (CARF), Institute for Future Environments, Queensland University of Technology, Brisbane, QLD, Australia

[6]Australian Proteome Analysis Facility (APAF), Macquarie University, North Ryde, NSW, Australia

[7]Bioinformatics Division, The Walter and Eliza Hall Institute of Medical Research, Parkville, VIC, Australia

[8]Microbial Screening Technologies, Smithfield, NSW, Australia

[9]Chemistry and Biomolecular Sciences, Faculty of Science, Macquarie University, North Ryde, NSW, Australia

[10]Department of Cell and Molecular Biology, Uppsala University, Uppsala, Sweden

***Corresponding author:** E-mail: emery.s@wehi.edu.au.

Associate editor: Claudia Russo

PRIDE Data Reviewer Login Details
Project Name: Quantitative proteomics of *Giardia* exposures to HKMT inhibitors
Project accession: PXD016747
Username: reviewer03656@ebi.ac.uk
Password: ytC0i80t
Project Name: Arginine and lysine methylproteomes of *Giardia duodenalis*
Project accession: PXD016813
Username: reviewer56627@ebi.ac.uk
Password: dOkSWMPD

Abstract

Methylation is a common posttranslational modification of arginine and lysine in eukaryotic proteins. Methylproteomes are best characterized for higher eukaryotes, where they are functionally expanded and evolved complex regulation. However, this is not the case for protist species evolved from the earliest eukaryotic lineages. Here, we integrated bioinformatic, proteomic, and drug-screening data sets to comprehensively explore the methylproteome of *Giardia duodenalis*—a deeply branching parasitic protist. We demonstrate that *Giardia* and related diplomonads lack arginine-methyltransferases and have remodeled conserved RGG/RG motifs targeted by these enzymes. We also provide experimental evidence for methylarginine absence in proteomes of *Giardia* but readily detect methyllysine. We bioinformatically infer 11 lysine-methyltransferases in *Giardia*, including highly diverged Su(var)3-9, Enhancer-of-zeste and Trithorax proteins with reduced domain architectures, and novel annotations demonstrating conserved methyllysine regulation of eukaryotic elongation factor 1 alpha. Using mass spectrometry, we identify more than 200 methyllysine sites in *Giardia*, including in species-specific gene families involved in cytoskeletal regulation, enriched in coiled-coil features. Finally, we use known methylation inhibitors to show that methylation plays key roles in replication and cyst formation in this parasite. This study highlights reduced methylation enzymes, sites, and functions early in eukaryote evolution, including absent methylarginine networks in the Diplomonadida. These results challenge the view that arginine methylation is eukaryote conserved and demonstrate that functional compensation of methylarginine was possible preceding expansion and diversification of these key networks in higher eukaryotes.

Key words: **methylproteome, methylarginine, methyllysine, Diplomonadida, Metamonada, *Giardia*.**

Introduction

Posttranslational modifications (PTMs) of proteins coordinate cell development, signaling, and gene regulation (Larsen et al. 2016; Blanc and Richard 2017; Murn and Shi 2017). Methylations (Me) of lysine (Lys/K/K-Me) and arginine (Arg/R/R-Me) are common PTMs of eukaryotic proteins, catalyzed by methyltransferases (MTases) enzyme classes that arose during the evolution of eukaryotes and are best known for epigenetic regulation (Lanouette et al. 2014; Murn and Shi 2017). Improved technologies have identified abundant methylated substrates in higher eukaryotes (Lanouette et al. 2014; Moore and Gozani 2014) and roles for methylation in metabolism, translation, and the modulation of RNA-binding proteins (Erce et al. 2012; Castello, Horos, et al. 2016; Murn and Shi 2017; Hamey and Wilkins 2018). These discoveries have driven identification of new protein MTase classes (Cloutieret al. 2013; Hamey and Wilkins 2018) and their extrachromatin specificities (Dillonet al. 2005; Moore and Gozani 2014; Hamey et al. 2018), with relevance for cell development in all eukaryotes, and clinical relevance for neurological disorders and cancer in humans (Guccione and Richard 2019; Han et al. 2019).

Eukaryotic methylproteomes are best characterized in mammals and yeast. These methylproteomes exhibit numerous evolutionary expansions of methylated substrates (Larsen et al. 2016) and MTases (Bachand 2007; Lanouette et al. 2014), with vertebrate methylproteomes further featuring multisite and state (mono-, di-, and tri-) methyl-regulation (e.g., histones [Lanouette et al. 2014]). *Saccharomyces cerevisiae* (budding yeast) represents the most complete eukaryotic methylation model with simpler, conserved methylproteomes (Low and Wilkins 2012). However, yeasts have already accumulated paralogous MTases with expanded architectures and new regulators (Iyer et al. 2008), and diverged from the methylation networks which evolved with eukaryote-specific protein methylation (R-Me, K-Me). This complexity presents obstacles in identifying regulators of nonhistone and nonnuclear protein methylation (Murn and Shi 2017), deciphering functional significance between methylated and nonmethylated protein interactomes (Cornett et al. 2019), and understanding PTM coregulation (e.g., phosphorylation [Larsen et al. 2016]). Identifying a simpler eukaryotic model would aid in unraveling this complexity and define eukaryote-conserved functions for further study.

Protists have evolved from the oldest eukaryote lineages and are ideal species to deconstruct this complexity. Many protists have simplified eukaryotic traits derived from deep-branching origins and genomes shaped by secondary losses during ecological adaptations (Leger et al. 2017), as is the case for parasitic protists (Morrison et al. 2007). *Giardia duodenalis* (order Diplomonadida) belongs to one of the deepest-branching phyla, the Metamonada, and is a parasite of the small intestine of mammals. In humans, *Giardia* infection causes ~200 million cases of diarrheal disease per year (Lane and Lloyd 2002), with infection and transmission via virulent trophozoites and environmentally resilient cysts, respectively.

Metamonads have many reduced or eliminated eukaryote-conserved gene families (Morrison et al. 2007; Leger et al. 2017), including protein MTases (Iyer et al. 2008; Fisk and Read 2011). There are five MTase Classes (I–V), and those targeting proteins are usually Class I (seven-stranded β-sheet [7βS]) and Class V ("Su(var)3-9, Enhancer-of-zeste and Trithorax," SET-domain) (Schubert et al. 2003). Early investigations of *Giardia* protein MTases suggest an absence of arginine methyltransferase (PRMT) classes (Iyer et al. 2008; Fisk and Read 2011) and divergent Class V n-lysine methyltransferases (K-MTases) (Sonda et al. 2010). These findings highlight reduced methylproteomes in *Giardia* accordant with its deep-branching origins but which have been implicated in regulating parasite development, antimicrobial resistance, and virulence (Sonda et al. 2010; Carranza et al. 2016; Salusso et al. 2017; Emery et al. 2018). *Giardia* MTases are essentially unexplored but could represent novel targets to disrupt parasites and their life cycles.

Giardia is a protist whose evolution precedes the expanded methylproteome of yeast and vertebrates, with anticipated unique adaptions to its host gastrointestinal niches. It is an ideal lineage to contrast with model organisms, including humans, to better understand the evolved complexity of higher eukaryotic methylproteomes. Here, we characterize the methylproteome of *G. duodenalis* and explore its evolution within the Metamonada. Notably, we localize a lack of PRMT enzymes and their preferred motifs in Diplomonadida species and do not detect methylarginine in the *Giardia* proteome by multiple methods, confirming absence of arginine methylation networks—a first among eukaryotes. Using mass spectrometry (MS) and novel methyl-site filtering strategies, we identified over 200 high-confidence K-Me sites in *Giardia*. These provide the evolutionarily earliest evidence of conserved eukaryotic elongation factor 1 alpha (eEF1a) lysine methylation and abundant species-specific methyllysine adaptations, enriched in coiled-coils within cytoskeletal proteins. Lastly, we disrupt in vitro models of parasite transmission and virulence using methylation inhibitors, demonstrating functional roles for K-Me in *Giardia* cytoskeletal regulation and encystation. This comprehensive functional and evolutionary exploration of the *Giardia* methylproteome provides potential avenues to specifically inhibit parasites and new insight into the evolution of protein methylation within the eukaryota.

Results

In Silico Curation of the Methyltransferases

We bioinformatically curated protein MTases in five Metamonada species using custom hidden Markov models (HMMs) for Class I (7βS, PRMT) and Class V (SET-domain) domains. In *Giardia*, MTase-domain-containing proteins were examined for MTase class folds using structures predicted from 3D-structural modeling of the *Giardia* proteome (Ansell et al. 2019). We identified 13 protein MTases in *G. duodenalis*, of which 11 were putative K-MTases (6 Class V; 5 Class I), and the remainder were homologs of N-terminal protein methyltransferase 1 (NTM1; GL_12215) and leucine

carboxyl methyltransferase (PPM1; GL_10516). No PRMT domains were identified in *Giardia* proteins.

Diverged Class V K-MTases

Class V K-MTases contain a SET-domain "pseudoknot" structure with separate substrate- and S-adenosyl methionine (SAM)-binding clefts facilitating progressive lysine methylation (mono-, di-, and tri-; K-MMe, K-DMe, and K-TMe) (Schubert et al. 2003). We confirmed six previously reported Class V K-MTases using custom HMM (Iyer et al. 2008), with no evidence of additional candidates (fig. 1*A* and supplementary table 1, Supplementary Material online). All six *Giardia* proteins encoded conserved "pseudoknot" residues (Dillon et al. 2005) (supplementary fig. S1, Supplementary Material online). Phylogenetic analysis of SET-domain sequences from *Giardia* and higher eukaryotes revealed only two (GL_9130, GL_17036) clustered with SET-domain families of higher eukaryotes, whereas the remainder formed *Giardia*-specific branches, indicating divergence and radiation (fig. 1*B* and supplementary fig. S3, Supplementary Material online). Overall, *Giardia* SET-domain proteins have a reduced domain architecture (fig. 1*A*), likely impacting on protein–protein interactions and facilitating broader substrate specificity. For example, despite strong phylogenetic association between SET-domains of GL_9130 and human SETD2/NSD1, there are multiple additional domains within the >2,000 AA residues encoded in human orthologs (fig. 1*B*).

Structural modeling of the *Giardia* proteome (Ansell et al. 2019) predicted full-sequence structures with a Class V pseudoknot fold for GL_9130, GL_221691, GL_17036, and GL_13838. Here, we also modeled SET-domain-only structures for GL_8921 and GL_6407 with predicted Class V folds. All six predicted structures (fig. 1*C*) were most similar to experimentally determined histone *n*-lysine methyltransferases (HKMTs) structures available in the Protein Data Bank (PDB) (supplementary table 2, Supplementary Material online). Within predicted structures, SET-domain proteins (GL_9130, GL_221961, GL_8291, and GL_13838) were distinct from SET & Myeloid-Nervy-DEAF1 (MYND) domain proteins (SMYD; GL_17036, GL_6407). Detailed domain searching in GL_17036 and GL_6407 identified SMYD domain signatures (PTHR12197) but not a MYND zinc finger, which is common for divergent SMYD proteins (Calpena et al. 2015). Interestingly, the GL_221691-predicted structure had domain folds matching both a SET-domain and SET- and RING-ASSOCIATED (SRA) domain, which could permit dual methylation of both proteins (SET) and nucleic acids (SRA) (Johnson et al. 2007) (supplementary fig. S4, Supplementary Material online). Sequence analyses of *Giardia* SET-domains further supported SET and SMYD classifications (supplementary fig. S1, Supplementary Material online), with the F/Y switch present in GL_9130 (F) and GL_13838 (Y). These F/Y residues confer methyllysine-product specificity (mono-, di-, and tri) (Dillon et al. 2005). The lack of this switch and reduced domain architecture in other *Giardia* K-MTases may reduce their substrate specificity.

Novel Class I K-MTases and Conserved eEF1a Methylation

Class I MTases have a 7βS fold and can methylate nucleic acids, proteins, or metabolites (Schubert et al. 2003). We identified five novel MTases in *Giardia* through HMMs for eukaryotic Class I K-MTases (Tempel et al. 2009; Kernstock et al. 2012; Cloutier et al. 2013) which had predicted protein structures with 7βS-folds matching PDB structures of methyllysine-specific Class I MTases (Tempel et al. 2009; Kernstock et al. 2012; Cloutier et al. 2013), particularly eEF1a-K-MTases (Hamey and Wilkins 2018; Jakobsson, Małecki, and Falnes 2018) (fig. 2). Similar sequence orthologs were identified in other Metamonada species (supplementary table 2, Supplementary Material online) and support deep-branching origins of these eEF1a-K-MTase families in eukaryotes.

Two *G. duodenalis* Class I MTases (GL_100959, DHA_151673) were homologous to family 16 (MTF16) METTL21 K-MTases (Hamey, Hart-Smith, et al. 2016). Both shared sequence and predicted structural homology with human METTL21D (fig. 2*A* and supplementary fig. S5, Supplementary Material online) (Kernstock et al. 2012), and DHA_151673 encoded the MTF16 catalytic methylation motif, "D/ExxF/Y" (Kernstock et al. 2012; Hamey, Hart-Smith, et al. 2016). In humans, METTL21D methylates K315 of Valosin-containing protein (VCP), but this lysine residue is not conserved in the *Giardia* VCP homolog (GL_16867; supplementary fig. S6, Supplementary Material online). However, MTF16 K-MTases can also target eEF1a (Hamey, Winter, et al. 2016; Hamey et al. 2017; Malecki et al. 2017), and *Giardia* DHA_151673 shared 37% sequence similarity with Efm7 from *S. cerevisiae*, which catalyzes eEF1a N-terminal methylation (supplementary fig. S7, Supplementary Material online). Three additionally identified *Giardia* genes (GL_3948, GL_4349, and GL_5013) may methylate eEF1a, given their sequence and structural homology to eEF1A-KMT4 within the Efm4 family of eEF1A-K-MTases (fig. 2*B* and supplementary fig. S8, Supplementary Material online) (Jakobsson et al. 2017; Hamey and Wilkins 2018). The current annotation of GL_4349 and GL_5013 as "Endothelin-converting enzyme 2" coincides with eEF1A-KMT annotations within human ECE2 protein prior to its reannotation (Tempel et al. 2009).

The domain architecture of the predicted protein structure of *Giardia* eEF1a is similar to human eEF1a (fig. 3*A*) and encodes lysine residues analogous to methylation targets in other eukaryotes (fig. 3*B* and supplementary fig. S9, Supplementary Material online) (Hamey and Wilkins 2018). We detected K55 dimethylation of *Giardia* eEF1a based on antibody-enrichment of K-Me peptides coupled to liquid chromatography (LC)-MS/MS. The K55 site is also methylated in humans by eEF1A-KNMT (Jakobsson, Małecki, Halabelian, et al. 2018) from the Efm4 family; therefore, *Giardia* Efm4-like MTases (fig. 2*B*) might methylate K55 of *Giardia* eEF1a. We also detected N-terminal trimethylation of *Giardia* eEF1a using targeted MS (Hamey, Winter, et al. 2016) (supplementary fig. S10, Supplementary Material online). Given DHA_151673 shared homology with Efm7 from

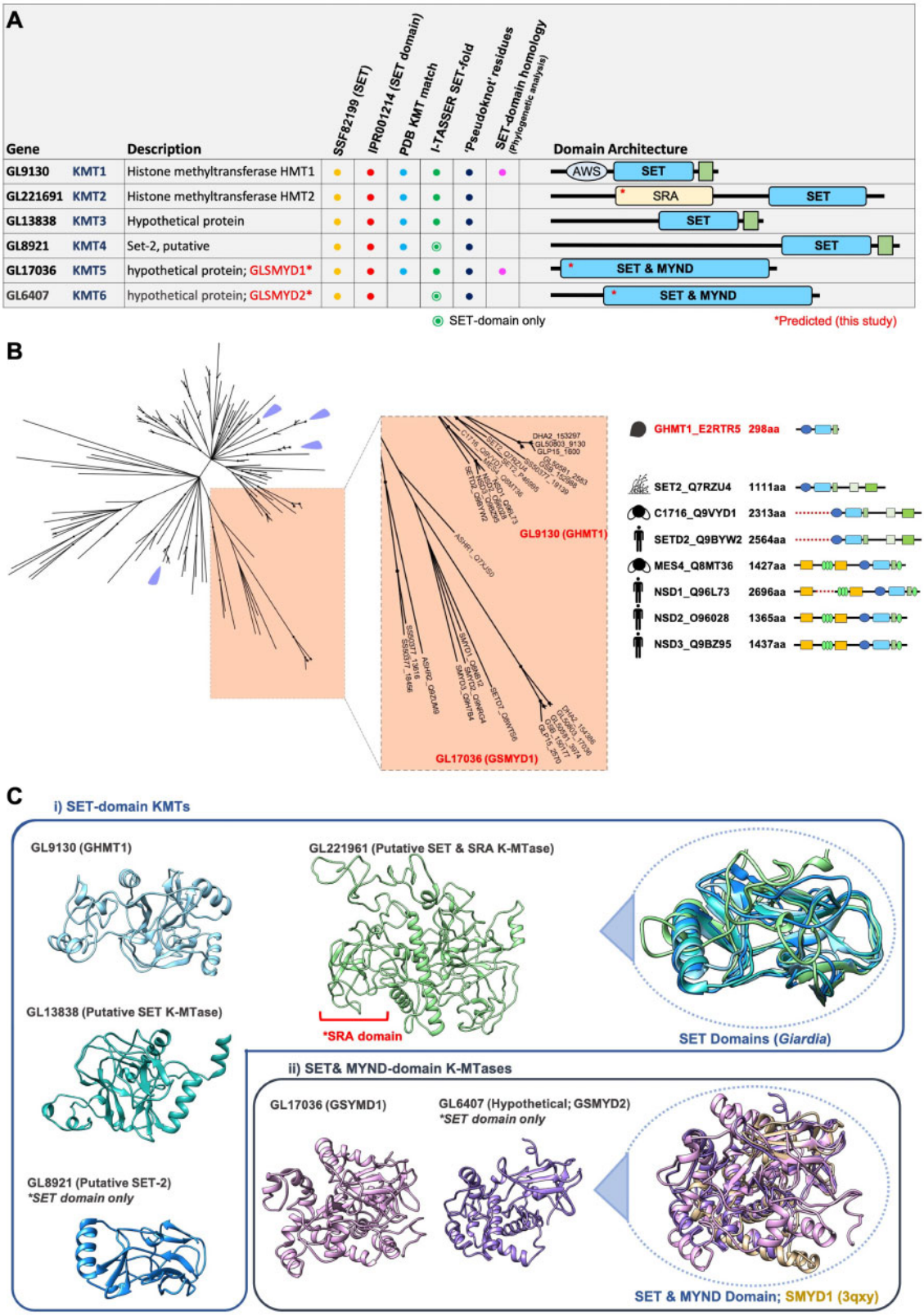


FIG. 1. *Giardia* Class V SET-domain methyltransferases. (*A*) Bioinformatic evidence for Class V methyltransferase annotation in *Giardia*, including SET-domain annotations, sequence-based homology to PDB structures of Class V HKMTs via EuPATHdb, and classification of Class V SET-domain "pseudoknot" evidence at sequence and structural levels. SET-domains homology based on phylogenetic analyses from reviewed HKMTs of model eukaryotes are also considered. Annotated domains and domain architecture relative to protein length are shown in the right column. Additional protein information is detailed in supplementary table 1, Supplementary Material online. (*B*) Phylogenetic analysis of the SET-domain from the six *Giardia* Class V K-MTases with SET-domain sequences from >100 HKMT proteins submitted to the SwissProt database from *Homo sapiens*, *Arabidopsis thaliana*, *Drosophila melanogaster*, *Neurospora crassa*, and *Saccharomyces cerevisiae*. Branch support >80 is depicted with a ●, and

S. cerevisiae (Hamey, Winter, et al. 2016), DHA_151673 may catalyze observed N-terminal methylation. Considering N-terminal methylation is shared between yeast, humans, and *Giardia*, this modification might be conserved in eukaryotes. However, this site has only been identified through targeted MS of ASP-N derived peptides in yeast and humans (Hamey, Winter, et al. 2016), as well as *Giardia* in this study. N-terminal methylation blocks Edman degradation sequencing and to date has not been identified in MS analyses of tryptic peptides, both of which are the main sources of eEF1a methyl-sites (Hamey and Wilkins 2018). Therefore, we hypothesize this methyl-site is underreported across eukaryotes given that standard methods have been unlikely to identify it. Although K79-TMe is considered eukaryote conserved (Hamey and Wilkins 2018; Jakobsson, Małecki, and Falnes 2018), no N6-adenine DNA methyltransferase-like orthologs (Dzialo et al. 2014; Hamey, Winter, et al. 2016) were identified in metamonad species (supplementary table 2, Supplementary Material online), and we did not detect K79 methylation in *Giardia*. Together, bioinformatically curated Class I K-MTases and experimentally detected eEF1a K-Me sites in *Giardia* highlight early evolution of protein translation regulation in eukaryotes.

