bioRxiv preprint doi: https://doi.org/10.1101/264291; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Structure of a novel dimeric SET domain methyltransferase that regulates cell motility

Yulia Pivovarova[1], Jun Liu[2], Johannes Lesigang[1], Oliver Koldyka[3], Rene Rauschmeier[1], Ke Hu[2,*] and Gang Dong[1,*]

Running title: Crystal structure of Toxoplasma gondii AKMT

[1] Max F. Perutz Laboratories, Medical University of Vienna, Vienna Biocenter (VBC), Dr. Bohr-Gasse 9, 1030 Vienna, Austria
[2] Department of Biology, Indiana University, Bloomington, IN 47405, USA
[3] FH Campus Wien, Vienna, Austria

[*] Correspondence to: kehu@indiana.edu; gang.dong@meduniwien.ac.at

bioRxiv preprint doi: https://doi.org/10.1101/264291; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Abstract

Lysine methyltransferases (KMTs) were initially associated with transcriptional control through their methylation of histones and other nuclear proteins, but have since been shown to regulate many other cellular activities. The apical complex lysine (K) methyltransferase (AKMT) of the human parasite *Toxoplasma gondii* was recently shown to play a critical role in regulating cellular motility. Here we report a 2.1-Å resolution crystal structure of the conserved and functional C-terminal portion (aa289-709) of *T. gondii* AKMT. AKMT dimerizes via a unique intermolecular interface mediated by the C-terminal TPR (tetratricopeptide repeat)-like domain together with a specific zinc-binding motif that is absent from all other KMTs. Disruption of AKMT dimerization impaired both its enzyme activity and egress from infected host cells *in vivo*. Overall, our findings reveal AKMT as the founding member of a new subclass of KMTs that adopt a novel dimeric conformation to regulate cell motility.

Key words:
AKMT/crystallography/lysine methylation/methyltransferase/SET domain/structure

bioRxiv preprint doi: https://doi.org/10.1101/264291; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Introduction

Lysine methyltransferases (KMTs) were first found to be involved in transcriptional control through their methylation of nuclear proteins including histones and various transcription factors [1-3]. In recent years, the roles of KMTs in other cellular activities have started to emerge [4-7]. A relatively new member of the family, the apical complex lysine (K) methyltransferase (AKMT) was found in the phylum Apicomplexa, which includes more than six thousand species of intracellular parasites [6, 8]. Many of these parasites pose serious health threats to humans. *Plasmodium falciparum* infections claim nearly half a million lives each year [9]. Another apicomplexan parasite, *Toxoplasma gondii*, permanently infects nearly 20% of the global population. *Toxoplasma* infections in immunocompromised individuals (*e.g.* organ transplant recipients or AIDS patients) have devastating consequences, including the development of lethal toxoplasmic encephalitis due to extensive cerebral lesions [10, 11]. Similarly, maternally-transmitted *T. gondii* infection of unprotected fetuses leads to congenital toxoplasmosis [12]. To sustain an infection, the parasites need to regulate their motility appropriately in response to environmental changes to complete the lytic cycle, during which they actively invade a host cell, replicate, and disseminate by egress [11, 13-16].

Previously, we identified AKMT [for apical complex lysine (K) methyltransferase] as a key motility regulator which controls the transition from immotile to motile behavior in the lytic cycle of *T. gondii* [6]. In immotile intracellular parasites AKMT is localized to the cytoskeletal apical complex, a set of structures at the tip of the parasite that play an important role in host cell invasion [6, 17, 18]. An influx of Ca^{2+} is the signal that normally activates parasite motility, and when cytoplasmic [Ca^{2+}] increases, AKMT disperses throughout the parasite in response [6]. The AKMT knockout (Δ*akmt*) parasites generally fail to switch from the immotile to the motile state in response to this stimulus. This results in a dramatic delay in parasite dispersion during egress and a 10-fold decrease in (re)invasion efficiency, severely impairing the lytic cycle. The enzymatic activity of AKMT is important for its function in the parasite, as an enzymatically dead AKMT allele (H447V) does not complement the motility defect of the Δ*akmt* parasites. Although the native substrate(s) of AKMT are not yet known, cortical force measurement by laser trap analyses suggests that it might directly or indirectly regulate actin polymerization [19].

AKMT contains a SET-domain, which is an evolutionarily conserved structural unit that was originally identified in Suppressor of variegation [Su(var)3-9], Enhancer of zeste [E(z)]

bioRxiv preprint doi: https://doi.org/10.1101/264291; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

and Trithorax [20, 21]. Our previous phylogenetic analysis showed that AKMT orthologs form a clade close to, but distinct from, the SMYD (SET- and myeloid-Nervy-DEAF-1 (MYND)-domain) family of KMTs [22]. Importantly, while AKMT is found in all sequenced genomes of apicomplexans, there are no mammalian AKMT orthologs. This makes it a possible target for designing new, parasite-specific drugs.