Losses of Arginine Methylation Networks from the Diplomonadida

Arginine methylation is catalyzed by nine PRMT families in humans (Bachand 2007). Type I PRMTs (PRMT1–4, 6, and 8) catalyze mono- (R-MMe) and asymmetric dimethylation (ADMA). Type II enzymes (PRMT5, 9) catalyze mono- and symmetrical dimethylation (SDMA), whereas Type III enzyme (PRMT7) catalyzes monomethylation (Blanc and Richard 2017). In unicellular eukaryotes, PRMT1 (Type I) and PRMT5 (Type II) orthologs are conserved, and, together, this combination is minimally required to catalyze all R-MMe, ADMA, and SDMA methyl-forms (Bachand 2007; Fisk and Read 2011).

PRMTs are absent in all published *Giardia* spp. genomes (Morrison et al. 2007; Iyer et al. 2008; Adam et al. 2013). Although previous HMM explored CARM1 (PRMT Type I) domains in clinical protists, only *Giardia* and *Trichomonas* were included from the Metamonada (Iyer et al. 2008). To assess whether PRMT absence was specific to *Giardia* or occurred in other Metamonada, we built new HMMs including all nine PRMT families of Type I and II classes and searched additional Metamonada genomes which had not been explored for methyltransferases. We included genomes within the four Metamonada lineages as described by Leger et al. (2017), which included species of the Diplomonadida (*G. duodenalis*, *Spironucleus salmonicida*, and *Trepomonas* spp.), Fornicata (*Kipferlia bialata* [Tanifuji et al. 2018]), Preaxostyla (*Monocercomonoides* sp. [Karnkowska et al. 2016]), and known PRMTs of the Parabasalia (*Trichomonas vaginalis* [Iyer et al. 2008; Fisk and Read 2011]). This identified PRMT1 and PRMT5 orthologs in three of four lineages, excepting the Diplomonadida (fig. 4A and supplementary fig. S11, Supplementary Material online). These findings suggest that the absence of PRMTs is restricted to the Diplomonadida and relate to genome reduction that has occurred during evolutionary divergences of Diplomonadida from Fornicata (Leger et al. 2017).

Glycine-neighboring-arginine RGG/RG motifs are preferred methylation sites for eukaryote-conserved PRMT1 and PRMT5 (Thandapani et al. 2013; Blanc and Richard 2017), with >1,000 sites occurring in some proteomes (Guo et al. 2014; Larsen et al. 2016). RGG-containing R-Me targets include fibrillarin, Gar1, 40S ribosomal protein S2 (RPS2), RNA helicases, and RNA-binding proteins (Thandapani et al. 2013; Yagoub et al. 2015; Chong et al. 2018). We hypothesized that RGG motifs are less conserved and reduced in number in species lacking PRMTs (fig. 4A). We searched 14 eukaryote genomes for defined RGG/RG motifs (Thandapani et al. 2013) (supplementary table 4, Supplementary Material online) and found that the RGG motif was uniquely lacking in Diplomonadida proteomes, including Gar1 and fibrillarin orthologs (fig. 4B). Interestingly, we observed conserved disordered regions in Gar1 and fibrillarin orthologs, irrespective of RGG/RG motifs (supplementary fig. S12, Supplementary Material online). In other instances, arginine residues in broadly eukaryote-conserved R-Me targets were substituted for lysine residues in the Diplomonadida (supplementary fig. S13, Supplementary Material online), including *Giardia* Histone 3 at K43, K50, K54, K55, and K70 (fig. 4C).

We also investigated presence of methylarginine-interacting Tudor domains in model eukaryotes, parasitic protists, and Diplomonadida (supplementary fig. S14, Supplementary Material online), noting that *Giardia* has K-Me-interacting, chromo-recognition domains (Iyer et al. 2008). Ancestral staphylococcal nuclease domain-containing 1 (SND1) protein is conserved in eukaryotes (Liu et al. 2010) and has ribonuclease functions in RNAi complexes (Caudy et al. 2003). SND1 contains a single Tudor domain, which can bind methylarginine when at least three-fourths aromatic phenylalanine/tyrosine residues are present (Liu et al. 2010). We observed multiple nonaromatic substitutions in Tudor domains of the Diplomonadida, *Entamoeba*, and

Fig. 1. Continued

purple arrows indicate orphan clades of *Giardia duodenalis* SET-domain sequences that lack significant homology to SET-domains from representative, eukaryotes species. A full tree can be viewed in supplementary figure S1, Supplementary Material online. Only SET-domains from GL_9130 and GL_17036 have branch homology to SET-domains from higher organisms, although reduced additional domain architecture is noted in GL_9130 compared with SET2 and NSD1 orthologs (shown on the right). (*C*) Protein structures of Class V K-MTases in *Giardia* predicted using the program I-TASSER. There are four SET-domain-containing K-MTases in *Giardia* (GL_9130, GL_13838, GL_221961, and GL_8921), which have the pseudoknot fold present in the Class V SET-domain architecture (Schubert et al. 2003). I-TASSER also predicted a SET- and RING-ASSOCIATED domain (SRA) for the SET-domain in GL_221961. *Giardia* encodes two SET and MYND domain Class V methyltransferases for GL_17036 and GL_6407 (SET-domain only) which are consistent with the human SMYD1 structure (PDB:3qxy; SET-domain shown only).

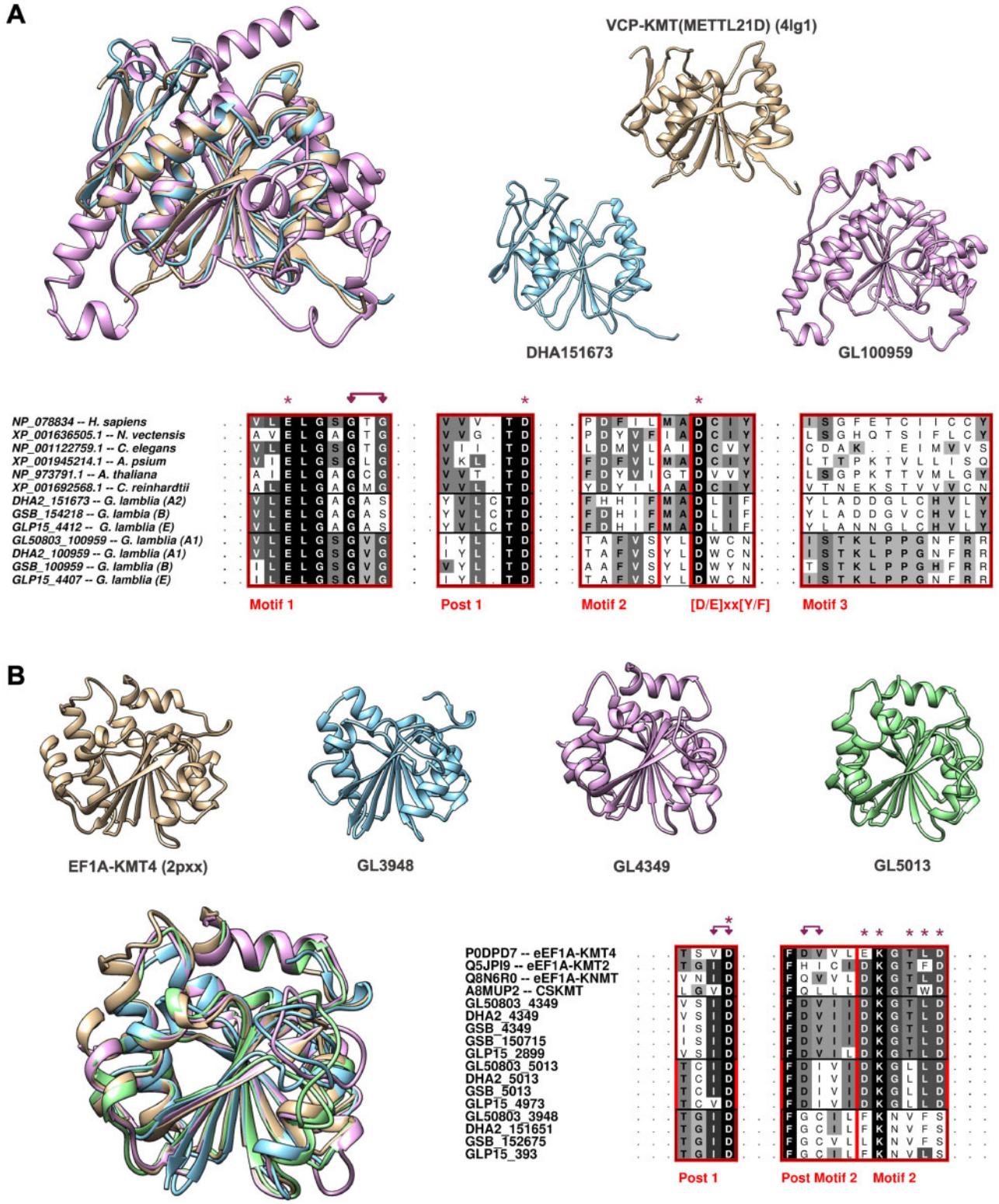
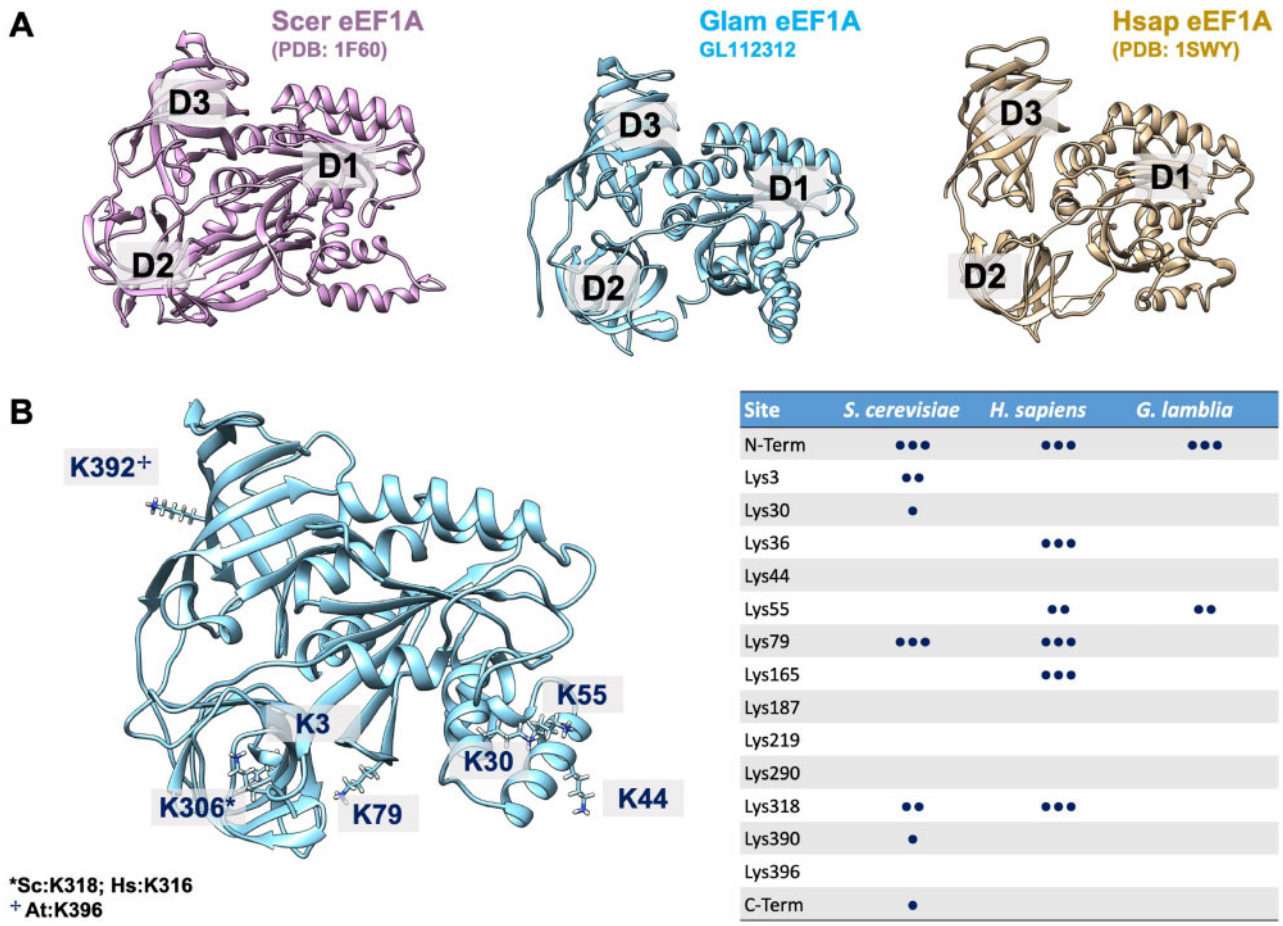


FIG. 2. *Giardia* Class I 7βS putative K-MTases. (*A*) Structure of VCP-KMT from humans (brown) compared with structural orthologs from *Giardia* DHA_151673 (blue) and GL_100959 (purple) predicted by I-TASSER. Sequence alignment of VCP-KMT from humans and other orthologs compared with *Giardia* and its subspecies orthologs share sequence homology in motif, post, and catalytic regions. DHA_151673 shares the [D/E]xx[y/F] motif essential for methyltransferase activity. (*B*) Structure of eEF1a-KMT4 from humans, previously annotated as ECE2 (Tempel et al. 2009), compared with putative eEF1a-K-MTases from *Giardia* (GL_3948, GL_4349, and GL_5013). Sequence homology is shown for key motif regions with eEF1a-KMT4 and other Efm4 family K-MTases for eEF1a.



Site	*S. cerevisiae*	*H. sapiens*	*G. lamblia*
N-Term	●●●	●●●	●●●
Lys3	●●		
Lys30	●		
Lys36		●●●	
Lys44			
Lys55		●●	●●
Lys79	●●●	●●●	
Lys165		●●●	
Lys187			
Lys219			
Lys290			
Lys318	●●	●●●	
Lys390	●		
Lys396			
C-Term	●		

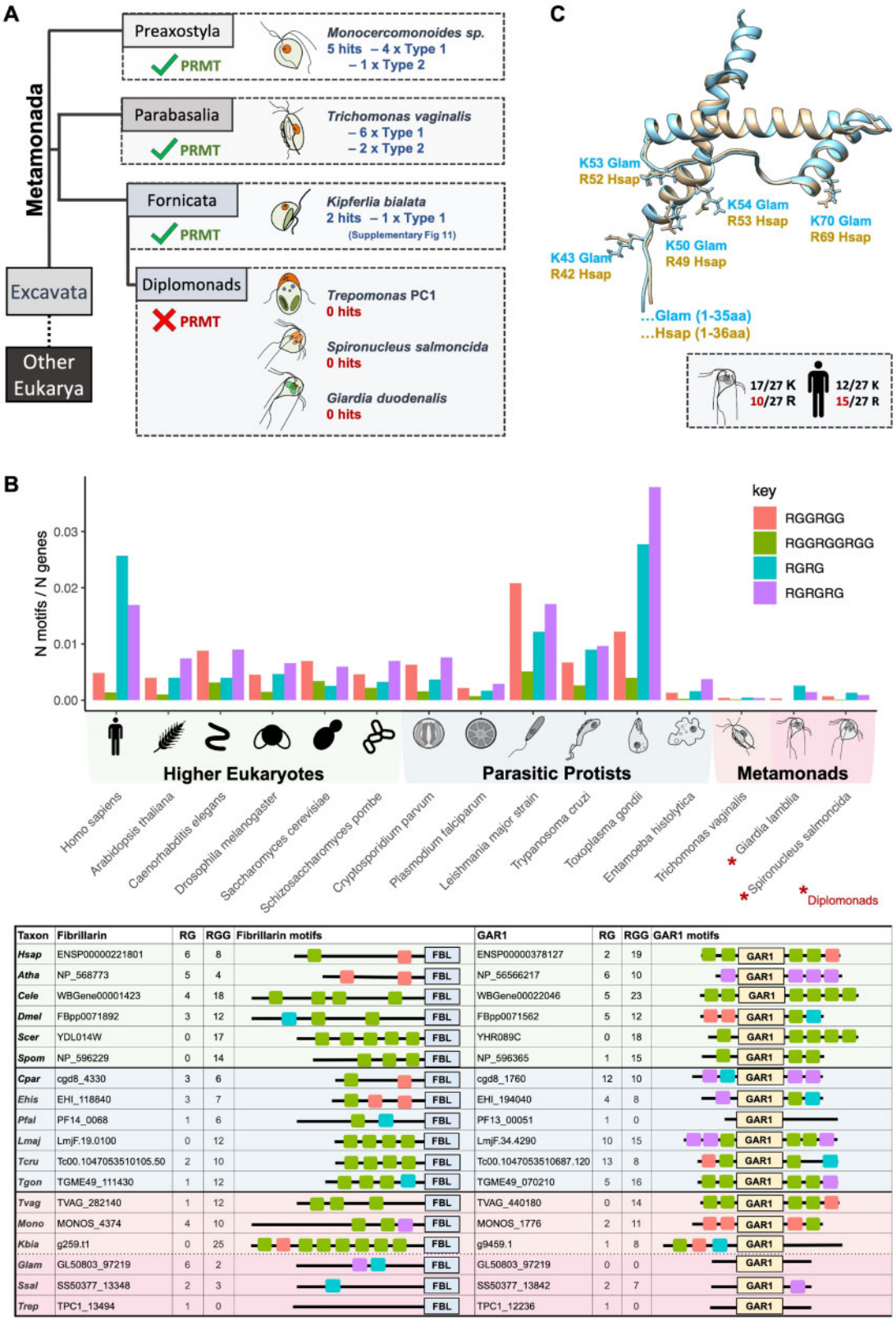
Fig. 3. *Giardia* eEF1a structure and conserved lysine residues. (*A*) Structural comparison of the *Giardia* eEF1a (GL_112312) structure predicted using I-TASSER with eEF1a structures from humans and yeast. Domain (D1, D2, and D3) architecture is observed as closer to humans than yeast. (*B*) Conserved methylated lysine residues (Hamey and Wilkins 2018) of *Giardia* eEF1a based on MSA (supplementary fig. S7, Supplementary Material online) shown on the predicted structure (left) and site mapping of K-Me from humans and yeast compared with *Giardia* (● = mono; ●● = di; and ●●● = tri) for K55 dimethylation and N-terminal trimethylation.

Trichomonas spp. (supplementary fig. S14, Supplementary Material online) and propose that no methylarginine-recognition domains occur in these species. These deep-branching protists support the proposal that *Schizosaccharomyces pombe* SND1, which also has nonaromatic Tudor domain substitutions, exhibits an "ancestral fold" preceding the evolution of methylarginine-binding in higher eukaryotes (Chen et al. 2011).

Giardia Has Detectable Methyllysine, but Not Methylarginine

We analyzed in vitro concentrations of *Giardia* R-Me and K-Me with amino acid analysis (AAA) (supplementary table 5, Supplementary Material online). First, TUV detection (260 nm) of K-MMe estimated 0.27 and 0.33 µmol/g in *Giardia* trophozoite and cyst lysates, respectively, and 0.50 µmol/g in HeLa cell-lysate controls. We detected all three methylarginine peaks in HeLa lysate, and no peaks in *Giardia* (Supplementary Material online). The HeLa R-MMe and ADMA levels were calculated at 0.526 and 5.89 µmol/g, respectively, but SDMA levels were below limits for reliable quantitation. Since estimated K-MMe levels in *Giardia* lysates were near the TUV limit of detection (LOD), we performed more sensitive AAAs with a single quadrupole (QDa) mass detector operating in selected ion recording (SIR) mode (fig. 5A). Noting that all PRMT enzymes catalyze R-MMe, we selected the 359.2-Da parent ion for R-MMe (3.31 min), and the 501.2-Da parent ion for K-MMe (7.01 min) (supplementary table 5, Supplementary Material online). R-MMe and K-MMe gave LODs of 20 and 10 fmol/µl, respectively. No R-MMe peak was detected in *Giardia* lysates, and R-MMe concentrations for HeLa lysate was 0.86 µmol/g. In contrast, we detected K-MMe in both *Giardia* and HeLa protein lysates at concentrations of 0.32 and 0.26 µmol/g in trophozoite and cyst lysates, respectively, and 0.56 µmol/g in HeLa. On immunoblot, we detected only K-Me modified proteins in *Giardia*, whereas all six R-Me and K-Me methyl-forms were detected in HeLa lysates (fig. 5B).

Subsequently, we tested for methylation by immunoaffinity purification (IAP) followed by LC-MS/MS of R-MMe and K-Me modified peptides from lysates of trophozoites ($n = 4$ and $n = 9$, respectively) and cysts (K-Me only; $n = 1$) (fig. 5



Taxon	Fibrillarin	RG	RGG	Fibrillarin motifs	GAR1	RG	RGG	GAR1 motifs
Hsap	ENSP00000221801	6	8	FBL	ENSP00000378127	2	19	GAR1
Atha	NP_568773	5	4	FBL	NP_56566217	6	10	GAR1
Cele	WBGene00001423	4	18	FBL	WBGene00022046	5	23	GAR1
Dmel	FBpp0071892	3	12	FBL	FBpp0071562	5	12	GAR1
Scer	YDL014W	0	17	FBL	YHR089C	0	18	GAR1
Spom	NP_596229	0	14	FBL	NP_596365	1	15	GAR1
Cpar	cgd8_4330	3	6	FBL	cgd8_1760	12	10	GAR1
Ehis	EHI_118840	3	7	FBL	EHI_194040	4	8	GAR1
Pfal	PF14_0068	1	6	FBL	PF13_00051	1	0	GAR1
Lmaj	LmjF.19.0100	0	12	FBL	LmjF.34.4290	10	15	GAR1
Tcru	Tc00.1047053510105.50	2	10	FBL	Tc00.1047053510687.120	13	8	GAR1
Tgon	TGME49_111430	1	12	FBL	TGME49_070210	5	16	GAR1
Tvag	TVAG_282140	1	12	FBL	TVAG_440180	0	14	GAR1
Mono	MONOS_4374	4	10	FBL	MONOS_1776	2	11	GAR1
Kbia	g259.t1	0	25	FBL	g9459.1	1	8	GAR1
Glam	GL50803_97219	6	2	FBL	GL50803_97219	0	0	GAR1
Ssal	SS50377_13348	2	3	FBL	SS50377_13842	2	7	GAR1
Trep	TPC1_13494	1	0	FBL	TPC1_12236	1	0	GAR1

FIG. 4. Bioinformatic analysis of R-Me enzymes and substrates in *Giardia* and diplomonads. (*A*) HMM analysis of PRMT enzymes within the phylogeny of the Metamonada (phylogeny and lineages taken from Leger et al. [2017]). PRMT enzymes from *T. vaginalis* noted from Fisk and Read (2011). PRMTs were identified in the metamonads *Monocercomonoides* sp. and *Kipferlia bialata* using HMMer, with Type I and II identifications based on sequence and structural (I-TASSER) homology. No PRMT enzymes were identified in the diplomonads *Giardia*, *Trepomonas* spp., or *Spironucleus salmonicida*. (*B*) Presence and number of RGG/RG motifs (motif classification from Thandapani et al. [2013]) in genomes of model eukaryote species, protozoan parasites, and metamonads (above). The number and organization of RGG/RG motifs in conserved orthologs of fibrillarin (left) and GAR1 (right) in eukaryote model species, protozoan parasites, and metamonad species (below). RGG/RG motifs are conserved

and supplementary table 6, Supplementary Material online). We applied novel detection parameters postdatabase searching to filter false-positive and mislocalized methyl-sites (Hart-Smith et al. 2016) (fig. 5C and supplementary fig. S15, Supplementary Material online), as methyl-stable isotope labeling in culture approaches is incompatible with *Giardia* in vitro culture. Mouse liver peptides ($n = 2$) were used as an IAP positive control, and resultant-identified mouse methyl-sites were used to assess filtering (supplementary table 7, Supplementary Material online). Base threshold filters, which required one site-fragment ion adjacent to the methyl-site, were applied in mouse samples and retained 61/67 and 17/17 known ("true") mouse R-MMe and K-Me sites, respectively. Six R-MMe sites retained in base filtering were unreported in mouse or equivalent human orthologs, and considered "false" sites. Stricter methyl-sites filters requiring two adjacent site-fragment ions reduced all K-Me and all but 1 R-Me unknown ("false") mouse methyl-sites but eliminated many of the "true" methyl-sites retained in base threshold filters.

In total, 38 R-MMe sites were identified in *Giardia* with base threshold filters, with two of these sites retained with stricter filtering requiring two adjacent site-fragment ions. *Giardia* R-Me sites lacked RGG in their site windows (+/−15aa), whereas the 67 filtered R-MMe mouse sites had strong enrichment for RGG up- and downstream as expected (supplementary fig. S16, Supplementary Material online). PRMT1 and PRMT5 preferentially methylate high-abundance RGG sites (Thandapani et al. 2013; Blanc and Richard 2017) resulting in RG-rich sequences in experimental methyl-peptide enrichments (Guo et al. 2014; Yagoub et al. 2015; Larsen et al. 2016), which correlates to the computationally inferred lack of PRMT and RGG motifs in *Giardia* (fig. 4). Taken together, these findings indicate that *G. duodenalis* lacks methylarginine PTM networks.

Methyllysine Sites in Coiled-Coils and Cytoskeletal Proteins in *Giardia*

We identified 202 K-Me sites from 160 proteins using base threshold filters in *Giardia* (supplementary table 6, Supplementary Material online). These filters were tested for high-confidence (Hart-Smith et al. 2016) using mouse methyl-sites (fig. 5C), with similar proteins and trends observed at each stringency level when applied to *Giardia* K-Me sites (fig. 6A and B). Interestingly, only 58/160 (37%) and 40/160 (26%) *Giardia* K-Me proteins have orthologs in *Homo sapiens* and *S. cerevisiae*, respectively. This highlights that most *Giardia* K-Me proteins are species- or lineage-specific as compared with methyl-substrates in model eukaryotes.

Giardia K-Me sites are enriched for acidic residues, particularly glutamatic acid (Glu/E) at −4, +1, +3, and +4 positions (fig. 6A). We specifically filtered for false-positive methyl-sites mislocalized from acidic residues (Hart-Smith et al. 2016) (Glu, Aspartic Acid) and noted Glu/E enrichment increased with stricter filtering. Functional relationships between the 160 K-Me proteins were visualized using STRING (fig. 6C) and overlapped with functional annotation clusters identified using DAVID (supplementary table 8, Supplementary Material online). This included functions in protein translation, including ribosomal proteins (GL_19436, GL_10367, GL_17244, and GL_14620), translation initiation factor IF-2 (GL_115229) as well as eEF1a (GL_112312). A cluster of RNA metabolism and cell cycle K-Me proteins in *Giardia* are noted as typically R-Me modified in humans and yeast (Yagoub et al. 2015; Larsen et al. 2016) and included NOP5 (GL_5359, snoRNA binding), Pescadillo homolog (GL_16313, rRNA maturation), an RNA-binding protein (GL_3993), and a nucleolar GTPase (GL_16498). Similarly, aminoacyl tRNA synthetases/ligases are enriched R-Me substrates in humans (Larsen et al. 2016) but are enriched in *Giardia* K-Me proteins (GL_22204, GL_354268, and GL_15708). K-Me sites in two-thirds enzymes in the *Giardia* arginine catabolism pathways (KEGG: gla00220; GL_10311, GL_16443) indicated possible roles for regulating parasite virulence (Stadelmann et al. 2013). Lastly, we confirmed roles of K-Me in epigenetic regulation (Carranza et al. 2016), identifying a chromodomain helicase (GL_112397) and H3K37-TMe (equivalent K36).

Coiled-coils and ankyrin (ANK) repeat proteins were significantly overrepresented in K-Me proteins (fig. 6B). Coiled-coils consist of heptad repeats (denoted a-b-c-d-e-f-g) configured into an alpha-helix, where "a" and "d" are nonpolar residues at the hydrophobic core, and "e" and "g" are polar residues forming electrostatic interactions between helices (Mason and Arndt 2004). These preferred residues account for unusual sequence motifs of K-Me sites in *Giardia*, as many K-Me sites are localized to annotated coiled-coil regions of *Giardia* proteins, including 67/202 (33.2%), 28/63 (44.4%), and 11/21 (52.4%) sites at the three filtering levels. Using COILS (Lupas et al. 1991), we identified specific heptad sequences with K-Me modifications in *Giardia* (fig. 6D and supplementary table 9, Supplementary Material online) and demonstrated heptad preference for Glu/E and Leu/L residues coincides with their enrichment in *Giardia* K-Me sequence motifs (fig. 6A). We also noted that K-Me heptad sites occurred on solvent-accessible "c" and "f" residues ($n = 50/62$), which are excluded from hydrophobic or electrostatic interactions driving alpha-helix formation (Mason and Arndt 2004).

Although coiled-coil K-Me sites span diverse proteins, many in *Giardia* are associated with the cytoskeleton. NimA-related kinases (NEK) and 21.1 proteins were enriched among *Giardia* K-Me proteins and are extensively radiated *Giardia* gene families containing coiled-coil and ANK-repeat features (Manning et al. 2011) (fig. 6B). *Giardia* encodes the

Fig. 4. Continued
in eukaryotes, protozoans, whereas RGG motifs are absent in Diplomonadida (supplementary table 4, Supplementary Material online). (C) Predicted structure of *Giardia* H3 variant (GL_135231; blue) compared with human H3.1 (brown). Conserved arginine-to-lysine converted sites in *Giardia* compared with humans are annotated with the structure, with the full MSA shown in supplementary figure S11, Supplementary Material online.

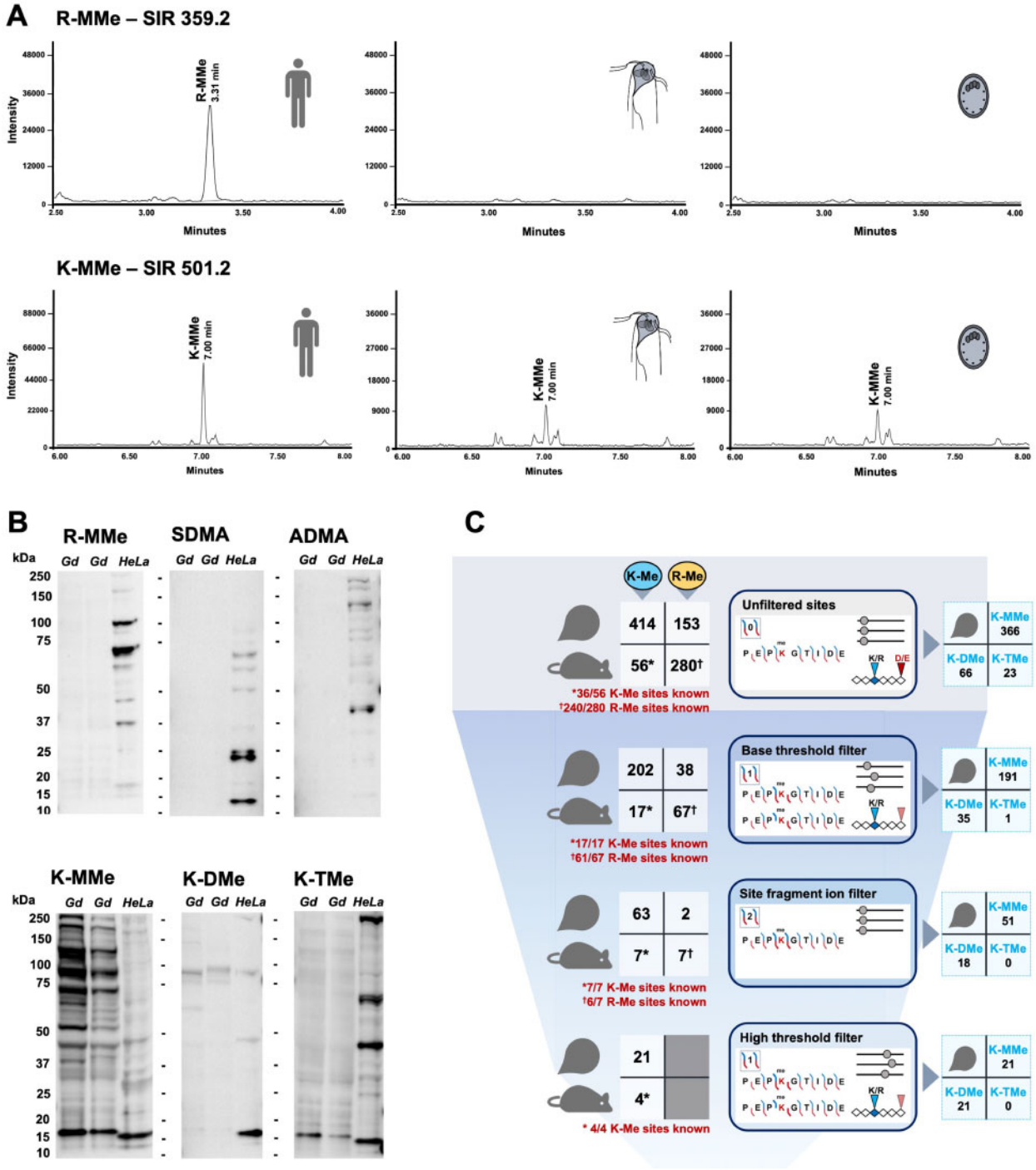


FIG. 5. In vitro analysis of R-Me and K-Me in *Giardia*. (*A*) Ion chromatograms from AAA on a UPLC-QDA system of R-MMe (SIR: 359.2) and K-MMe (SIR: 501.2) in *Giardia* protein lysates (trophozoite and cyst) with HeLa lysate as a positive control. R-MMe peak is detected in HeLa but not *Giardia*, whereas K-MMe peaks are detected in *Giardia* and HeLa lysates. Quantitation of K-MMe and R-MMe can be found in supplementary data 5, Supplementary Material online. (*B*) Immunoblots of R-Me methyl-forms (R-MMe, ADMA, and SDMA) and K-Me methyl-forms (K-MMe, K-DMe, and K-TMe) in total protein lysate from log-phase *Giardia* trophozoites (lane 1), encysting *Giardia* trophozoites (lane 2), and HeLa as a positive control (lane 3). No R-Me modified proteins are detected in *Giardia*, whereas all K-Me (mono-, di-, and tri-) modifications are observed. HeLa lysate contains all six R-Me and K-Me methyl-forms. (*C*) Newly designed MS/MS and site-level filtering for methyl-sites identified from IAP enrichment to reduce FDR. Mouse methyl-sites identified using the same IAP procedure and reagents as a positive control. Filtering strategies (supplementary fig. S15, Supplementary Material online) exclude mis-localizations for aspartic acid (D) and glutamic acid (E) methylation and imposed cutoffs for methyl-site score, peptide length, unique peptide count, and number of methyl states. A filter requiring presence of MS/MS fragment ions adjacent to the K-Me/R-Me site was also used. A total of three levels of filtering are reported, and the number of retained sites in mouse and *Giardia* is shown. Note, the "high-threshold filter" required at least two methyl states per site, therefore no R-MMe sites could be identified, and this filter was specific for K-Me sites. Known mouse R/K-Me sites were taken for reported methyl-sites for mouse, or equivalent sites in humans, based from "PhosphositePlus."