In this study, we report a 2.1-Å resolution crystal structure of the structured C-terminal region of AKMT (aa289-709) without bound cofactors (apo form). The structure reveals a unique handshake-like dimerization interface that is absent in any other SET-domain-containing KMTs, including the SMYD proteins- the nearest homologs of AKMT. Further, compared with those in the SMYD proteins, which methlylate unstructured short peptides, the substrate binding groove of AKMT is unusally wide, indicating that it is capable of methylating large globular substrates. Dimerization of AKMT is mediated by a unique zinc-binding motif together with the C-terminal TPR (tetratricopeptide repeat)-like domain. In *vitro* assays demonstrated that dimerization is important for the enzymatic activity of AKMT, while *in vivo* egress assays showed that dimerization is also required for efficient activation of parasite motility. Taken together, our findings reveal a novel homodimeric SET domain-containing KMT that might have important implications in the evolution of a large group of parasites as well as KMTs.

Results

Overview of the crystal structure of *T. gondii* AKMT

The 80-kDa AKMT is a modular protein; bioinformatics analyses suggested that most of the N-terminal part of the protein (aa1-300) is intrinsically disordered, while the C-terminal folded portion (aa301-709) was predicted to contain four structural domains (**Fig. S1A, Fig 1A**). Sequence alignment indicated that the C-terminal region is well conserved among apicomplexan species, whereas the N-terminal regions is highly variable (**Fig. S1B**). Limited proteolysis using various proteases (chymotrypsin, trypsin, thermolysin, etc.) on purified recombinant AKMT consistently generated a ~50-kDa protease-resistant core of only the C-terminal region (data not shown). This region thus appeared to constitute the structural and functional core of the protein and was made the focus of our structural studies.

bioRxiv preprint doi: https://doi.org/10.1101/264291; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

A panel of N-terminally truncated AKMT constructs including aa289-709, aa295-709, and aa301-709 were expressed and purified in bacteria. Purified recombinant proteins, both with and without reductive lysine methylation, were subjected to crystallization trials. The lysine methylated AKMT-ΔNTD (aa289-709) yielded crystals that diffracted X-rays to 2.1-Å resolution with space group *P1* ($a = 61.94$ Å, $b = 89.37$ Å, $c = 91.72$ Å; $\alpha = 108.69°$, $\beta = 101.32°$, $\gamma = 103.41°$). Structure was determined using the single-wavelength anomalous dispersion (SAD) method, exploiting the anomalous signal from the bound zinc ions in the protein. The final structure had a R_{work} and R_{free} of 19.1% and 22.6%, respectively (**Table 1**).

Consistent with a previous report [22], AKMT-ΔNTD was revealed to contain a SET domain (aa290-486) followed by a post-SET motif (aa487-517), together with a TPR-like C-terminal domain (CTD, aa551-709). Interestingly, the structure additionally contained a three-stranded antiparallel β-sheet (β10-12, aa518-550) between the post-SET motif and the TPR-like CTD (**Fig. S1B)**, which is absent in all other SET-domain-containing KMTs except for AKMT orthologs and thus named the AKMT-specific insertion (ASI) motif (**Fig. 1A, B**). AKMT-ΔNTD consisted of eighteen α helices and thirteen β strands (**Fig. S2**), which were folded into two topologically distinct lobes (**Fig. 1B, C**). The final 2Fo-Fc map overall had an excellent quality (**Fig. 1D**), and the two bound zinc ions in the protein were well defined (**Fig. 1E**). Additionally, the methylated side chains of most lysine residues were resolvable (**Fig. S3A, B**).

AKMT forms a homodimer

The refined structural model contained four molecules per asymmetric unit, consisting of residues 289-707, 292-709, 289-708, and 289-709, which were designated as chains A, B, C and D, respectively (**Fig. 2A**). Chains A and D also contained vector-derived residues "GSHM" and "M", respectively, at their N-termini. The final model additionally contained five sulfate (SO_4^{2-}) ions, eight glycerol molecules, 21 ethylene glycols, and one diethylene glycol (**Fig. S3C**). The four molecules in the asymmetric unit of the crystal were virtually the same with root-mean-square deviation (RMSD) of ~0.2 Å for the Cα atoms, except for a small rigid-body shift of the N-terminal extensions between chains A/D and B/C, which is consistent with the pseudo-twofold axis related equivalence of the A-B and D-C dimers (**Fig. 2B**).

bioRxiv preprint doi: https://doi.org/10.1101/264291; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The tetramer in the final model had four intermolecular dimeric interfaces of two types, with the A-B dimer being structurally equivalent to the D-C dimer, and the A-C dimer equivalent to the D-B dimer. The two types of dimeric interface were both pronounced and had a similar buried surface areas (BSA) of ~1800 $Å^2$ (**Fig. S4**), and it was unclear whether either or both of these interfaces are physiological or just a crystallization artifact. To reveal the native oligomerization state of AKMT, we carried out static light scattering (SLS) analyses on both full-length AKMT and AKMT-ΔNTD. The results showed that both proteins formed homodimers (**Fig. 2C**, green and blue traces), suggesting that a dimeric conformation is the native state of AKMT. Therefore, only one of the two dimeric interfaces present in the tetrameric crystal structure is physiologically relevant.

To determine which of the two dimeric interfaces (A-C/D-B, or A-B/D-C) in the crystal structure represents the native one, we first disrupted the A-C dimer interface by deleting the N-terminal segment (i.e. α1 and β1) of the structure (designated as AKMT-ΔNTD-short, aa309-709) (**Fig. 2A**, upper insert). The resulting protein generated a similar dimer to the wild-type (**Fig 2C**, magenta trace), suggesting that this interface is not the native inter-dimer interface. Based on our analyses of the intermolecular interactions in the crystal structure using PDBePISA [23], eight of the ten direct hydrogen bonds and all salt bridges in the A-B/D-C interfaces were mediated by the side chains of three charged residues (R521, R672, and D676) plus the polar residue N679 (**Fig. S4**). We therefore mutated all these residues to alanine to try to disrupt the interaction between chains A and B (**Fig. 2A**, lower insert). The resulting mutant (AKMT-ΔNTD-4A: R521A/R672A/D676A/N679A) was found to become monomeric in SLS analyses, indicating that the A-B interface is indeed the native dimeric interface (**Fig. 2C**, red trace). For convenience, the construct AKMT-ΔNTD-4A will be referred to as the monomeric mutant in this study.

To further verify the dimeric conformation of AKMT, we additionally assayed AKMT-ΔNTD in solution using small-angle X-ray scattering (SAXS). The results showed that the theoretical scattering of the A-B dimer was in excellent agreement with the experimental data (χ^2 = 1.59), whereas the simulated curve of the A-C dimer was not (χ^2 = 10.99) (**Fig. 2D**). Furthermore, the A-B dimer fitted perfectly into the calculated *ab initio* model reconstructed from the SAXS data (**Fig. 2E**). Therefore, we concluded that AKMT forms a homodimer in its native state, and the dimerization is mediated by its C-terminal region. The N-terminally mediated dimerization in the crystal structure was probably an artifact caused by domain swapping during crystallization.

bioRxiv preprint doi: https://doi.org/10.1101/264291; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The dimeric interface of AKMT contains a unique zinc-binding motif

Both post-SET and ASI motifs in AKMT bound a zinc ion, and the zinc ion in each case was coordinated by a cluster of four cysteine residues, C449/C505/C507/C510 in the post-SET motif and C522/C525/C544/C547 in the ASI motif (**Fig. 3A, B**). Notably, one of the cysteine residues (C449) participating in zinc binding in the post-SET motif is located in the SET domain. The ASI motif is mostly buried at the A-B dimeric interface, and mutating these zinc-binding cysteine residues in the ASI made the protein insoluble (data not shown). This suggests that the ASI motif and the interactions it mediates at the dimerization interface play a critical role in maintaining the structural stability of AKMT.

A closer examination of the structure revealed that AKMT formed an "open hand"-like structure, with the ASI motif forming the “fingers”, and the central and C-terminal parts of the TPR-like domain forming the “palm” and the “thumb”, respectively (**Fig. 3C, D**). In the homodimer, the two hands were orientated in a handshake manner, with the “fingers” of one hand reaching the base of the “thumb” on the other and the two “palms” directly facing each other (**Fig. 3E**). The homodimer of AKMT was stabilized by more than twenty hydrogen bonds (directly or mediated by water molecules), two salt bridges, and many hydrophobic interactions (**Fig. 3F**; **Fig. S4**).

AKMT binds cofactor in submicromolar range

The lysine methyltransferase activity of AKMT has been previously tested *in vitro* using an artificial histone substrate and a ^{3}H-labeled cofactor, S-adenosyl-L-methionine (SAM) [6, 22]. These assays demonstrated that AKMT is an active lysine methyltransferase capable of binding SAM, which serves as a methyl group donor in the methylation reaction. The crystal structure of AKMT reported here was obtained by crystallizing AKMT in the absence of cofactors. To verify the binding of AKMT to cofactors and to understand how the interaction occurs, we first carried out isothermal titration calorimetry (ITC) experiments. Recombinant full-length AKMT bound the cofactor SAM with a dissociation constant (K_d) of 0.62 μM; AKMT-ΔNTD showed a similar affinity of K_d = 0.63 μM (**Fig. 4A**). AKMT-ΔNTD bound the cofactor analog sinefungin (SFG) with a much higher affinity (K_d = 0.26 μM), likely due to

bioRxiv preprint doi: https://doi.org/10.1101/264291; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

extra hydrogen bonds formed between the additional amide protons on SFG and surrounding residues of AKMT (**Fig. 4A**).