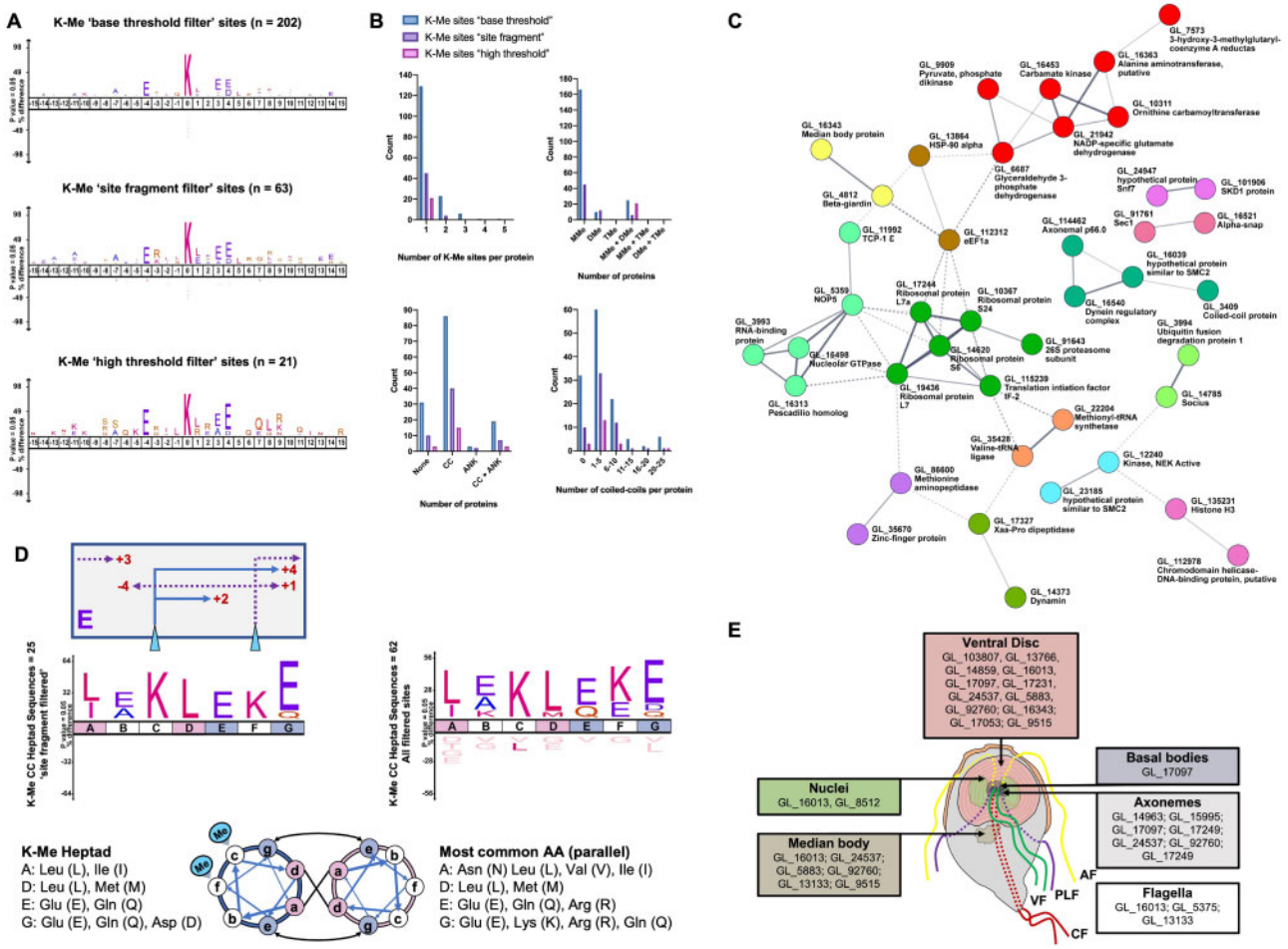


FIG. 6. Analysis of *Giardia* K-Me sites and proteins. (*A*) Sequence logo for K-Me sites at each of the three levels of K-Me site filtering. (*B*) Distribution of K-Me sites per protein, K-Me methyl states, and the number of coiled-coil (CC) and ANK repeats in K-Me proteins as well as the numbers of CCs. (*C*) Visualization of functional connections and relationships between base threshold filtered K-Me proteins mapped using STRING. (*D*) Sequence logo for CC heptads which contain a *Giardia* K-Me site, which accounts for the enrichment of Glu(E) acidic residues in the K-Me sequence motif in (*A*) (above). A parallel dimeric coiled-coil (below) representing the heptad coil sequence and preferred residues in K-Me heptads compared with known residues as well as the preference for K-Me on "c" and "f" solvent exposed residues. (*E*) Subcellular localization of K-Me proteins based of GFP-tagging of cytoskeletal proteins from Hagen et al. (2011).

most NEKs of any organism studied to date (Manning et al. 2011), and most are genus specific with ANK-repeats not present in any other eukaryotic NEKs. When the ~200 *Giardia* NEKs were compared with the 11 human NEKs (Manning et al. 2011), only one had a human ortholog, and 145/198 (73.2%) *Giardia* NEKs had incomplete kinase catalytic triads. We observed complete triads in 8/9 methylated *Giardia* NEKs. Within these, 6/10 K-Me sites occurred within coiled-coils, and none in ANK-repeats. This is despite coiled-coils are less prevalent in *Giardia* NEKs (39/198, 19.7%) than ANK-repeats (130/198, 65.7%). In humans, many NEK-protein interactions involve coiled-coils regions (Meirelles et al. 2014) (e.g., NEK1 [Surpili et al. 2003]), where substrate-binding may facilitate "opening" of coiled-coils/leucine-zippers toward kinase activation and phosphorylation (Croasdale et al. 2011; Meirelles et al. 2014). In this context, K-Me in catalytically active *Giardia* NEKs may regulate phosphorylation through modulating coiled-coil conformations and protein–protein interactions.

NEK kinases, ANK-repeat (21.1), and coiled-coil gene families are hypothesized to regulate the *Giardia* microtubule-based cytoskeleton and its associated structures (Hagen et al. 2011; Manning et al. 2011; Smith et al. 2012; Hennessey et al. 2019). Multiple K-Me proteins in this study localize to the cytoskeleton (Hagen et al. 2011), especially the ventral disk (fig. 6*E*). *Giardia* K-Me proteins included microtubule motor proteins (GL_17265, GL_100906, GL_173333, and GL_8460), seven putative spindle-pole proteins, and tubulin-specific molecular chaperones (GL_1649), including prefoldin subunits (GL_11713, GL_112870) (Millan-Zambrano and Chavez 2014). Many K-Me sites in microtubule-associated proteins occur within coiled-coil heptads, supporting roles for K-Me regulation in the *Giardia* cytoskeleton and cell cycle, potentially through coiled-coil methylation.

Probing Methylation in *Giardia* Using Selective Chemical Inhibitors

Gene knockout is exceedingly difficult in *Giardia*, and gene silencing has variable efficacy (Krtkova and Paredez 2017; Marcial-Quino et al. 2017), therefore we screened a panel of methylation inhibitors to disrupt *Giardia* MTases. Inhibitors were screened in a growth-based, cell viability assay against replicating trophozoites, with complete details of the compound panel and screen detailed in the Supplementary Material online. We observed no antigiardial activity for any compound known to inhibit PRMT enzymes or R-Me function specifically, coinciding with absence of these enzymes and residues in *Giardia* (figs. 4 and 5). However, we observed moderate activity for two Class V KMT-inhibitors, BIX-01294 (IC_{50} = 20.68 μM) and Chaetocin (IC_{50} = 50.31 μM) against replicating trophozoites (supplementary fig. S17, Supplementary Material online). These two inhibitors were then assessed during encystation, as trophozoites were induced in vitro and differentiated into cysts. Both inhibitors were tested during in vitro differentiation at concentrations below their IC_{50} for replicating trophozoites, and the number of cysts counted to assess if the encystation process was interrupted or prevented. Both inhibitors reduced numbers of cysts by 46.1 to 56.5% at 2.5 to 20 μM for BIX-01294, and 37.9 to 71.1% at 10 to 40 μM for Chaetocin as compared with controls (fig. 7A and supplementary table 10, Supplementary Material online) and indicated that these inhibitors were effective against encysting parasites at doses ~5–10-fold lower than their trophozoite IC_{50}.

BIX-01294 and Chaetocin have highest affinity and submicromolar IC_{50} values for HKMTs GLP/G9a and SU(var)39 (Chang et al. 2009) but are also reported to bind and inhibit other mammalian SET-domain protein families with lower specificity and higher IC_{50} values (Cherblanc et al. 2013; Morishita et al. 2017). We propose that sequence and structural divergences in most *Giardia* SET-domains (fig. 1) reduce inhibitor binding affinities, resulting in higher observed trophozoite IC_{50} values. However, increased efficacy for differentiating trophozoites suggests that both inhibitors have improved affinity for GL_9130, which is both encystation regulated (Einarsson et al. 2016; Salusso et al. 2017) and shows greatest SET-domain homology to mammalian HKMTs among *Giardia* Class V K-MTases (fig. 1B).

KMT-Inhibitors Arrest Log-Phase Trophozoite Growth and Reduce Cyst Formation

Giardia Class V K-MTases (besides GL_9130) are transcriptionally enriched during in vitro log-growth, cell stress, or drug treatment (Ansell et al. 2016, 2017). Therefore, we exposed log-phase trophozoites to investigated sublethal inhibitor concentrations and assayed cell viability and adherence at 10 and 24 h (fig. 7B and supplementary table 10, Supplementary Material online). Compared with dimethyl sulfoxide (DMSO) controls at 10 h, there were 38.0% and 50.1% fewer total trophozoites and 51.8% and 19.7%, fewer adhered trophozoites in BIX-01294- and Chaetocin-treated cultures, respectively. Total viable cells remained >90% in both treatments, indicating reduced cell replication, and not cell death, caused the reduced parasite numbers. BIX-01294-treated cultures recovered cell numbers by 24 h. Chaetocin-treated cultures still had 36.5% fewer cell numbers at 24 h, indicating irreversible growth arrest as observed for *Trypanosoma* (Zuma et al. 2017).

Log-phase trophozoite growth does not require epigenetic remodeling, and we observed no significant changes to H3 methylation in trophozoites following KMT-inhibitor treatment (fig. 7C). Rather, KMT-inhibitors disrupted cytoskeletal structures in trophozoites (fig. 7D), particularly the ventrolateral flange which mediates trophozoite adherence (Nosala and Dawson 2015). BIX-01294-treatment affected the trophozoite surface and produced a "rolling" of the ventrolateral flange over the ventral disk. In contrast, Chaetocin-treated trophozoites exhibited severe disruptions in their surface, ventrolateral flange and flagella, and by 10 h accumulated large numbers of intracellular vesicles (supplementary fig. S18, Supplementary Material online).

In contrast, increased efficacy of KMT-inhibitors against encysting *Giardia* agrees with reports of modified chromatin structure (Einarsson et al. 2016) and remodeling of histone H3 methyl-marks during differentiation (Sonda et al. 2010; Carranza et al. 2016), particularly increased H3K9-TMe (Carranza et al. 2016). We observed sublethal treatment of either KMT-inhibitor in encysting trophozoites reduced H3 K-MMe, H3K4-DMe, and H3K9-TMe during early (8 h) encystation (fig. 7C). Further, cysts generated during KMT-inhibitor exposures had less total H3 K-MMe, suggesting other H3 methyl-marks were reduced (supplementary fig. S19, Supplementary Material online). Significant cytoskeletal remodeling is required throughout the differentiation process (Midlej and Benchimol 2009), and KMT-inhibitors also disrupted restructuring of *Giardia* trophozoites during encystation (fig. 7E). The observation that KMT-inhibitors disrupted this process and log-phase replication suggests that they may impact methylation of cytoskeletal proteins containing multiple K-Me target sites (fig. 5), as well as epigenetic regulation occurring during encystation.

Quantitative Proteomics of KMT-Inhibitor Treatment during Growth and Differentiation

We explored differential protein expression (fig. 8A) during sublethal exposures of KMT-inhibitors (BIX-01294 = 10 μM, Chaetocin = 20 μM) during log-phase (growth; TYI) and encystation (differentiation; EC), as compared with DMSO-treated controls. We robustly detected a total of 18,527 peptides from 1,981 proteins in log-phase, and 19,852 peptides from 2,097 proteins in encysting cultures, with cell media (growth vs. encystation) significantly affecting unique peptide detection and necessitating separate quantitative analyses of drug effects (fig. 8A and supplementary fig. S20, Supplementary Material online). Positive correlations for peptide fold changes between inhibitors in the same cell phase (fig. 8B), but negative correlations for the same inhibitor between growth and differentiation (supplementary fig. S21, Supplementary Material online), suggest that inhibitor MTase-targets are determined by cell phase. Upon BIX-

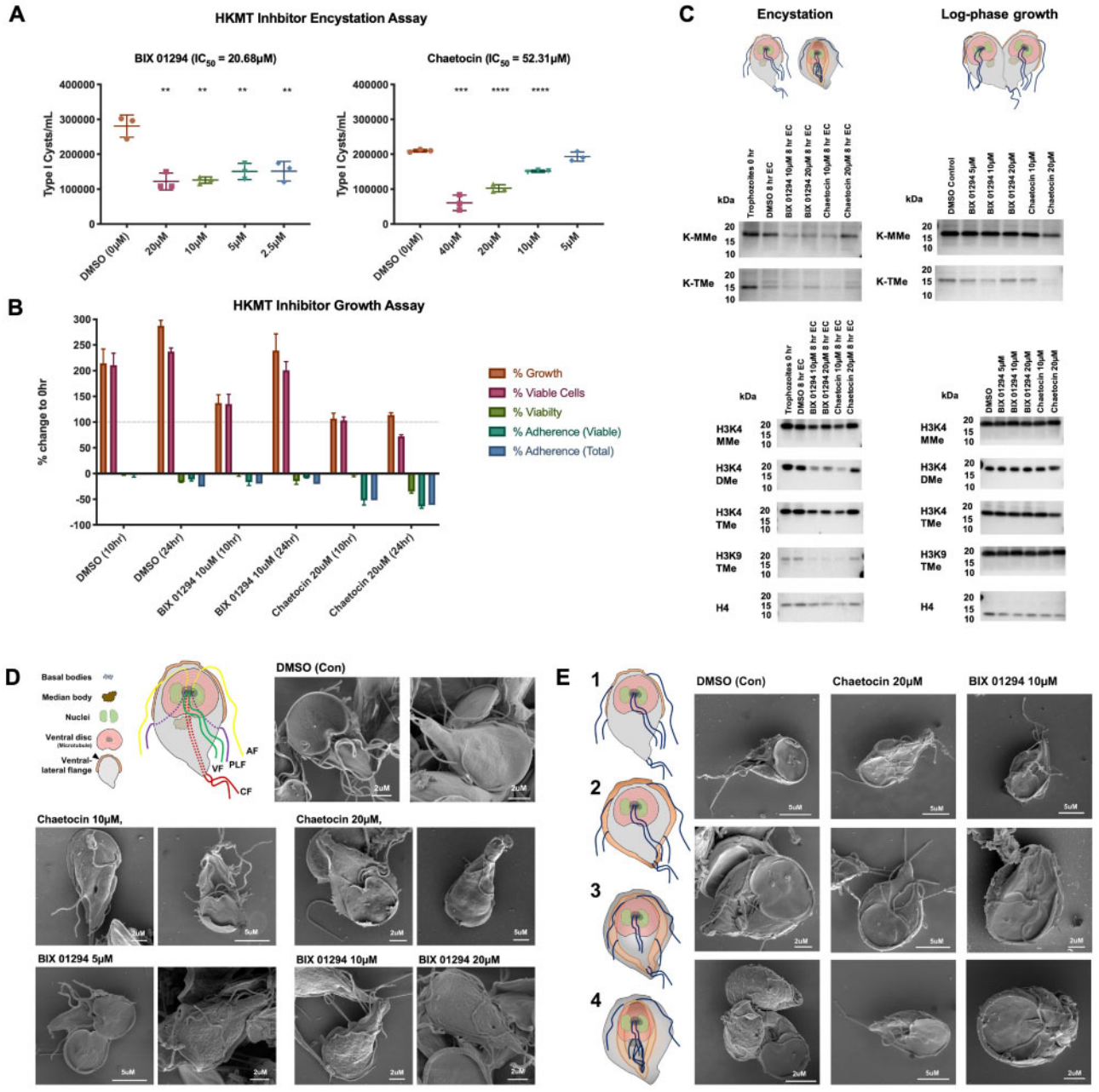


FIG. 7. Activity of KMT-inhibitors during growth and differentiation assays. (*A*) Numbers of Type 1 cysts observed after KMT-inhibitor treatment during in vitro encystation against differentiating trophozoites. Encysting cultures were treated with BIX-01294 and Chaetocin at trophozoite-sublethal concentrations based on their calculated IC_{50} from trophozoite growth assays (BIX-01294 = 20.68 μM, Chaetocin 50.31 μM). BIX-01294 significantly reduced total Type I cysts by 56.5–46.1% of controls (DMSO vehicle) between 20 and 2.5 μM, whereas Chaetocin by 71.1–37.9% between 40 and 10μM. (*B*) Effects of trophozoite-sublethal doses of KMT-inhibitors (BIX-01294 = 10 μM, Chaetocin = 20 μM) on *Giardia* growth, viability, and adherence during log-phase growth. Changes and 10 and 24 h are expressed as percentage of counts taken from triplicate cultures at 0 h. (*C*) Immunoblots of the histone H3 variant for total K-MMe and K-TMe, as well as H3K4-MMe, H3K4-DMe, H3K4-TMe, and H3K9-TMe marks in KMT-inhibitor-treated encysting and log-phase cultures treated. (*D*) SEM images of trophozoites during log-phase growth treated with sublethal concentrations of KMT-inhibitors show cytoskeletal deformities. The overall cytoskeletal structures and morphology of the trophozoite are depicted in the top left. (*E*) SEM images of encysting trophozoite cultures with sublethal concentrations of KMT-inhibitors. Trophozoites undergoing encystation undergo coordinated morphological changes as depicted on the left, including the increase in the ventrolateral flange that folds over the ventral disk and encloses cytoskeletal structures including the flagella. Trophozoites treated with KMT-inhibitors, particularly Chaetocin, cannot coordinate these cytoskeletal rearrangements required for encystation.

01294-treatment, 56 proteins were up- and 36 downregulated in log-phase trophozoites, and 32 up- and 12 downregulated during encystation, relative to the DMSO controls. Chaetocin-treatment induced upregulation of 195 and downregulation of 121 proteins in log-phase trophozoites, with 63 up- and 12 downregulated during encystation.