The monomeric mutant AKMT-ΔNTD-4A bound SAM and SFG with K_d values of 0.86 μM and 0.71 μM, respectively (**Fig. 4B**). The mutant therefore showed only slightly reduced cofactor binding affinity compared to AKMT-ΔNTD based on the measured K_d values. However, the N values for both mutant/cofactor interactions (N = 0.27 and 0.51 for interaction with SAM and SFG, respectively), which represent the molar ratio between the cofactor and the target protein, were much lower than those of the AKMT-ΔNTD (N = ~0.9). This was probably because some of the protein in the ITC reaction chamber was unable to bind cofactors. A low N value (0.39) in the reaction between full-length AKMT and cofactor SAM was also observed, which was most likely caused by the co-purified impurities as seen in Fig. 2C. Such impurities could not be detected on a SDS-PAGE, suggesting that they are probably not proteins. Nevertheless, the impurities had ultraviolet absorbance that influenced protein concentration measurements, and thus indirectly reduced the amount of active full-length AKMT used for calculation in the ITC reaction. Notably, no such impurities were present in all other AKMT proteins including the mutant AKMT-ΔNTD-4A (**Fig. 2C**).

Comparison of the structure of AKMT with its structural homolog SMYD2 in complex with the cofactor SAM revealed a putative cofactor binding site in AKMT (**Fig. S5A, B**). In the AKMT structure we observed that the residue H447, which is highly conserved across the phylum Apicomplexa, not only forms a hydrogen bond via its side chain with the backbone amide proton of the neighboring residue Y481, which helps define the cofactor binding pocket, but is also in close proximity (3.4-4.3 Å) to the zinc-binding site in the ASI motif (**Fig. 4C**). Mutation of H447 to valine (H447V) would be predicted to disrupt its interaction with Y481, and perturb the structure around the putative cofactor binding site. This explains our previously published observation that the H447V mutant completely abolished the enzymatic activity and *in vivo* function of AKMT [6]. Such a mutation possibly also destabilizes the nearby ASI motif that forms a part of the dimerization interface of AKMT.

The stoichiometry for SFG:AKMT-ΔNTD in the ITC experiment was approximately 0.9:1, indicating that both AKMT molecules in the homodimer are capable of binding cofactors. As an additional test, we carried out mass spectrometry (MS) experiments. The molecular mass of AKMT-ΔNTD measured at a denaturing condition was in excellent agreement with the theoretical molecular weight calculated based on the primary sequence (**Fig. 4D**). Further MS

bioRxiv preprint doi: https://doi.org/10.1101/264291; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

analyses of SFG-bound AKMT-ΔNTD at native state showed that the majority of the protein was a dimer loaded with two SFG molecules plus four zinc ions (**Fig. 4E**). Subsequent tests for collision-induced dissociation of ligand from native SFG-bound AKMT-ΔNTD, which were analyzed by tandem mass spectrometry, revealed that increasing trap collision energy (CE) led to dissociation of SFG from the dimer, but affected neither the stability of the dimer nor its binding to zinc ions (**Fig. 4F**).

Dimerization of AKMT is important for maintaining its thermostability

The weaker cofactor binding and lower molecular ratio (*i.e.* N values) of the monomeric mutant AKMT-ΔNTD-4A compared to the AKMT-ΔNTD protein suggested that the mutant is possibly misfolded and/or structurally unstable. To test these possibilities, we first carried out circular dichroism (CD) analysis for rapid determination of the folding properties of the purified proteins [24]. The results showed similar curves for the wild-type and the monomeric mutant, suggesting that the mutant is folded similarly to the wild-type (**Fig. 4G**).

We further checked the stability of the monomeric mutant by differential scanning fluorimetry (DSF), a technique that generates a melting curve for a target protein by monitoring the increase in the fluorescence of a dye binding to hydrophobic parts of the protein while the protein gradually unfolds in increasing temperature [25]. The results showed that wild-type AKMT-ΔNTD had a similar melting temperature (Tm) to the full-length AKMT (46°C vs 45°C), but the monomeric mutant had a drastically reduced Tm (38°C) compared to the wild-type protein (**Fig. 4H**). These results demonstrated that disruption of the dimerization interface renders the protein structurally less stable.

Dimerization of AKMT is important for its function *in vitro* and *in vivo*

To test the effects of dimerization on the *in vitro* methylation activity, we compared the activities of the WT, AKMT-ΔNTD, and AKMT-ΔNTD-4A constructs in methylation reactions (**Fig. 5A**). The reactions constituted a series of twelve 2-fold dilutions of the enzyme starting from ~7.6 μM (for AKMT-ΔNTD) or ~7.8 μM (for AKMT-ΔNTD-4A). Each reaction also contained 4.1 μM of the artificial substrate, human histone H3.3, and 9.2 μM of the tritium labeled co-factor, ^{3}H-SAM. The difference between the two truncations was pronounced in reactions with less than 7.6×2^{-7} μM of the enzyme (*i.e.* [enzyme]:[substrate] < ~1:70), in which

bioRxiv preprint doi: https://doi.org/10.1101/264291; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the amount of enzyme was likely to be limiting (**Fig. 5A**, right-hand lower panel). The methylation level of AKMT-ΔNTD-4A at 7.8×2^{-8} μM was lower than that in the AKMT-ΔNTD reaction with 7.6×2^{-11} μM enzyme, indicating that AKMT-ΔNTD-4A is less active than AKMT-ΔNTD as a methyltransferase.