A

	# peptides	# proteins	Chaetocin				BIX-01294			
			# peptides		# proteins		# peptides		# proteins	
			up	down	up	down	up	down	up	down
Log-phase (growth)	18 527	1981	2902	1394	195	121	230	401	56	30
Encysting (differentiation)	19 852	2097	459	28	63	12	228	43	32	12

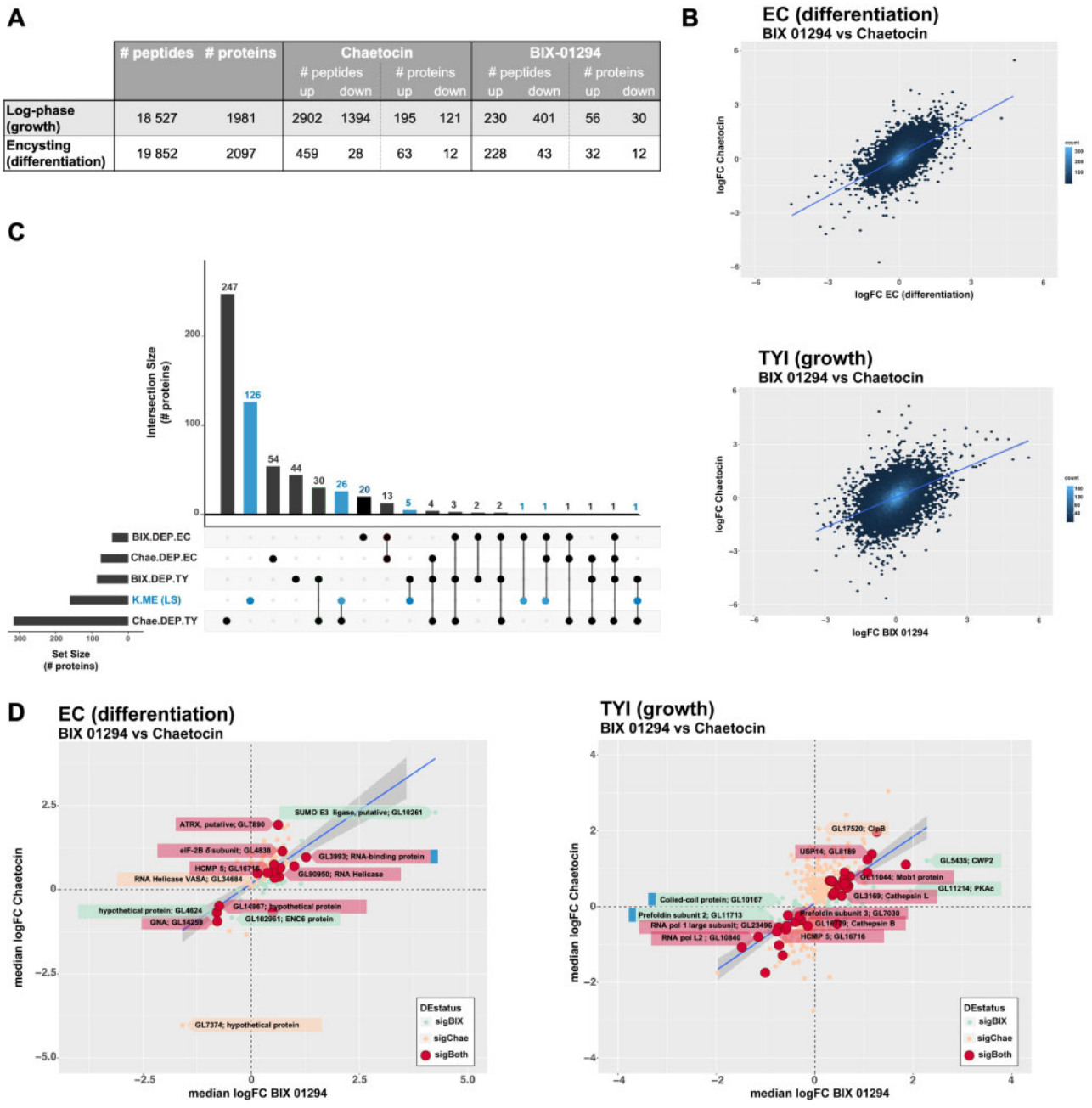


FIG. 8. Label-free quantitative proteomics of KMT-inhibitor-treated cultures during log-phase (growth) and encystation (differentiation). (*A*) Protein and peptide count of KMT-inhibitor exposed trophozoites in log-phase (growth/TYI) and encysting (differentiation/EC) cultures after normalization and imputation. Differentially expressed peptides were identified with *limma* (Ritchie et al. 2015) and the DEPs were identified using peptide-set enrichment analysis (PSEA) with *fry* (Wu et al. 2010; Alhamdoosh et al. 2017) based on BIX-01294 and Chaetocin exposure. (*B*) Correlation between log 2 median protein FC based on BIX-01294 and Chaetocin exposure during encystation ($r^2 = 0.37$, *P* value = 0) and log-phase growth ($r^2 = 0.20$, *P* value = 0) cell phases shows a positive correlation. (*C*) UpSet plot depicting overlap between DEPs for BIX-01294 and Chaetocin treatment in growth (TYI) and encystation (EC) and *Giardia* K-Me proteins identified with base threshold filters (blue). (*D*) Median FC of DEPs of inhibitor for encystation (left) and growth (right). BIX-01294 DEPs are shown in green, Chaetocin DEPs are shown in orange, and shared are depicted in red. Proteins of interest have been annotated, with a blue tag indicating that the DEP is also a K-Me substrate identified in this study.

Overall, 35 (9.5%) differentially expressed proteins (DEPs) were shared between inhibitor treatments during log-phase (fig. 8*C* and supplementary table 11, Supplementary Material online). We observed several consistent functional trends among these DEPs (fig. 8*D*), including differential expression of kinesin and dynein motor proteins, some of which were common (GL_17478, 16480), and others which were unique to Chaetocin- (GL_100906, GL_112846) or BIX-01294-treatment (GL_16945, GL16540). Prefoldin subunits, which are tubulin chaperones (Millan-Zambrano and Chavez 2014) that were also identified among *Giardia* K-Me proteins, were downregulated during inhibitor treatment (GL_7030,

GL_11713), particularly with BIX-01294-treatment. RNA-polymerase subunits were enriched among downregulated proteins during BIX-01294-treatment, including subunits shared with Chaetocin (GL50803_14413, GL50803_10840), and those observed only in BIX-01294-treatment (GL50803_17448, GL50803_23496). Mob1-like protein (GL_11044) was upregulated in both log-phase exposures. Mob1-like protein regulates activity of kinases in mitotic checkpoints for mitotic exits and interacts with spindle-pole bodies (Chow et al. 2010), the latter of which were enriched among *Giardia* K-Me proteins.

In addition to changes observed for both inhibitors, DEPs from BIX-01294-treatment were enriched for gene ontology (GO) terms associated with microtubule motility (fig. 8*D*). We also observed upregulation of protein kinase A catalytic subunit (PKAc, GL_11214), which localizes to basal bodies for flagella locomotion, motility, and excystation (Abel et al. 2001), and of cyst-wall protein 2 (CWP, GL_5435), which is low-abundance in trophozoites during in vitro culture and epigenetically induced during encystation (Sonda et al. 2010). In comparison to BIX-01294, Chaetocin treatment during log-phase resulted in severe cytoskeletal deformities and irreversible growth arrest (fig. 7*B* and *D*), possibly explaining higher numbers of DEPs observed (fig. 8*A*). Gene set changes for upregulated DEPs during Chaetocin treatment implicated protein-folding and protein-trafficking and included thioredoxin-domain proteins (GL_9045, GL_9355), protein disulfide isomerases (GL_9413, GL_103713), Peroxiredoxin 1 (GL_15383, GL_14521), as well as heat shock proteins and DnaJ (GL_15148, GL_88765, GL_16412, GL_98054, and GL13864). We also noted upregulation of ClpB protein (GL_17520), which assists protein-folding under stress conditions and protein disaggregation (Lee et al. 2003). Observed accumulation of intracellular vesicles in Chaetocin-treated trophozoites (supplementary fig. S18, Supplementary Material online) coincides with upregulation of vesicle and exo/endocytosis transport proteins (GL_102108, GL_16773, GL_8329, GL_11953, GL96994, and GL50803_14373). Reduced protein translation was reflected during Chaetocin treatment in 13 downregulated ribosomal structural constituents. When we analyzed DEPs for cell-localization (Hagen et al. 2011), we recorded only two 21.1 protein DEPs in BIX_01294-treatment localizing to the ventrolateral crest (GL_14872, GL_13766) and the abundant, multiply methylated K-Me plasma membrane protein (GL_10167). In contrast, 14 DEPs during Chaetocin treatment localized to wide-ranging cytoskeletal structures. We hypothesize that these correspond to changes in the ventrolateral flange in BIX-012940, and severity of cytoskeletal deformation during Chaetocin treatment (fig. 7*D*).

We observed downregulation of known encystation markers (Einarsson et al. 2016) during inhibitor treatment in encystation (fig. 8*D*). Significantly downregulated proteins included Glucose 6-phosphate *N*-acetyltransferase (GNA; GL_14529) (Einarsson et al. 2016) and ENC6 (GL_102961) (Que et al. 1996) during BIX-01294-treatment, and a late-encystation marker GL_7374 during Chaetocin treatment (Einarsson et al. 2016). This suggests that disrupting K-MTases and methylation dysregulates encystation. RNA and DNA helicases (GL_7890, GL_34684, and GL_90950) were upregulated upon exposure to one or both inhibitors, potentially perturbation gene regulation. GL_3993, which encodes multiple RNA-recognition motifs and a K-Me site, was upregulated in both treatments. Further, multiple *Giardia*-specific High-Cysteine Membrane Proteins (HCMPs) were upregulated exclusively during Chaetocin treatment (supplementary table 11, Supplementary Material online), highlighting potential roles for methylation regulating expression of this complex gene family, noting that histone acetylation regulates VSP-family variant antigen-switching in *Giardia* (Carranza et al. 2016), and histone methylation regulates complex gene families in *Plasmodium falciparum* (Jiang et al. 2013). These data suggest distinct roles for methylation in trophozoites (i.e., cell division and metabolism) compared with cysts (i.e., chromatin regulation), and during cell cycle phases.

In Silico Docking of *Giardia* Class V K-MTases with KMT-Inhibitors

BIX-01294 and Chaetocin inhibit catalytic SET-domains by binding to the active site (histone substrate-competitive) or SAM-binding site (coenzyme-competitive), respectively (Chang et al. 2009; Cherblanc et al. 2013). To assess their targets in *Giardia* and predict their affinity, we modeled drug–ligand interactions with *Giardia* Class V K-MTases.

First, we validated our in silico molecular models through comparisons with interactions of BIX-01294 binding G9a/GLP determined in a published crystal structure (Chang et al. 2009). We calculated binding affinity at −7.2 kcal/mol for G9a/GLP with BIX-01294 and correctly predicted 4/6 known BIX-01294-binding residues determined from the crystal structure (Chang et al. 2009) (fig. 7*A*). We then performed in silico docking between BIX-01294 and predicted structures for four *Giardia* Class V K-MTases with full-length protein structures (GL_9130, GL_17036, GL_221961, and GL13838), with predicted binding affinities of −9.7, −8.7, −7.2, and −5.6 kcal/mol, respectively (supplementary fig. S22, Supplementary Material online). These affinities are consistent with their relative homology to SET-domains of higher eukaryotes (fig. 1*B*), with highest affinity predicted for GL_9130. This correlates with the 5–10-fold increased efficacy of BIX-01294 during encystation (fig. 7*A*), when GL_9130 is transcriptionally upregulated (Einarsson et al. 2016).

Noting this, we undertook detailed modeling of molecular interactions between GL_9130, SAM, and both drug ligands (fig. 9). The closest sequence-structural match for GL_9130 is NSD1, which has structural differences compared with G9a in its lysine channel where BIX-01294 binds (Qiao et al. 2011; Morishita et al. 2017). Therefore, we compared GL_9130 with published molecular models of NSD1 binding BIX-01294 in active- and cofactor-site conformations (Morishita et al. 2017) (fig. 9*B*), which are notably different to G9a molecular models (fig. 9*A*). Our models for GL_9130 predicted high binding affinities for both inhibitors, with predicted binding residues for GL_9130 corresponding to active-site and cofactor-site configurations for BIX-01294 and Chaetocin, respectively, as based on NSD1 (Qiao et al. 2011) (fig. 9*C* and Supplementary

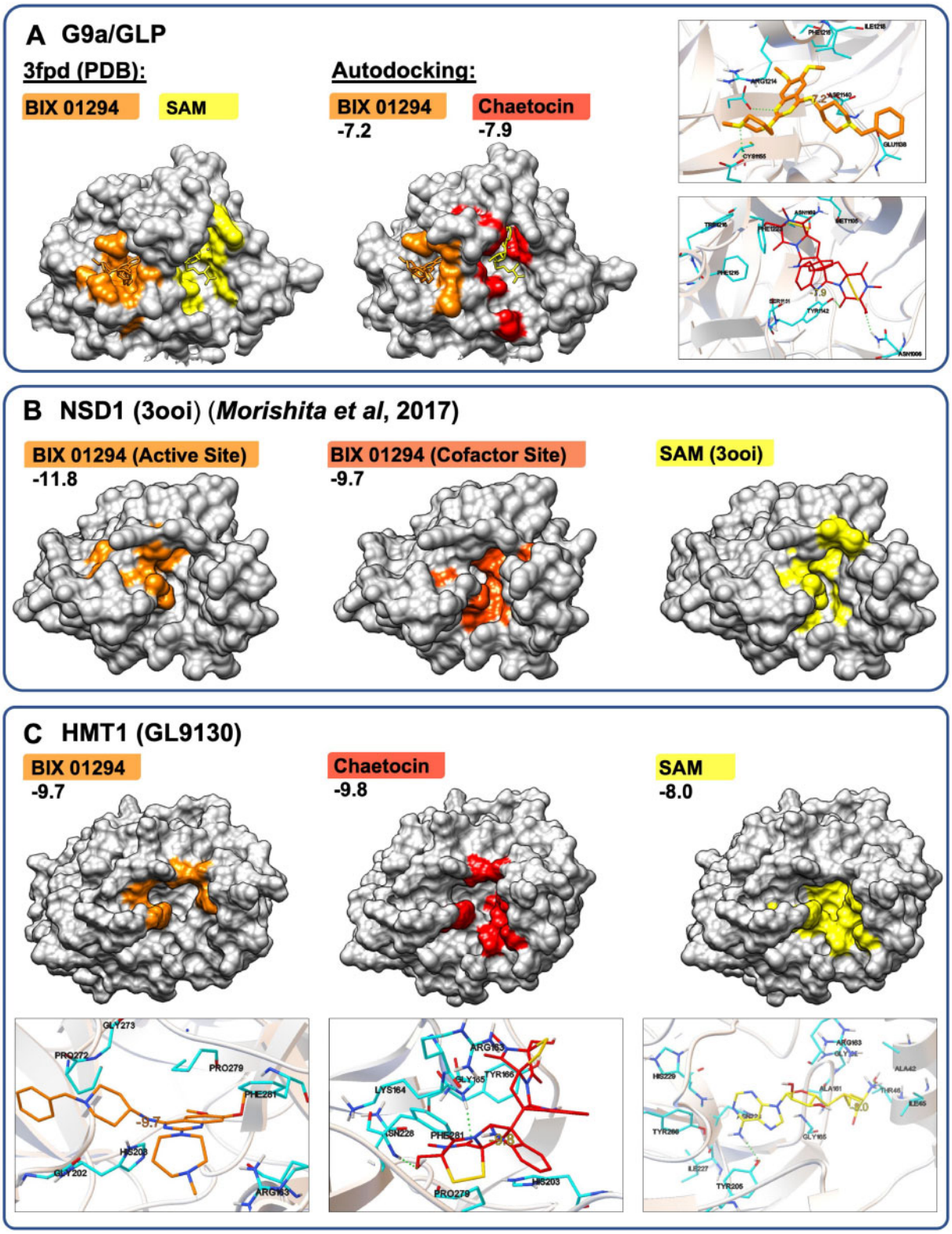


FIG. 9. Autodocking models and molecular details of BIX-01294 and Chaetocin binding to human G9a/GLP and NSD1 and *Giardia* GL9130. (*A*) The crystal structure of G9a/GLP (left) with residues in from cocrystallization binding BIX-01294 (orange) and SAM (yellow) shown is shown on the left and compared with our autodocking models of G9a/GLP binding BIX-01294 (orange) and Chaetocin (red). Predicting binding site and residues in autodocking models overlap with active-site and cofactor-binding sites from the in vitro crystal structure. (*B*) Autodocking models of NSD1 binding BIX-01294 taken from Morishita et al. (2017) show the residues and binding site of BIX-01294 in active- and cofactor-sites. NSD1 is the closest structural homolog for GL9130. The binding site of SAM is taken from residue interactions identified in the NSD1 crystal structure (PDB:3ooi). (*C*) Autodocking models of BIX-01294, Chaetocin, and SAM for GL9130. Binding site and residues are shown for each molecule, which shows that BIX-01294 binds in active-site configuration as compared with NSD1 models, and Chaetocin binds separate residues overlapping with predicted SAM-binding residues in both GL9130 and NSD1.

Material online). These predicted binding sites correspond to modes of action for BIX-01294 (Chang et al. 2009) and Chaetocin (Cherblanc et al. 2013), whereas the predicted high binding affinities support increased efficacy of inhibitors during encystation via GL_9130-targeting.

Discussion

Lysine methylation in *Giardia* is functionally differentiated between life cycle stages, being largely directed toward cytoskeletal, cell cycle and metabolic regulation in trophozoites, and shifting toward chromatin regulation during encystation. The enrichment of coiled-coil proteins among K-Me substrates from trophozoite lysates was striking, particularly K-Me sites in coiled-coils at solvent-facing positions. Coiled-coil peptide domains are molecular spacers (e.g., kinesins in the kinetochore) and can position catalytic and substrate subunits at precise distances (e.g., structural maintenance of chromosome proteins) (Bremang et al. 2013). Coiled-coils can also transduce signals by undertaking conformational changes, as in dynein motor proteins (Kon et al. 2009) and human NEK2 (Croasdale et al. 2011), or facilitate protein–protein interactions, also observed in human NEK2 (Surpili et al. 2003; Meirelles et al. 2014). Coiled-coil K-Me was observed in *Giardia* orthologs of kinesins, motor proteins and structural maintenance of chromosomes, and in *Giardia*-specific gene families linked to cytoskeletal regulation, including lineage-expanded NEKs. In *Chlamydomonas*, axonemal methylarginine regulates flagella length and resorption (Werner-Peterson and Sloboda 2013; Mizuno and Sloboda 2017), with PRMT translocation to flagella catalyzing ADMA within coiled-coils of CCDC65 and CCDC40 in the nexin-dynein regulatory complex. A *Giardia* species-specific K-MTase recognizing coiled-coil heptad sequences (Mason and Arndt 2004) or the coiled-coiled alpha-helical secondary structure is likely and may localize to cytoskeletal structures.

R-Me has a rich role in higher-order eukaryotes (Larsen et al. 2016), with hypothesized expansions and functional gains in RNA metabolism and spliceosomal introns. Deep-branching parasitic protists (*Giardia*, *Trichomonas*, *Spironucleus*, and *Entamoeba*) generally lack introns (Nixon et al. 2002), have lower numbers of RGG/RG motifs, and precede methylarginine-recognition domains. Deep-branching eukaryote species also feature lower numbers of encoded arginine residues in orthologous positions with lower rates of methylation (Larsen et al. 2016), whereas both are expanded in vertebrates alongside gains in PRMT enzymes and families. During these expansions, methylarginine regulation gained essential functions; PRMT1 knockdown is embryonic lethal in mice (Pawlak et al. 2000), whereas knockdown of the orthologous PRMT in yeast, HMT1p, is nonlethal (Chia et al. 2018). We hypothesize that methylarginine regulation emerged in reduced, nonessential roles in deep-branching protists and is absent from, and functionally compensated for, in the Diplomonadida by alternative PTMs. The presence of lysine and methyllysine in *Giardia* in otherwise conserved methylarginine sites and substrates for RNA-binding, processing, and transport proteins, and in the H3 variant, is consistent with this transition. The absence of PRMT and methylarginine in the Diplomonadida is also consistent with their highly compact genomes and functional minimization (Morrison et al. 2007; Leger et al. 2017). This alternative PTM regulation, and its impact on other PTMs warrants further investigation, particularly in light of arginine citrullination in *Giardia* (Touz et al. 2008), which is a reported antagonist of methylarginine in chromatin proteins (Fuhrmann and Thompson 2016).