To test the effects of the dimerization-blocking mutations on parasite egress, we transiently expressed full-length AKMT-4A and AKMT-WT proteins with a fused eGFP tag at the N-terminus in Δ*akmt* parasites (**Fig. 5B-D**) to assess the complementation efficacy of expressing these alleles without long-term selection in a population. The eGFP-AKMT-WT construct localized to the apical complex in intracellular parasites and translocated in response to the addition of the calcium ionophore A23187 (**Fig. 5B,** upper panels). Similarly, the eGFP-AKMT-4A mutant also localized and redistributed upon stimulus in the same manner as the WT protein (**Fig. 5B**, lower panels). However, eGFP-AKMT-4A expressing parasites displayed a slower response to the A23187 stimulus than those expressing eGFP-AKMT-WT. As previously reported [6], the untransfected Δ*akmt* parasites had severe defects in A23187-induced egress (parasites in 40 out of 42 vacuoles failed to egress). This defect was complemented by eGFP-AKMT-WT expression: eGFP-AKMT-WT-expressing parasites egressed within ~11 minutes after A23187 addition in ~89% of the vacuoles (40 out of 45). Within the same time frame, eGFP-AKMT-4A-expressing parasites in ~74 % of the vacuoles (28 out of 38) egressed (**Fig. 5C**). Furthermore, for those that did egress, the eGFP-AKMT-4A expressing parasites egressed with a marked delay compared with those expressing eGFP-AKMT-WT (**Fig. 5D**). The median length of time for the parasites to respond to the addition of A23187 to egress was ~244 seconds for eGFP-AKMT-WT expressors, and ~353 seconds for eGFP-AKMT-4A expressors, respectively. These results indicated that AKMT-4A is functionally compromised *in vivo*.

AKMT is a unique homodimeric lysine methyltransferase

SET domains have been found in over a thousand proteins of diverse functions in organisms ranging from viruses to mammals [26]. Most SET domain-containing KMTs exist as monomers, but there are exceptions. Previous biochemical studies have demonstrated the self-association of two SET domains in human ALL-1 and in *Drosophila* TRITHORAX and ASH1 proteins [27]. A crystal structure of the viral SET (vSET) protein from the Paramecium bursaria chlorella virus, which has explicit methyltransferase activity for host histone H3 lysine

bioRxiv preprint doi: https://doi.org/10.1101/264291; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

27, displayed a side-by-side homodimer [28] (**Fig. 6A, B**). The dimer was believed to be similar to that formed by the TRITHORAX and Ash1 proteins on the basis of the conserved interface residues among them [29]. However, in contrast to these dimeric KMTs, the dimerization of AKMT is distinct in several aspects. Firstly, the SET domain in AKMT is not involved in its dimeric interaction. Secondly, unlike the side-by-side arrangement in vSET, the dimerization of AKMT is via a C- and C-terminal (C-C) interaction mediated by the unique ASI motif together with the superhelical TPR-like CTD (**Fig. 6C**). Thirdly, the dimer of AKMT is structurally much more stable than that in the vSET structure, as indicated by the doubled BSA and the dramatically lower Gibbs free energy ($\Delta^{i}G$) in the AKMT dimer (**Fig. 6D, E**).

Besides AKMT, there are many other unrelated proteins that also form homodimers via diverse interactions between two identical TPR or TPR-like domains. For examples, the anaphase-promoting complex/cyclosome (APC/C) subunit Cut9 from *Schizosaccharomyces pombe* forms an N- and N-terminal (N-N) intertwined dimer [30] (**Fig. S6A**), the light chain of the mammalian microtubule motor kinesin-1 forms a C-C dimer [31] (**Fig. S6B**), and the superhelical TPR domain of human O-linked GlcNAc transferase forms a middle-to-middle (M-M) dimer [32] (**Fig. S6C**). However, the dimerization of AKMT is clearly different from all these reported TPR-mediated homodimers. Firstly, the TPR-like domain in the AKMT dimer is fully engaged to allow maximal interaction along the two conformationally complementary surfaces (**Fig. 6E**), unlike other TPR dimers mentioned above that are formed via the interaction of only parts of the two involved TPR domains (**Fig. S6A-C**). Secondly, besides the TPR-like domain, the ASI motif is also profoundly involved in the dimerization of AKMT. In fact, the two ASI motifs in the AKMT dimer not only provide a similar buried surface area to that of the TPR-like domains, but also contribute most of the hydrogen bonds and salt bridges between the two molecules in the complex (**Fig. S4**). Deletion of the ASI motif would render the AKMT dimer significantly less stable (**Fig. 6D, E**). Therefore, AKMT is a unique homodimeric KMT that is stably maintained via an extensive interacting network provided by both the ASI motif and the TPR-like domain.

Structural comparison of AKMT with SMYD proteins

Previous phylogenetic analysis indicated that the AKMT and SMYD families branched off from a common node of SET domain-containing methyltransferases during evolution [22]. In the crystal structure, we found that AKMT shares an overall similar domain arrangement to SMYD proteins, including an N-terminal SET domain, a post-SET cysteine cluster, and a C-

bioRxiv preprint doi: https://doi.org/10.1101/264291; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

terminal TPR-like domain (**Fig. 7A**). In fact, many parts of AKMT, accounting for approximately 50% of the sequence in the crystal structure reported here, are very similar to the corresponding regions in SMYD proteins (**Fig. S7**). Using the flexible alignment mode on the RAPIDO server [33], which is a tool for rapid alignment of protein structures in the presence of domain movements, the RMSDs of the locally aligned rigid bodies between AKMT and SMYD proteins were found to be approximately 1.08-1.34 Å (**Fig. S7A**), which are nearly as good as those among different SMYD proteins that share much higher sequence identities with one another (30-35%) than with AKMT (~18%). Superposition of the corresponding rigid bodies between AKMT and SMYD3 showed clear similarities of the structures (**Fig. S7B-H**).

Despite the high degree of local alignment with the SMYD proteins, AKMT has a number of unique structural features which is consistent with a clear divergence between these two subfamilies after the speciation of the apicomplexan parasites [22]. First of all, the unique zinc-binding ASI motif in AKMT is not found in any SMYD proteins (**Fig. 7A, Fig. S1B**). Secondly, AKMT and orthologs do not have the signature MYND (Myeloid-Nervy-DEAF1) domain of SMYD proteins, which is a zinc finger motif embedded in the SET domain and located at the tip of the N-terminal lobe to gauge the entrance to the groove between the N- and C-terminal lobes (**Fig. 7B**). Previous studies suggested that the MYND domain serves as a protein-protein interaction module mainly in the context of transcriptional regulation [34, 35]. Lack of a MYND domain in AKMT might correlate with its specific function outside the nucleus [6].

Thirdly, the groove between the N- and C-terminal lobes in AKMT is considerably wider than that in all known SMYD structures, due to both the larger distance between its SET and TPR-like domains and the absence of MYND domain from its N-terminal lobe (**Fig. 7B**). The grooves in SMYD1-3 were proposed to accommodate their U-shaped substrates of various sizes [36]. Similarly, the larger open groove in AKMT may have evolved to target a unique substrate that regulates the motility switch of the parasite during egress. Such a variation in the open grooves of AKMT and SMYDs is coupled with a dramatic flip of a "stirring" loop (aa425-438 in AKMT) protruding out from the hinge connecting the N- and C-terminal lobes, close to the putative substrate binding site in AKMT (**Fig. 7C**). While the loops in all SMYDs adopt a similar conformation of a hairpin-like two-stranded antiparallel β sheet, the one in AKMT is irregularly folded and tilts towards the N-terminal SET domain (**Fig. 7D**). Sequence alignments show that the "stirring" loop in AKMT is 4-5 residues shorter than those in SMYD1-3 and highly conserved in all AKMT orthologs (**Fig. 7E**), suggesting that the loop might play a unique

bioRxiv preprint doi: https://doi.org/10.1101/264291; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

and important role in AKMT function, such as to regulate the docking and/or selection of the so-far-unknown substrates. The difference of the substrate binding grooves in AKMT and SMYD proteins is better viewed in the superimposed structures (**Fig. 7F**).

Finally, AKMT forms a distinct dimer that has not been seen in any SMYD proteins. This can be explained by the difference in the structural features at the C-terminal end of the proteins (**Fig. 8**). SMYD1-3 do not appear to possess the necessary features that are compatible for dimer formation. While the surface area at the C-terminal end of the AKMT structure is largely hydrophobic (**Fig. 8A, B**), the corresponding regions in the structures of SMYD1-3 are mostly covered by charged residues (**Fig. 8C-E**). Further, the buried interfaces on the two subunits of the AKMT dimer are also spatially and electrostatically complementary with each other, whereas those in SMYD1-3 structures do not support the docking of one molecule on top of the other to form a C- to C-terminal dimer as that of AKMT.