The absence of methylarginine networks from *Giardia* offers an intriguing model for RNA-binding proteins (RBPome). RBPs are enriched for RGG/RG repeats in genome analyses (Thandapani et al. 2013) and in experimental RNA-interactome captures (Castello et al. 2012), and rely on methylarginine modifications to prevent RNA–protein interactions and directly regulate RNA-binding events (Murn and Shi 2017). Given that motifs in conserved RNA-binding orthologs are remodeled in the Diplomonadida without affecting structurally disordered regions, *Giardia* provides a singular model for studying the RBPome without methylarginine modifications and its sequence-targets, and exploring roles of alternative RNA-binding motifs (Castello et al. 2012; Castello, Fischer, et al. 2016), disordered regions (Calabretta and Richard 2015), and novel regulators in these systems.

The reduced methylproteomes of *Giardia* raise questions regarding what is considered canonical or conserved. Abundant methyllysine in *Giardia* microtubule regulation and its cytoskeleton are species-specific and are key adaptations to the gastrointestinal tract during evolution (Nosala and Dawson 2015). In contrast, methyllysine in protein translation is conserved in deep-branching protists, but additional eEF1a enzymes and sites in yeast (Hamey and Wilkins 2018) highlight their expansion after metamonads. Resolving the chronology of these events will require incorporation of additional basal eukaryotic genomes and experimental data. The comprehensive *Giardia* methylproteome presented here validates protists as models for early eukaryotic cell development, with simpler systems for understanding complex PTM regulation in higher eukaryotes, particularly potential in *Giardia* for methylarginine and RNA-binding proteins interactions (Chong et al. 2018).

Materials and Methods

Bioinformatic Curation of *Giardia* Methyltransferases

Giardia (WB) gene accession numbers are contracted from GL50803 to GL. The *Giardia* WB genome (ATCC 50803) and additional genomes were accessed through GiardiaDB.org, including *Giardia* Assemblage A2 isolate DH (DHA_), Assemblage B isolate GS (GL50581_), Assemblage B isolate GS_B (GSB_), Assemblage E isolate P15 (GLP15_), and the diplomonad *Sp. salmonicida* (SS50377_) genomes.

We employed HMM-based algorithms using HMMER (Version 3.0) for protein MTases in Metamonada species (Finn et al. 2011). To construct HMM models, we curated amino acid sequences of protein methyltransferases in model organisms as described by Petrossian and Clarke (2011). Protein subfamilies included SET-domain MTases, 7βS K-

MTases, and PRMT domains. Protein sequences were downloaded from UniProt (Release 2019_03) and independent multiple sequence alignments were performed for each MTase subfamily using the program MUSCLE. MSAs were used to construct HMM profiles for each subfamily using the hmmbuild algorithm of HMMER3.0, and then to detect domains using the hmmsearch algorithm against complete sets of protein coding sequences from *G. duodenalis* WB genome, *Sp. salmonicida* (Xu et al. 2014), *Trepamonas* PC1 (Xu et al. 2016), *K. bialata* (Tanifuji et al. 2018), and *Monocercomonoides* spp. genomes (Karnkowska et al. 2016). Domains with *E* values $<10^{-3}$ were retained for further analysis. Identified domains were manually inspected, and putative domains confirmed using the domain mapping programs HMMER (v3.2.1), SMART-EMBL, and PROSITE (Release 2019_04).

Further confirmation was obtained by structural protein predictions using the I-TASSER server (Yang et al. 2015) as well as MSAs constructed from MTase sequences of *Giardia* and other representative species using CLUSTALW. MSAs were displayed using the Aline program (version 1.0.025).

Protein structures previously predicted for *Giardia* using I-TASSER (Ansell et al. 2019) were downloaded from Predictein (http://www.predictein.org/giardia_duodenalis, last accessed Febuary 2020). Structures of all other Metamonada proteins or trimmed domains from *Giardia* were predicted using I-TASSER. Closest structural homologs (in the PDB) from these are provided in supplementary data 1, Supplementary Material online.

Class V, SET-Domain K-MTases

SET-domain proteins were identified based on HMMer analyses. Corresponding computationally predicted structures for Class V K-MTases (GL_9130, GL_13838, GL_17036, GL_6407, GL_8921, and GL_221691) were examined for SET-fold architecture. Two predicted full-length structures lacking a SET-domain fold were resubmitted as trimmed, SET-domain sequences (GL_8291, GL_6407) to I-TASSER (Yang et al. 2015). Domain architectures were downloaded from *Giardia*DB.org, compared with UniProt, and independently analyzed via InterProScan (Jones et al. 2014). MSAs were constructed from trimmed SET-domains and compared with representative SET-domain sequences from reviewed HKMTs encoded in *H. sapiens* for conserved cofactor-binding residues, catalytic residues, and SET "pseudoknot" residues defined by Dillon et al. (2005).

To explore eukaryotic Class V K-MTases lineages in *Giardia*, we mined annotated proteomes for a broad representation SET-domain sequences from eukaryotes (including SwissProt reviewed SET-domain proteins from *Arabidopsis thaliana*, *Drosophila melanogaster*, *H. sapiens*, *Neurospora crassa*, and *S. cerevisiae* submitted to the SwissProt database). These were compared with putative SET-domain proteins from *Giardia* Assemblage A isolate WB, additional *Giardia* subspecies/assemblage orthologs and *Sp. salmonicida*. SET-domain-containing proteins were based on three independent domain models (PF00856, Pfam database; PS50280, ProSite database; and SM00317, SMART database) using InterProScan v.5.15.54 (Jones et al. 2014). The SET-domain was extracted from each protein, using annotations based on Pfam domains first, then ProSite domains (if no Pfam domain was found) and lastly, SMART domains. A total of 139 SET-domain proteins were identified and aligned using the program MAFFT v.6.864b (L-INS-i option, using the parameters –localpair and –maxiterate 16; Katoh and Standley 2013). The domain alignment was further refined using the program MUSCLE v.3.7 (-refine option; Edgar 2004) and assessed manually. The MSA of the 139 SET-domains can be found as Supplementary Material online. A phylogenetic tree was constructed from the resulting alignment (393 positions) using the program MrBayes v.3.2.2 (Ronquist et al. 2012), employing the following parameters: prset aamodelpr = mixed; lset rates = invgamma; mcmc ngen = 1,000,000; samplefreq = 100; nchains = 4; sumt relburnin = yes; burninfrac = 0.25; and contype = halfcompat.

Class I, 7βS K-MTases

Novel 7βS K-MTases were identified based on HMMer models. Putative *Giardia* 7βS-K-MTase primary sequence domain architectures were downloaded from *Giardia*DB.org, verified to contain the SAM-dependent methyltransferase annotation (IPR029063) and/or family 16 methyltransferase (IPR019410), whereas computationally predicted structures were assessed for Class I fold, and PBD structural matches to known Class I K-MTases.

Two family 16 putative methyltransferases identified in *G. duodenalis* (GL_100959, DHA_151673) were compared via MSA with their closest PDB structural match VCP-KMT (METTL21D; PDB accession code 4lg1) and additional METTL21D orthologs identified in Kernstock et al. (2012) using ClustalW. Catalytic and cofactor-binding residues were identified based on Kernstock et al. (2012). Orthologs from other *G. duodenalis* subspecies were also included. Three proteins (GL_5013, GL_4349, and GL_3948) with homology to eEF1a-KMT4 (ECE2-KMT; 2pxx) were compared with their subspecies orthologs to the *H. sapiens* Efm4 family eEF1A-K-MTases, including EFMT2, EFMT4, EF-KNMT, and CS-KMT sequences via MSA, and motif regions and putative catalytic and cofactor-binding residues were identified (Jakobsson et al. 2017).

Protein PRMTs

PRMT enzymes were identified based on HMMer analyses. HMM hits for PRMT domain-containing proteins from *K. bialata* (g12581, g9272) and *Monoceromonoides* sp. (MONOS_12731, MONO_2266, MONO11719, and MONOS_2509) were analyzed using I-TASSER structural modeling. PRMT composition for *T. vaginalis* was taken from published homology analyses (Iyer et al. 2008; Fisk and Read 2011).

RGG/RG Motif Analysis in Diplomonads, Protozoan Parasites, and Higher Eukaryotes

Regular expressions of Tri-RGG, Di-RGG, Tri-RG, and Di-RG were taken from Thandapani et al. (2013) and searched via orthoMCL (Chen et al. 2006) using the "Protein Motif

Pattern" search function across 14 genomes. The genome reference and database sources are listed in supplementary table 12, Supplementary Material online. OrthoMCL group identifiers, motif count, and motif location, along with PFAM domain and domain descriptions, were taken for each hit. Gar1 (OG5_128232) and fibrillarin (OG5_127191) ortholog groups were further analyzed of RGG/RG motif arrangement. Orthologs from *Monocercomonoides* sp. (MONO_4374, MONO_1776), *K. bialata* (g259.1, g9459.1), and *Trepomonas* PC1 (TPC1_13494, TPC1_12236) were identified and included. Structural disorder metrics for Gar1 and Fibrillarin sequences were predicted with PrDOS (Ishida and Kinoshita 2007).

Analysis of Conserved Lysine and Arginine Sites in the Histone 3 Variant

MSA of H3.1 variant from 16 species included sequences derived from higher eukaryotes and model organisms, parasitic protists, metamonads, and diplomonads. H3 variant gene ascensions from these species are listed in supplementary table 12, Supplementary Material online. Arginine and lysine residues with >0.75 similarity were considered conserved.

SND1 Orthologs

Ancestral SND1 protein orthologs (OG5_128197) from 16 species including higher eukaryotes, model eukaryotes, and parasitic protists were aligned using CLUSTALW. Gene ascensions for SND1 protein orthologs from these species are listed in supplementary table 12, Supplementary Material online. Aromatic residues implicated in methylarginine binding in the Tudor domain (F740, Y746, Y763, and Y766) and residues involved in hydrogen bonding (N768) were examined (Liu et al. 2010). Protein domain architectures were obtained using OrthoMCL and verified in the UniProt database.

In Vitro Trophozoite Culture and Encystation of *Giardia*

Cultures of *Giardia* were maintained in flat-sided tubes (Nunclon, delta) filled with complete TYI-S33 medium containing 6 mM glucose, 10% heat-inactivated bovine serum, and 0.5 g/l of bovine bile (Luján and Svärd 2011). For experiments necessitating high-protein yield, *Giardia* were cultured upright and stationary in custom borosilicate glass 500 ml "outside-in" roller-bottles filled and capped tightly.

For encystation and cyst production, confluent cultures were chilled on ice and subcultured into pre-encystation, low-bile TYI-S33 (0.1 g bile per liter) and grown to subconfluence (<50%) for 24 h. At 24 h, pre-encystation media was decanted to leave adhered trophozoites, and tubes were filled with encystation medium at pH 7.8 containing 5 g bile per liter. To obtain pure, Type I cysts (Luján and Svärd 2011), sedimented material after 48 h was incubated in dH_2O overnight at 4 °C to lyse nonencysted trophozoites. This was repeated twice, for 72 h total.

AAA of *Giardia* Protein for Methyllysine and Methylarginine

Full methodology for AAA can be found in Supplementary Material online. Briefly, total hydrolysis of the *Giardia* and HeLa protein lysates was performed by gas-phase acid hydrolysis. Hydrolyzed samples, BSA control, and amino acid standards (including standard solutions of R-MMe, K-MMe, ADMA, and SDMA [Sigma-Aldrich]) were precolumn derivatized with AQC reagent (6-aminoquinoyl-*N*-hydroxysuccinimidylcarbamate) using the AccQ-Tag Ultra derivatization kit (Waters Corporation, USA). An internal standard (norvaline) was added to each sample.

Amino acid separation and quantitation was performed on an ACQUITY UPLC system (Waters Corporation) equipped with a TUV (260 nm) or QDa single quadrupole mass detectors. The column was a Waters BEH C18 run at 60 °C and 0.7 ml/min over a 12-min gradient with a 1 μl injection volume. Quantitation by UV of the methylated amino acids was by six-point calibration (0.1–10 pmol/µl, $R^2 >$ 0.999). Quantitation by QDa detector was performed in SIR mode in positive mode (cone voltage 10 V, capillary voltage 0.8 kV) with quantitation of the 359.2-Da ion (R-MMe) and 501.2-Da ion (K-MMe). Calibration was from 20 to 1,000 fmol/µl for K-MMe and 50 to 1,000 fmol/µl for R-MMe ($R^2 = 0.99$). The QDa detector LOD was determined from USP s/n and visual inspection.

IAP of K-Me- and R-MMe-Modified Peptides

IAPs were performed using the PTMScan pan-methyl lysine (Cell Signalling Technologies [CST]) and PTMScan monomethylarginine (CST) according to manufacturer's protocol. Briefly, trophozoites were solubilized in 9 M urea in 20 mM N-2-hydroxyethylpiperazine-N-ethanesulfonic acid (HEPES) (pH 8) containing 1 mM sodium orthovanadate, 2.5 mM sodium pyrophosphate, 1 mM β-glycerolphosphate, and 5 mM Dithiothreitol (DTT) with gentle sonication. This was reduced with DTT for 30 min at 37 °C, followed alkylation with iodoacetamide for 15 min. Protein concentration was verified by bicinchoninic acid (BCA) assay (Pierce), and 10 mg of protein underwent predigestion with Lys-C (Wako) for 3 h at 37 °C at a ratio of 1:100 enzyme to protein, followed by Trypsin (Promega) digestion overnight at 1:100. Digestion was terminated via acidification with trifluoroacetic acid (TFA) to 1%. Peptide purification was performed using a Sep-PAK Classic C18 column (Waters) and peptides dried via lyophilization.

K-Me and R-MMe peptides were enriched via IAP using PTMScan Pan-Methyl IAP beads, or PTMScan Mono-Methylarginine IAP beads (CST), according to manufacturer's instructions. Mouse liver peptides (CST) were used as a positive control for IAPs and performed parallel to *Giardia* pull-downs. After enrichment, samples were desalted using in-house C18 stage tips, dried down in a vacuum centrifuge, and secondary trypsin digestion was performed as recommended by the manufacturer. The sample was acidified, desalted on a C18 tip and vacuum centrifuged to dryness (as above), and peptides dissolved in 0.1% TFA and 2% acetonitrile (ACN) for LC-MS/MS.

MS of K-Me-Enriched IAP Fractions

All LC-MS/MS analyses of IAP-enriched fractions were performed on an Orbitrap Fusion Lumos (Thermo) interfaced with a nano-LC system (Ultimate 3000 RSLC, Dionex) equipped with an Acclaim Pepmap nano-trap column (Dionex—C18, 100 Å, 75 μm × 2 cm) and an Acclaim Pepmap RSLC analytical column (Dionex—C18, 100 Å, 75 μm × 50 cm). From *Giardia*, a total of ten K-Me IAP fractions from independent biological replicates were analyzed with duplicate technical injections ($n = 9$ trophozoite, $n = 1$ cyst), along with four R-MMe IAP fractions from independent biological replicates ($n = 4$ trophozoite). A total of two K-Me and two R-MMe IAP fractions from control mouse liver peptides (CST) were also analyzed. K-Me IAP fractions were analyzed on a 210-min gradient ranging 3–90% ACN in 0.1% formic acid (FA), whereas R-Me IAP fractions were analyzed on a 60-min gradient ranging 3–90% ACN in 0.1% FA. Eluting peptides were ionized using positive ion mode nano-ESI with 120,000 resolution scanning from 350 to 1,550 *m*/*z* in full scan with lockmass of 445.12003 (K-Me samples) and 401.92272 (R-Me samples). Data-dependent MS2 was carried out using the 3-s maximum cycle method. MS2 was carried out in HCD mode with resolution of 15,000, AGC target of 40,000 (R-Me samples) and 50,000 (K-Me samples) and activation Q of 0.25 for ions above 50,000. Dynamic exclusion was enabled (exclusion duration = 45 s).

Database Searching and K-Me Site Mapping and Filtering

All raw data analyses of K-Me and R-MMe IAP fractions were performed with Maxquant (Tyanova et al. 2016) software (v1.6.3.4). Search parameters included trypsin digestion allowing maximum of three missed cleavages, minimum peptide length of seven amino acids, carbamidomethylation as a fixed modification, and protein N-terminal acetylation, methionine oxidation as variable modifications. For K-Me IAP samples, searches included Methyl(K), DiMethyl(K), TriMethyl(K), and Methyl(D/E) as variable modifications, and R-MMe IAP samples Methyl(K/R) and Methyl(D/E). Peptide and protein FDR of 0.01 were set, based on number of reverse hits. For *Giardia* methyl searches, sequences from the *G. duodenalis* WB isolate genome (*Giardia*DB.org, ATCC 50803 version 43) were used with sequences containing common contaminants, and human sequences from the UniProt database (downloaded May 2019) for further exclusion of media- and serum-based K-Me contaminants. Raw data from control mouse IAP fractions were searched as above against sequences from the UniProt Mouse fasta database (downloaded May 2019). The complete raw files and search results can be accessed via the ProteomeXchange Consortium via the PRIDE partner repository with the data set identifier PXD016813.

MaxQuant output tables ("msms.txt," "peptides.txt," and site tables—e.g., "Methyl (K)Sites.txt," "Dimethyl (K)Sites.txt," and "Trimethyl (K)Sites.txt") were processed and filtered by means of a custom R script (available at https://github.com/JoshuaHamey/MethylSite-Filtering), with full details of confidence filtering strategies found in the Supplementary Material online. Sequence logo and motif analyses were performed using the iceLOGO web server (Colaert et al. 2009). Analyses of significantly enriched GO Terms, protein domains, and KEGG were performed using DAVID (Huang et al. 2007). Protein–protein interactions were visualized using STRING (Szklarczyk et al. 2019), as well as significant functional enrichments. Predicted coiled-coil and ANK repeat regions were mapped via UniProt annotations (downloaded June 2019) and *Giardia* heptads mapped using COILS (Lupas et al. 1991).

Parallel Reaction Monitoring Analysis of eEF1A N-Terminal Methylation

Whole protein lysate from midlog-phase *Giardia* trophozoites was separated by sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) as previously described (Emery et al. 2018) and the gel band corresponding to eEF1A (~50 kDa) was excised and digested with Asp-N. The N-terminal Asp-N peptide of *Giardia* eEF1A (GKEKKHINLVVIGHV) was then analyzed by LC-parallel reaction monitoring according to previous methods (Hamey, Winter, et al. 2016), except that the analysis was performed on an Orbitrap Fusion Lumos Tribrid mass spectrometer (Thermo Fisher Scientific). Each cycle (~0.4 s duration) consisted of one precursor scan, acquired in the Orbitrap (300–1,500 *m*/*z*, resolution = 60,000), followed by three MS2 scans, acquired in the Orbitrap (HCD fragmentation at NCE = 30, 150–1,750 *m*/*z*, resolution = 30,000), targeting precursors at *m*/*z* 857.03, 571.69, and 429.02, which corresponds to the trimethylated GKEKKHINLVVIGHV peptide in its doubly, triply, and quadruply charged states. Extracted ion chromatograms of transitions from the trimethylated peptide to the y14 ion in its unmethylated or trimethylated state were obtained using Qual Browser in Thermo Xcalibur 2.2 SP1.48.