Discussion

In this work we describe our structural and biochemical characterizations of *T. gondii* AKMT. We discovered that the protein contains a number of structural features that are well-conserved among AKMT orthologs, but not found in other KMT families. In particular, the groove between the N- and C-terminal lobes in AKMT is much wider than that in any known SMYD structures. Such a difference might reflect the need to accommodate different types of substrates. SMYD proteins regulate chromatin remodeling mainly by methylating histone lysines H3K4 (SMYD1-3) [37] and H3K36 (SMYD2) [38], both of which are located in the structurally disordered N-terminal tail of H3. It has been shown that SMYD proteins also methylate a number of non-histone targets, such as p53 [39], retinoblastoma tumor suppressor RB1 [40], MAP3K2 kinase [41], as well as vascular endothelial growth factor receptor-1 [42]. Similar to H3, targeted lysines in these non-histone proteins are positioned in either unstructured regions or extended loops. The disordered nature of targeted regions in these substrates allow them to accommodate a compact U-shaped structure to fit into the narrow substrate binding site between the N- and C-terminal lobes in the SMYD proteins [43-45].

In contrast, mass spectroscopy analysis of *Xenopus laevis* histone H3.3 methylated by AKMT *in vitro* revealed an unusual methylation site, K123, which is located at the C-terminal end of a 10-Å long α-helix right in the core of the folded structure [6]. This finding implies that in order to get access to residue K123 in H3.3, AKMT has to accommodate the α-helix

bioRxiv preprint doi: https://doi.org/10.1101/264291; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

containing it or even the whole folded structure into its substrate-binding site. Moreover, it was recently reported that AKMT is essential for recruiting the glideosome-associated connector (GAC) to the apical localization, and depletion of AKMT led to the disappearance of several large proteins in a western blot using α-H4K20Me3 antibodies to detect methylated lysine residues [46]. The identities of these high molecular weight proteins and whether they are real AKMT substrates are still yet to be determined. Nevertheless, it is conceivable that the wide-open groove leading to the putative substrate binding site in AKMT might be necessary for AKMT to accommodate either a large substrate or its subdomains (up to 30 Å in diameter), or several substrates with various sizes and shapes, in which case a wider groove would provide more flexibility for AKMT to target different substrates.

Notably, although the monomeric mutant AKMT-ΔNTD-4A showed the same overall conformation as the wild-type protein (**Fig. 4G**), the Tm measured by DSF dropped by ~8°C (**Fig. 4H**), suggesting that homodimerization is essential for the thermostability of AKMT. Another reason for AKMT to form a homodimer might be to cooperatively regulate cofactor and/or substrate binding in order to increase enzyme efficiency. Indeed, our ITC data showed that the AKMT-ΔNTD-4A mutant has compromised cofactor binding, with more than half of the protein failing to interact with cofactors (**Fig 4B**). MS analyses of recombinant AKMT in native conditions demonstrated that AKMT always exists as a stable dimer with all four zinc ions tightly bound, but the bound cofactors could be stripped off by increasing trap collision energy (**Fig. 4E, F**), which shows that binding of cofactors is much less robust than the zinc ions and thus might be cooperatively regulated.

Besides regulating cofactor binding, homodimerization of AKMT might additionally regulate substrate recognition. Both negative and positive cooperativity in multimeric enzymes have been seen previously. The dimeric vSET protein has a negative cooperativity between the two active sites when methylating its substrate H3K72 [29]. Such an auto-regulatory mechanism is beneficial to this viral KMT, as using only one active site at a time helps to increase its processivity on chromatin. Whether AKMT uses a similar strategy in methylating its substrates is unclear and needs to be further investigated in the future.

Our *in vitro* enzymatic activity assays, as well as *in vivo* egress rescue tests, revealed that the mutant with an impaired dimerization interface (i.e. mutant 4A) is less active than the wild-type, indicating that dimerization enhances its activity (**Fig. 5**). The fact that the AKMT-4A mutant manifested only a modest defect *in vivo* compared to the wild type was possibly due to

bioRxiv preprint doi: https://doi.org/10.1101/264291; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

partial disruption of the dimerization of AKMT, as the AKMT-ΔNTD dimer is maintained by a large number of other interactions besides the hydrogen bonds and salt bridges disrupted in the monomeric mutant (**Fig. 3E)**. Indeed, size exclusion chromatography (SEC) analyses of AKMT-ΔNTD-4A showed that a small fraction of the protein eluted as dimers at >10 mg/ml (data not shown). The *ab initio* shape reconstructed with SAXS data for AKMT-ΔNTD-4A at a similar concentration in a batch measurement displays an elongated tail connected to the spherical envelope that is comparable to the AKMT monomer (data not shown). It suggests that while most of AKM-4A protein forms monomers, a small fraction could form dimers, or else there is a quick exchange between monomer and dimer at an elevated concentration. As mentioned above, mutations of any of the four conserved cysteine residues in the ASI motif resulted in insoluble proteins, possibly due to the deleterious effect caused by the complete disruption of the dimeric interaction. All these suggest that the 4A mutant probably only partially disrupts the dimeric conformation of AKMT.

It is also worth noting that while a delay in egress does not significantly impact the parasite survival under the lenient culture conditions in the laboratory, in an animal host, this delay may render the parasite vulnerable to the attack from the immune system and can be lethal to the parasite. For instance, the loss of perforin (a pore-forming protein secreted by the parasite during egress to disrupt the membrane of the parasitophorous vacuole) delays parasite egress, but does not have any apparent impact on parasite growth in tissue culture [47]. However, it completely abolishes parasite virulence in mouse infections.