In Vitro Viability Assays of Methylation Inhibitors against *Giardia* Trophozoites

The Epigenetic Inhibitor library from SelleckChem (Compounds Australia) was screened against replicating trophozoites at 10 μM in a miniaturized 384-well ATP luminescence based Cell Titre Glo (Promega) assay modified from previous methods (Ansell et al. 2017). Single-concentration screens were first performed in duplicate, with compounds with kill or reproducible partial inhibition activity taken for dose–response estimations from 30 μM in 2-fold, ten-point dilution series. All compounds were diluted in DMSO, with negative controls containing equivalent DMSO concentration, as well as positive "kill" control compounds of metronidazole (20 μM) and albendazole (2 μM).

Several methylation inhibitors were screened at higher concentrations in trophozoites in medium-throughput 94-well ATP luminescence based Cell Titre Glo assays as previously described (Ansell et al. 2017). These included BIX-01294, AMI-1, 5′-deoxy-5′(methylthio)adenosine, S-(5adenosyl)-L-homocysteine, adenosine-2′,3′-dialdehyde, pargyline, and trans-2-phenylcyclopropylamine obtained from Sigma (USA), as well as Chaetocin and Sinefungin obtained from BioAustralis (Australia). All solutions were solubilized in DMSO, and stocks used fresh. For KMT and PRMT inhibitors,

trophozoite sensitivity was assessed up to 100 μM concentrations, whereas for histone lysine demethylase inhibitors compounds were tested up to 1,000 μM concentration. All assays were carried out in triplicate, and active compounds repeated in 2-fold series dilution to calculate IC_{50} (Ansell et al. 2017).

In Vitro KMT Drug Treatments for Log-Phase Cultures

Log-phase trophozoite cultures were seeded prior to exposure to achieve midconfluence, log-phase growth within 24 h. Trophozoite-sublethal exposures were calculated as half the IC_{50} against replicating trophozoites (Chaetocin = 20 μM; BIX 01294 = 10 μM), with control tubes seeded with equivalent DMSO to 0.001%. During growth counts, nonadhered trophozoites were counted by subsampling the media column and total trophozoites (adhered and nonadhered parasites) counted after icing tubes to detach adhered trophozoites. A baseline for trophozoite growth and adherence was established by triplicate counts at hour 0. Trophozoite viability was via Trypan Blue exclusion (Sigma).

For quantitative proteomics and growth counts, log-phase trophozoites were exposed to sublethal drug treatments (Chaetocin = 20 μM; BIX 01294 = 10 μM) and compared with a DMSO control at two time points (8 and 24 h). Trophozoites were harvested for quantitative proteomics as previously described (Emery-Corbin et al. 2018). For K-Me immunoblotting, log-phase trophozoites were exposed to multiple sublethal concentrations (Chaetocin = 10 μM, 20 μM; BIX 01294 = 5 μM, 10 μM, and 20 μM) as compared with a DMSO control for 15 h, and then detached and pelleted via centrifugation for protein extraction. For quantitative proteomics and immunoblotting, total adhered and nonadhered trophozoite material was used.

In Vitro KMT Drug Treatments for Encysting Cultures

Encystation was induced as described above, with KMT-inhibitors (or DMSO vehicle only) added into encystation media after the low-bile priming. For quantitative proteomics, trophozoites were exposed to sublethal drug treatments (Chaetocin = 20 μM; BIX 01294 = 10 μM) and compared with DMSO controls at two timepoints (8 and 24 h) in triplicate. For immunoblotting, encysting cultures were exposed to multiple trophozoite-sublethal concentrations (Chaetocin = 10 μM, 20 μM; BIX 01294 = 10 μM, 20 μM) for 8 h, and protein extracted for immunoblotting. To observe the effects of KMT-inhibitors on cyst formation, encystation in high-bile media was carried out for 48 h, and Type I cysts enriched by hypotonic lysis and counted after sedimentation. This material was also harvested for protein extraction for immunoblotting by combining triplicates.

Immunoblotting of K-Me

Anti-K-MMe, Anti-K-DMe, and Anti-K-TMe, as well as Anti-R-MMe and Anti-R-DMe (asymmetric and symmetric) were obtained from CST, whereas Anti-H3K4Mme, Anti-H3K4DMe, Anti-H3K4TMe, and Anti-H3K9TMe were obtained from Abcam as per Carranza et al. (2016). Where necessary, HeLa cell protein lysate was used as a positive control. Proteins were extracted and immunoblotted as previously described (Emery et al. 2018), at 1:1,000 for anti-K/R-Me antibodies, and 1:500 for Abcam anti-H3 methyl mark antibodies. Consistent protein loading (5 μg) was estimated using BCA assay (Pierce), verified after transfer using Ponceau S staining (Sigma), and anti-H4 (Abcam) used as a loading control. Protein–antibody interaction was detected with a horseradish peroxidase-conjugated immunoglobulin G secondary antibody via chemiluminescent reagent (LumiGLO, CST) on a BioRad ChemiDoc MP imaging system. Exposure times were optimized for each antibody; images were collected up to 180 s based on resolution and absence of oversaturation.

Scanning Electron Microscopy of KMT-Inhibitor-Treated Trophozoites

For morphological analysis, log-phase trophozoites were exposed to multiple sublethal concentrations (Chaetocin = 10 μM, 20 μM; BIX 01294 = 5 μM, 10 μM, and 20 μM) as compared with a DMSO control for up to 15 h prior to fixation. For encystation, encysting trophozoites were exposed to sublethal concentrations of KMT-inhibitors sublethal drug treatments (Chaetocin = 20 μM; BIX 01294 = 10 μM) for 8 and 24 h. Trophozoites were fixed in 2.5% glutaraldehyde in 1× phosphate buffered saline (PBS). Following fixation, samples were washed in 1× PBS, mounted on 10-mm round coverslips using poly-L-lysine, dehydrated through a series of ethanol solutions (30%, 50%, 70%, 90%, 100%, and 100%), and dried using a critical point dryer. Coverslips were mounted on stubs using adhesive carbon and coated in gold (Au) using a Leica gold coater. Trophozoites were imaged at 3 kV using the in-lens secondary electron detector on a Tescan MIRA3 field emission SEM.

Quantitative Proteomics of KMT-Inhibitor-Treated *Giardia*

Protein extraction was performed in 2.5% sodium dodecyl sulfate in 50 mM Tris (pH 8.8) (Sigma-Aldrich). Proteins were reduced with 10 mM dithiothreitol at 37 °C followed by alkylation in 15 mM iodoacetamide (with alkylation quenched with 5 mM dithiothreitol). Proteins were precipitated via methanol/chloroform, transferred to 8 M urea in 50 mM Tris (pH 8.8) and concentration quantitated by BCA assay (Pierce), followed by trypsin digestion of 20 μg protein overnight at 37 °C at a ratio of 1:100 enzyme to protein. Samples were acidified with TFA and peptides desalted using C18 solid-phase extraction stage tips, dried using a vacuum centrifuge, and reconstituted in 2% ACN, 0.1% TFA for LC-MS/MS.

LC-MS/MS was performed on an LTQ Orbitrap Elite (Thermo Scientific) with a nano-ESI interface in conjunction with an Ultimate 3000 RSLC nanoHPLC (Dionex Ultimate 3000) equipped with an Acclaim Pepmap nano-trap column (Dionex-C18, 100 Å, 75 μm × 2 cm) and an Acclaim Pepmap RSLC analytical column (Dinoex-C18, 100 Å, 75 μm × 15 cm). Peptides were analyzed on a 125-min flow gradient ranging 3–80% ACN in 0.1% FA as previously described (Emery-Corbin et al. 2018). The LTQ Orbitrap Elite spectrometer

was operated in the data-dependent mode and all spectra acquired as previously described (Emery-Corbin et al. 2018), with fragmentation of the 20 most intense precursors obtained using rapid collision-induced dissociation with normalized collision energy of 30 and activation q of 0.25.

Database Searching and Bioinformatic Analyses of KMT-Inhibitor-Treated *Giardia*

Maxquant (Tyanova et al. 2016) software was used for database searching and label-free quantitation, with the LFQ minimum ratio count set to 1, and matching between runs set to "match from and to." Log-phase trophozoites and encysting trophozoite raw data searches were performed separately. *Giardia duodenalis* WB isolate genome (*Giardia*DB.org, ATCC 50803 version 37) was used along with sequences containing common contaminants, with parameters of trypsin digestion with maximum two missed cleavages, carbamidomethylation as a fixed modification, protein N-terminal acetylation and methionine oxidation as variable modifications, peptide length of minimum seven amino acids, and peptide and protein false discovery rate set to 0.01. The search parameters, raw files, and search results can be accessed via the ProteomeXchange Consortium via the PRIDE partner repository with the data set identifier PXD016747.

Peptide intensities were summed, and reverse peptide decoys and contaminants removed. Due to large discrepancies in detected peptides, data from experiments in EC and TYI media were processed separately as follows. Unique peptides detected in at least 8 of 18 samples were log-transformed and quartile-normalized using limma software (Ritchie et al. 2015), and missing values were imputed using a method based on low-rank decomposition (available at https://github.com/DavisLaboratory/msImpute). Up to five surrogate variables were removed to account for unwanted variation, and differentially expressed peptides between chemical treatments were identified using linear models with empirical Bayes moderated *t*-statistics. DEPs were estimated using peptide-set enrichment analysis in a GSEA framework (Wu et al. 2010; Alhamdoosh et al. 2017).

In Silico Autodocking of KMT-Inhibitors with *Giardia* SET-Domain-Containing Proteins

In silico docking of small molecules within predicted protein structures was performed using AutoDockTools (ADT), v1.5.6. BIX-01294 docking was simulated within complete *Giardia* SET-domain-containing protein structures (GL9130, GL221691, GL17036, and GL17036), but not for GL8921 and GL6407 which lacked predicted SET-domains. BIX-01294 interactions were also simulated for *H. sapiens* G9a (PDB:3fpd) and compared with its crystal structures crystallized with BIX-01294 and SAM (Chang et al. 2009). Additionally, *H. sapiens* G9a protein and *Giardia* GL_9130 were docked with Chaetocin, and GL_9130 docked with SAM. 3D structures of *Giardia* SET-domain proteins were aligned with the SET-domain of G9a (residues 913–1193 Or 1126–1243) to locate active sites, and then BIX-01294, SAM, and Chaetocin molecules and predicted active sites were used for docking studies. ADT, version 1.5.6 was used to prepare proteins and the BIX-01294 molecules coordinate files separately by adding polar hydrogen molecules, partial charges, and assigning atom types. To build a torsion tree in ADT for BIX-01296, SAM, and Chaetocin in ADT, a root atom was chosen for flexibility, rotatable bonds were defined, and the grid box was constructed around active-site region of respective proteins. Then, docking was performed using ADT-vina version 2 and ADT was used to analyze and plot interactions between protein residues and inhibitor molecules.

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

This work, including the efforts of S.J.E.-C., was funded by a Jack Brockhoff Foundation Early Career (Grant No. JBF 4184, 2016). S.J.E.-C., B.R.E.A., B.B., S.T., S.H.Z., M.B., M.J.D., and A.R.J. are supported by the Victorian State Government Operational Infrastructure Support and Australian Government National Health and Medical Research Council Independent Research Institute Infrastructure Support Scheme. A.R.J. is also supported by an NHMRC Career Development Fellowship (APP1126395). M.B. is supported by an NHMRC Senior Research Fellowship (1102971). B.R.E.A. is supported by an Early Career NHMRC Fellowship (APP1157776). The authors acknowledge the facilities and the scientific and technical assistance supported by Microscopy Australia (MA) at the Central Analytical Research Facility (CARF), Institute for Future Environments, Queensland University of Technology, Brisbane, Australia. The authors also acknowledge the Melbourne Mass Spectrometry and Proteomics Facility of the Bio21 Molecular Science and Biotechnology Institute at the University of Melbourne, including Shuai Nie, Ching-seng Ang, and Nick Williamson, for the support of MS analysis. The authors also acknowledge the NCRIS-enabled facilities of the Australian Proteome Analysis Facility (APAF) of Macquarie University, Sydney. The authors acknowledge the facilities and staff of the Bioanalytical Mass Spectrometry Facility at the University of New South Wales.

Author Contributions

S.J.E.-C. was responsible for conceptualization, investigations, and methodology and was responsible for funding acquisition. S.J.E.-C. and A.R.J. were responsible for project administration. B.B., B.R.E.A., A.J.S., S.T., C.C., B.V.M., J.J.H., and S.G.S. contributed to specific methodology, investigations, and formal analyses within. M.J.D., M.B., S.H.Z., and B.R.E.A. contributed to formal analyses for quantitative proteomics, and S.H.Z. provided software for these analyses. J.J.H. provided software and performed formal analyses for methylation filtering, with contributions by M.R.W. E.L., D.V., and A.C. provided resources and methodology for drug-screening. S.J.E.-C. was responsible for the writing of the original draft and preparation, with contributions from A.R.J., J.J.H., M.R.W., B.R.E.A., and R.B.G.. S.J.E.-C., R.B.G., and A.R.J. were responsible for the

writing review and editing, with additional contributions for this from all authors.

References

Abel ES, Davids BJ, Robles LD, Loflin CE, Gillin FD, Chakrabarti R. 2001. Possible roles of protein kinase A in cell motility and excystation of the early diverging eukaryote *Giardia lamblia*. *J Biol Chem*. 276(13):10320–10329.

Adam RD, Dahlstrom EW, Martens CA, Bruno DP, Barbian KD, Ricklefs SM, Hernandez MM, Narla NP, Patel RB, Porcella SF, et al. 2013. Genome sequencing of *Giardia lamblia* genotypes A2 and B isolates (DH and GS) and comparative analysis with the genomes of genotypes A1 and E (WB and pig). *Genome Biol Evol*. 5(12):2498–2511.

Alhamdoosh M, Law CW, Tian L, Sheridan JM, Ng M, Ritchie ME. 2017. Easy and efficient ensemble gene set testing with EGSEA. *F1000Research* 6:2010.

Ansell BR, Baker L, Emery SJ, McConville MJ, Svard SG, Gasser RB, Jex AR. 2017. Transcriptomics indicates active and passive metronidazole resistance mechanisms in three seminal *Giardia* lines. *Front Microbiol*. 8:398.

Ansell BR, McConville MJ, Baker L, Korhonen PK, Emery SJ, Svard SG, Gasser RB, Jex AR. 2016. Divergent transcriptional responses to physiological and xenobiotic stress in *Giardia duodenalis*. *Antimicrob Agents Chemother*. 60(10):6034–6045.

Ansell BRE, Pope BJ, Georgeson P, Emery-Corbin SJ, Jex AR. 2019. Annotation of the *Giardia* proteome through structure-based homology and machine learning. *GigaScience* 8(1):giy150.

Bachand F. 2007. Protein arginine methyltransferases: from unicellular eukaryotes to humans. *Eukaryotic Cell* 6(6):889–898.

Blanc RS, Richard S. 2017. Arginine methylation: the coming of age. *Mol Cell* 65(1):8–24.

Bremang M, Cuomo A, Agresta AM, Stugiewicz M, Spadotto V, Bonaldi T. 2013. Mass spectrometry-based identification and characterisation of lysine and arginine methylation in the human proteome. *Mol Biosyst*. 9(9):2231–2247.

Calabretta S, Richard S. 2015. Emerging roles of disordered sequences in RNA-binding proteins. *Trends Biochem Sci*. 40(11):662–672.

Calpena E, Palau F, Espinos C, Galindo MI. 2015. Evolutionary history of the Smyd gene family in metazoans: a framework to identify the orthologs of human Smyd genes in *Drosophila* and other animal species. *PLoS One* 10(7):e0134106.

Carranza PG, Gargantini PR, Prucca CG, Torri A, Saura A, Svard S, Lujan HD. 2016. Specific histone modifications play critical roles in the control of encystation and antigenic variation in the early-branching eukaryote *Giardia lamblia*. *Int J Biochem Cell Biol*. 81:32–43.

Castello A, Fischer B, Eichelbaum K, Horos R, Beckmann BM, Strein C, Davey NE, Humphreys DT, Preiss T, Steinmetz LM, et al. 2012. Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell* 149(6):1393–1406.

Castello A, Fischer B, Frese CK, Horos R, Alleaume AM, Foehr S, Curk T, Krijgsveld J, Hentze MW. 2016. Comprehensive identification of RNA-binding domains in human cells. *Mol Cell* 63(4):696–710.

Castello A, Horos R, Strein C, Fischer B, Eichelbaum K, Steinmetz LM, Krijgsveld J, Hentze MW. 2016. Comprehensive identification of RNA-binding proteins by RNA interactome capture. *Methods Mol Biol*. 1358:131–139.

Caudy AA, Ketting RF, Hammond SM, Denli AM, Bathoorn AM, Tops BB, Silva JM, Myers MM, Hannon GJ, Plasterk RH. 2003. A micrococcal nuclease homologue in RNAi effector complexes. *Nature* 425(6956):411–414.

Chang Y, Zhang X, Horton JR, Upadhyay AK, Spannhoff A, Liu J, Snyder JP, Bedford MT, Cheng X. 2009. Structural basis for G9a-like protein lysine methyltransferase inhibition by BIX-01294. *Nat Struct Mol Biol*. 16(3):312–317.

Chen C, Nott TJ, Jin J, Pawson T. 2011. Deciphering arginine methylation: Tudor tells the tale. *Nat Rev Mol Cell Biol*. 12(10):629–642.

Chen F, Mackey AJ, Stoeckert CJ Jr, Roos DS. 2006. OrthoMCL-DB: querying a comprehensive multi-species collection of ortholog groups. *Nucleic Acids Res*. 34(90001):D363–D368.

Cherblanc FL, Chapman KL, Brown R, Fuchter MJ. 2013. Chaetocin is a nonspecific inhibitor of histone lysine methyltransferases. *Nat Chem Biol*. 9(3):136–137.

Chia SZ, Lai YW, Yagoub D, Lev S, Hamey JJ, Pang CNI, Desmarini D, Chen Z, Djordjevic JT, Erce MA, et al. 2018. Knockout of the Hmt1p arginine methyltransferase in *Saccharomyces cerevisiae* leads to the dysregulation of phosphate-associated genes and processes. *Mol Cell Proteomics* 17(12):2462–2479.

Chong PA, Vernon RM, Forman-Kay JD. 2018. RGG/RG motif regions in RNA binding and phase separation. *J Mol Biol*. 430(23):4650–4665.

Chow A, Hao Y, Yang X. 2010. Molecular characterization of human homologs of yeast MOB1. *Int J Cancer* 126(9):2079–2089.

Cloutier P, Lavallee-Adam M, Faubert D, Blanchette M, Coulombe B. 2013. A newly uncovered group of distantly related lysine methyltransferases preferentially interact with molecular chaperones to regulate their activity. *PLoS Genet*. 9(1):e1003210.

Colaert N, Helsens K, Martens L, Vandekerckhove J, Gevaert K. 2009. Improved visualization of protein consensus sequences by iceLogo. *Nat Methods*. 6(11):786–787.

Cornett EM, Ferry L, Defossez PA, Rothbart SB. 2019. Lysine methylation regulators moonlighting outside the epigenome. *Mol Cell* 75(6):1092–1101.

Croasdale R, Ivins FJ, Muskett F, Daviter T, Scott DJ, Hardy T, Smerdon SJ, Fry AM, Pfuhl M. 2011. An undecided coiled coil: the leucine zipper of Nek2 kinase exhibits atypical conformational exchange dynamics. *J Biol Chem*. 286(31):27537–27547.

Dillon SC, Zhang X, Trievel RC, Cheng X. 2005. The SET-domain protein superfamily: protein lysine methyltransferases. *Genome Biol*. 6(8):227.

Dzialo MC, Travaglini KJ, Shen S, Loo JA, Clarke SG. 2014. A new type of protein lysine methyltransferase trimethylates Lys-79 of elongation factor 1A. *Biochem Biophys Res Commun*. 455(3–4):382–389.

Edgar RC. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinf*. 5:113.

Einarsson E, Troell K, Hoeppner MP, Grabherr M, Ribacke U, Svard SG. 2016. Coordinated changes in gene expression throughout encystation of *Giardia intestinalis*. *PLoS Negl Trop Dis*. 10(3):e0004571.

Emery SJ, Baker L, Ansell BRE, Mirzaei M, Haynes PA, McConville MJ, Svard SG, Jex AR. 2018. Differential protein expression and post-translational modifications in metronidazole-resistant *Giardia duodenalis*. *GigaScience* 7(4):giy024.

Emery-Corbin SJ, Vuong D, Lacey E, Svard SG, Ansell BRE, Jex AR. 2018. Proteomic diversity in a prevalent human-infective *Giardia duodenalis* sub-species. *Int J Parasitol*. 48(11):817–823.

Erce MA, Pang CN, Hart-Smith G, Wilkins MR. 2012. The methylproteome and the intracellular methylation network. *Proteomics* 12(4–5):564–586.

Finn RD, Clements J, Eddy SR. 2011. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res*. 39(Suppl):W29–W37.

Fisk JC, Read LK. 2011. Protein arginine methylation in parasitic protozoa. *Eukaryotic Cell* 10(8):1013–1022.

Fuhrmann J, Thompson PR. 2016. Protein arginine methylation and citrullination in epigenetic regulation. *ACS Chem Biol*. 11(3):654–668.

Guccione E, Richard S. 2019. The regulation, functions and clinical relevance of arginine methylation. *Nat Rev Mol Cell Biol*. 20(10):642–657.

Guo A, Gu H, Zhou J, Mulhern D, Wang Y, Lee KA, Yang V, Aguiar M, Kornhauser J, Jia X, et al. 2014. Immunoaffinity enrichment and mass spectrometry analysis of protein methylation. *Mol Cell Proteomics* 13(1):372–387.

Hagen KD, Hirakawa MP, House SA, Schwartz CL, Pham JK, Cipriano MJ, De La Torre MJ, Sek AC, Du G, Forsythe BM, et al. 2011. Novel structural components of the ventral disc and lateral crest in *Giardia intestinalis*. *PLoS Negl Trop Dis*. 5(12):e1442.

Hamey JJ, Hart-Smith G, Erce MA, Wilkins MR. 2016. The activity of a yeast Family 16 methyltransferase, Efm2, is affected by a conserved

tryptophan and its N-terminal region. *FEBS Open Bio* 6(12):1320–1330.
Hamey JJ, Separovich RJ, Wilkins MR. 2018. MT-MAMS: protein methyltransferase motif analysis by mass spectrometry. *J Proteome Res*. 17(10):3485–3491.
Hamey JJ, Wienert B, Quinlan KGR, Wilkins MR. 2017. METTL21B is a novel human lysine methyltransferase of translation elongation factor 1A: discovery by CRISPR/Cas9 knockout. *Mol Cell Proteomics* 16(12):2229–2242.
Hamey JJ, Wilkins MR. 2018. Methylation of elongation factor 1A: where, who, and why? *Trends Biochem Sci*. 43(3):211–223.
Hamey JJ, Winter DL, Yagoub D, Overall CM, Hart-Smith G, Wilkins MR. 2016. Novel N-terminal and lysine methyltransferases that target translation elongation factor 1A in yeast and human. *Mol Cell Proteomics* 15(1):164–176.
Han D, Huang M, Wang T, Li Z, Chen Y, Liu C, Lei Z, Chu X. 2019. Lysine methylation of transcription factors in cancer. *Cell Death Dis*. 10(4):290.
Hart-Smith G, Yagoub D, Tay AP, Pickford R, Wilkins MR. 2016. Large scale mass spectrometry-based identifications of enzyme-mediated protein methylation are subject to high false discovery rates. *Mol Cell Proteomics* 15(3):989–1006.
Hennessey KM, Alas GCM, Rogiers I, Li R, Merritt EA, Paredez AR. 2019. Nek8445, a protein kinase required for microtubule regulation and cytokinesis in *Giardia lamblia*. *bioRxiv*. 719005.
Huang DW, Sherman BT, Tan Q, Kir J, Liu D, Bryant D, Guo Y, Stephens R, Baseler MW, Lane HC, et al. 2007. DAVID Bioinformatics Resources: expanded annotation database and novel algorithms to better extract biology from large gene lists. *Nucleic Acids Res*. 35(Suppl 2):W169–W175.
Ishida T, Kinoshita K. 2007. PrDOS: prediction of disordered protein regions from amino acid sequence. *Nucleic Acids Res*. 35(Web Server):W460–W464.
Iyer LM, Anantharaman V, Wolf MY, Aravind L. 2008. Comparative genomics of transcription factors and chromatin proteins in parasitic protists and other eukaryotes. *Int J Parasitol*. 38(1):1–31.
Jakobsson ME, Małecki J, Falnes PØ. 2018. Regulation of eukaryotic elongation factor 1 alpha (eEF1A) by dynamic lysine methylation. *RNA Biol*. 15(3):314–319.
Jakobsson ME, Małecki J, Nilges BS, Moen A, Leidel SA, Falnes PØ. 2017. Methylation of human eukaryotic elongation factor alpha (eEF1A) by a member of a novel protein lysine methyltransferase family modulates mRNA translation. *Nucleic Acids Res*. 45(14):8239–8254.
Jakobsson ME, Małecki JM, Halabelian L, Nilges BS, Pinto R, Kudithipudi S, Munk S, Davydova E, Zuhairi FR, Arrowsmith CH, et al. 2018. The dual methyltransferase METTL13 targets N terminus and Lys55 of eEF1A and modulates codon-specific translation rates. *Nat Commun*. 9(1):3411.
Jiang L, Mu J, Zhang Q, Ni T, Srinivasan P, Rayavara K, Yang W, Turner L, Lavstsen T, Theander TG, et al. 2013. PfSETvs methylation of histone H3K36 represses virulence genes in *Plasmodium falciparum*. *Nature* 499(7457):223–227.
Johnson LM, Bostick M, Zhang X, Kraft E, Henderson I, Callis J, Jacobsen SE. 2007. The SRA methyl-cytosine-binding domain links DNA and histone methylation. *Curr Biol*. 17(4):379–384.
Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, et al. 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30(9):1236–1240.
Karnkowska A, Vacek V, Zubacova Z, Treitli SC, Petrzelkova R, Eme L, Novak L, Zarsky V, Barlow LD, Herman EK, et al. 2016. A eukaryote without a mitochondrial organelle. *Curr Biol*. 26(10):1274–1284.
Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 30(4):772–780.
Kernstock S, Davydova E, Jakobsson M, Moen A, Pettersen S, Mælandsmo GM, Egge-Jacobsen W, Falnes PØ. 2012. Lysine methylation of VCP by a member of a novel human protein methyltransferase family. *Nat Commun*. 3(1):1038.
Kon T, Imamula K, Roberts AJ, Ohkura R, Knight PJ, Gibbons IR, Burgess SA, Sutoh K. 2009. Helix sliding in the stalk coiled coil of dynein couples ATPase and microtubule binding. *Nat Struct Mol Biol*. 16(3):325–333.
Krtkova J, Paredez AR. 2017. Use of translation blocking morpholinos for gene knockdown in *Giardia lamblia*. *Methods Mol Biol*. 1565:123–140.
Lane S, Lloyd D. 2002. Current trends in research into the waterborne parasite *Giardia*. *Crit Rev Microbiol*. 28(2):123–147.
Lanouette S, Mongeon V, Figeys D, Couture JF. 2014. The functional diversity of protein lysine methylation. *Mol Syst Biol*. 10(4):724.
Larsen SC, Sylvestersen KB, Mund A, Lyon D, Mullari M, Madsen MV, Daniel JA, Jensen LJ, Nielsen ML. 2016. Proteome-wide analysis of arginine monomethylation reveals widespread occurrence in human cells. *Sci Signal*. 9(443):rs9.
Lee S, Sowa ME, Watanabe YH, Sigler PB, Chiu W, Yoshida M, Tsai FT. 2003. The structure of ClpB: a molecular chaperone that rescues proteins from an aggregated state. *Cell* 115(2):229–240.
Leger MM, Kolisko M, Kamikawa R, Stairs CW, Kume K, Čepička I, Silberman JD, Andersson JO, Xu F, Yabuki A. 2017. Organelles that illuminate the origins of *Trichomonas* hydrogenosomes and *Giardia* mitosomes. *Nat Ecol Evol*. 1:0092.
Liu K, Chen C, Guo Y, Lam R, Bian C, Xu C, Zhao DY, Jin J, MacKenzie F, Pawson T, et al. 2010. Structural basis for recognition of arginine methylated Piwi proteins by the extended Tudor domain. *Proc Natl Acad Sci U S A*. 107(43):18398–18403.
Low JK, Wilkins MR. 2012. Protein arginine methylation in *Saccharomyces cerevisiae*. *FEBS J*. 279(24):4423–4443.
Luján HD, Svärd S. 2011. *Giardia*: a model organism. Berlin/Heidelberg, Germany: Springer Science & Business Media.
Lupas A, Van Dyke M, Stock J. 1991. Predicting coiled coils from protein sequences. *Science* 252(5009):1162–1164.
Malecki J, Aileni VK, Ho AYY, Schwarz J, Moen A, Sorensen V, Nilges BS, Jakobsson ME, Leidel SA, Falnes PO. 2017. The novel lysine specific methyltransferase METTL21B affects mRNA translation through inducible and dynamic methylation of Lys-165 in human eukaryotic elongation factor 1 alpha (eEF1A). *Nucleic Acids Res*. 45(8):4370–4389.
Manning G, Reiner DS, Lauwaet T, Dacre M, Smith A, Zhai Y, Svard S, Gillin FD. 2011. The minimal kinome of *Giardia lamblia* illuminates early kinase evolution and unique parasite biology. *Genome Biol*. 12(7):R66.
Marcial-Quino J, Gomez-Manzo S, Fierro F, Rufino-Gonzalez Y, Ortega-Cuellar D, Sierra-Palacios E, Vanoye-Carlo A, Gonzalez-Valdez A, Torres-Arroyo A, Oria-Hernandez J, et al. 2017. RNAi-mediated specific gene silencing as a tool for the discovery of new drug targets in *Giardia lamblia*; evaluation using the NADH oxidase gene. *Genes (Basel)* 8(11):303.
Mason JM, Arndt KM. 2004. Coiled coil domains: stability, specificity, and biological implications. *ChemBioChem* 5(2):170–176.
Meirelles GV, Perez AM, de Souza EE, Basei FL, Papa PF, Melo Hanchuk TD, Cardoso VB, Kobarg J. 2014. "Stop Ne(c)king around": how interactomics contributes to functionally characterize Nek family kinases. *World J Biol Chem*. 5:141–160.
Midlej V, Benchimol M. 2009. *Giardia lamblia* behavior during encystment: how morphological changes in shape occur. *Parasitol Int*. 58(1):72–80.
Millan-Zambrano G, Chavez S. 2014. Nuclear functions of prefoldin. *Open Biol*. 4(7):140085.
Mizuno K, Sloboda RD. 2017. Protein arginine methyltransferases interact with intraflagellar transport particles and change location during flagellar growth and resorption. *Mol Biol Cell* 28(9):1208–1222.
Moore KE, Gozani O. 2014. An unexpected journey: lysine methylation across the proteome. *Biochim Biophys Acta* 1839(12):1395–1403.
Morishita M, Mevius DEHF, Shen YP, Zhao SY, di Luccio E. 2017. BIX-01294 inhibits oncoproteins NSD1, NSD2 and NSD3. *Med Chem Res*. 26(9):2038–2047.

Morrison HG, McArthur AG, Gillin FD, Aley SB, Adam RD, Olsen GJ, Best AA, Cande WZ, Chen F, Cipriano MJ, et al. 2007. Genomic minimalism in the early diverging intestinal parasite *Giardia lamblia*. *Science* 317(5846):1921–1926.

Murn J, Shi Y. 2017. The winding path of protein methylation research: milestones and new frontiers. *Nat Rev Mol Cell Biol*. 18(8):517–527.

Nixon JE, Wang A, Morrison HG, McArthur AG, Sogin ML, Loftus BJ, Samuelson J. 2002. A spliceosomal intron in *Giardia lamblia*. *Proc Natl Acad Sci U S A*. 99(6):3701–3705.

Nosala C, Dawson SC. 2015. The critical role of the cytoskeleton in the pathogenesis of *Giardia*. *Curr Clin Microbiol Rep*. 2(4):155–162.

Pawlak MR, Scherer CA, Chen J, Roshon MJ, Ruley HE. 2000. Arginine *N*-methyltransferase 1 is required for early postimplantation mouse development, but cells deficient in the enzyme are viable. *Mol Cell Biol*. 20(13):4859–4869.

Petrossian TC, Clarke SG. 2011. Uncovering the human methyltransferasome. *Mol Cell Proteomics* 10(1):M110.000976.

Qiao Q, Li Y, Chen Z, Wang M, Reinberg D, Xu RM. 2011. The structure of NSD1 reveals an autoregulatory mechanism underlying histone H3K36 methylation. The. *J Biol Chem*. 286(10):8361–8368.

Que X, Svard SG, Meng TC, Hetsko ML, Aley SB, Gillin FD. 1996. Developmentally regulated transcripts and evidence of differential mRNA processing in *Giardia lamblia*. *Mol Biochem Parasitol*. 81(1):101–110.

Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. 2015. *limma* powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 43(7):e47.

Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Hohna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol*. 61(3):539–542.

Salusso A, Zlocowski N, Mayol GF, Zamponi N, Ropolo AS. 2017. Histone methyltransferase 1 regulates the encystation process in the parasite *Giardia lamblia*. *FEBS J*. 284(15):2396–2409.

Schubert HL, Blumenthal RM, Cheng X. 2003. Many paths to methyltransfer: a chronicle of convergence. *Trends Biochem Sci*. 28(6):329–335.

Smith AJ, Lauwaet T, Davids BJ, Gillin FD. 2012. *Giardia lamblia* Nek1 and Nek2 kinases affect mitosis and excystation. *Int J Parasitol*. 42(4):411–419.

Sonda S, Morf L, Bottova I, Baetschmann H, Rehrauer H, Caflisch A, Hakimi MA, Hehl AB. 2010. Epigenetic mechanisms regulate stage differentiation in the minimized protozoan *Giardia lamblia*. *Mol Microbiol*. 76(1):48–67.

Stadelmann B, Hanevik K, Andersson MK, Bruserud O, Svard SG. 2013. The role of arginine and arginine-metabolizing enzymes during *Giardia*—host cell interactions in vitro. *BMC Microbiol*. 13(1):256.

Surpili MJ, Delben TM, Kobarg J. 2003. Identification of proteins that interact with the central coiled-coil region of the human protein kinase NEK1. *Biochemistry* 42(51):15369–15376.

Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva NT, Morris JH, Bork P, et al. 2019. STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res*. 47(D1):D607–D613.

Tanifuji G, Takabayashi S, Kume K, Takagi M, Nakayama T, Kamikawa R, Inagaki Y, Hashimoto T. 2018. The draft genome of *Kipferlia bialata* reveals reductive genome evolution in fornicate parasites. *PLoS One* 13(3):e0194487.

Tempel W, Wu H, Dombrovsky L, Zeng H, Loppnau P, Zhu H, Plotnikov AN, Bochkarev A. 2009. An intact SAM-dependent methyltransferase fold is encoded by the human endothelin-converting enzyme-2 gene. *Proteins* 74(3):789–793.

Thandapani P, O'Connor TR, Bailey TL, Richard S. 2013. Defining the RGG/RG motif. *Mol Cell* 50(5):613–623.

Touz MC, Ropolo AS, Rivero MR, Vranych CV, Conrad JT, Svard SG, Nash TE. 2008. Arginine deiminase has multiple regulatory roles in the biology of *Giardia lamblia*. *J Cell Sci*. 121(17):2930–2938.

Tyanova S, Temu T, Cox J. 2016. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat Protoc*. 11(12):2301–2319.

Werner-Peterson R, Sloboda RD. 2013. Methylation of structural components of the axoneme occurs during flagellar disassembly. *Biochemistry* 52(47):8501–8509.

Wu D, Lim E, Vaillant F, Asselin-Labat ML, Visvader JE, Smyth GK. 2010. ROAST: rotation gene set tests for complex microarray experiments. *Bioinformatics* 26(17):2176–2182.

Xu F, Jerlstrom-Hultqvist J, Einarsson E, Astvaldsson A, Svard SG, Andersson JO. 2014. The genome of *Spironucleus salmonicida* highlights a fish pathogen adapted to fluctuating environments. *PLoS Genet*. 10(2):e1004053.

Xu F, Jerlstrom-Hultqvist J, Kolisko M, Simpson AG, Roger AJ, Svard SG, Andersson JO. 2016. On the reversibility of parasitism: adaptation to a free-living lifestyle via gene acquisitions in the diplomonad *Trepomonas* sp. PC1. *BMC Biol*. 14(1):62.

Yagoub D, Hart-Smith G, Moecking J, Erce MA, Wilkins MR. 2015. Yeast proteins Gar1p, Nop1p, Npl3p, Nsr1p, and Rps2p are natively methylated and are substrates of the arginine methyltransferase Hmt1p. *Proteomics* 15(18):3209–3218.

Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. 2015. The I-TASSER Suite: protein structure and function prediction. *Nat Methods*. 12(1):7–8.

Zuma AA, Santos JO, Mendes I, de Souza W, Machado CR, Motta MCM. 2017. Chaetocin—a histone methyltransferase inhibitor—impairs proliferation, arrests cell cycle and induces nucleolar disassembly in *Trypanosoma cruzi*. *Acta Trop*. 170:149–160.