Besides the folded region of AKMT that has been extensively discussed above, it is worth mentioning that AKMT additionally contains a structurally disordered N-terminal region (aa1-290) (**Fig. S1A, B**). Such an unstructured segment does not exist in any of the closely related SMYD proteins from animals. Previous studies on different AKMT truncations demonstrated that the N-terminal region makes only minor contributions to the specific cellular localization or the enzymatic activity of the protein [22]. Nevertheless, given so many negative charges spreading throughout the N-terminal sequence (31% of aa1-290 are D or E) (**Fig. S1B**), the unstructured N-terminal region of AKMT may play subtle regulatory roles in fine-tuning the function of the enzyme, which needs to be addressed in future studies.

The protein sequences for the distinct structural features of TgAKMT are highly conserved among the apicomplexans and their free-living relatives, but not found in animals. Future identification of AKMT substrates and structural studies of AKMT-substrate complex

bioRxiv preprint doi: https://doi.org/10.1101/264291; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

will shed light on the specificity of substrate binding for designing selective drugs and for informing the evolution of the function of the SMYD-related KMTs. High-resolution structures of AKMT in complex with cofactors should also provide structural information about the shape and chemical environment of the cofactor-binding site, which could be potentially used for the development of the cofactor-competitive selective inhibitors.

Materials and methods

Cloning and site-directed mutagenesis

Sequence encoding full-length *T. gondii* AKMT (TGGT1_216080) was amplified from cDNA using PCR and ligated into the pET15b vector (Novagen) between NdeI and BamHI sites at the 5' and 3' ends, respectively. The plasmid when expressed provides an N-terminal His_6 tag followed by a thrombin cleavage site prior to the target protein. All truncations of AKMT were subsequently generated in a similar manner using the precloned full-length AKMT construct as the PCR template. The AKMT-ΔNTD fragment was amplified using 5'-gcagctgcatatggccaacccttagctccatac-3' (forward) and 5'-gaccggatcctcaactggccggtaggcg-3' (reverse) primers and ligated into pET15b to produce pET15b-His6-AKMT-ΔNTD (aa289-709). The AKMT-ΔNTD-short insert was generated using the same reverse primer and the forward primer 5'-gcagctgcatatggtgccaggcaaggggcgatg-3' and ligated into pET15b to produce pET15b-His6-AKMT-ΔNTD-short (aa309-709). The AKMT-ΔNTD-4A mutant was generated in two steps. First, three mutations R672A/D676A/N679A were introduced by Megaprimer PCRs [48]. Forward primer 5'-ctgaacgtgatatctgcacactttgaggccgcgtatgctcttctgtac-3' and reverse primer 5'-gaccggatcctcaactggccggtaggcg-3' were used for generating the 142-bp megaprimer that was subsequently elongated in a second PCR with the same primers used for AKMT-ΔNTD amplification. In the second step, the fourth mutation, R521A, was introduced using a site-directed mutagenesis protocol (forward primer: 5'-caacgcccgagggtttgcatgccctctgtgtg-3', reverse primer: 5'-cacacagagggcatgcaaaccctcgggcgttg-3') to generate pET15b-His6-AKMT-ΔNTD-4A (aa289-709, R512A/R672A/D676A/N679A). All mutations were verified by DNA sequencing.

The plasmid pmin-eGFP-AKMT-4A was constructed using NEBuilder HiFi DNA assembly kit (NEB cat# E5520S) with the following three components: (1) the pmin-eGFP vector backbone released from pmin-eGFP-AKMT (aa1-300) [22] by BglII-AflII digestion; (2)

bioRxiv preprint doi: https://doi.org/10.1101/264291; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

PCR product that contained AKMT (aa1-298) sequence and appropriate overlapping regions for Gibson assembly [amplified from pmin-eGFP-AKMT as the template and 5'-ctgtacaagtccggactcagatct-3' and 5'-gcagtgtgtatggagctaaggggttggc-3'as the primers]; (3) PCR product that contained AKMT (aa299-709)-4A sequence and appropriate overlapping regions for Gibson assembly [amplified from pET15b-His_6-AKMT-ΔNTD-4A as the template and 5'-gggcagcttctgtttacttaagtcaactggccggtaggcgttcctcg-3' and 5'-ttagctccatacacactgccccagatt-3' as the primers].

Protein expression and purification

For purifying recombinant proteins, pET15b-His_6-AKMT-ΔNTD (aa289-709), pET15b-His_6-AKMT-ΔNTD-4A (aa289-709, R512A/R672A/D676A/N679A) and pET15b-His_6-AKMT-ΔNTD-short (aa309-709) constructs were used to transform competent *E. coli* BL21(DE3) cells. Bacterial cells were grown in LB medium containing 50 μg/ml ampicillin at 37 °C until their OD_{600} reached 0.6-1.0 (approximately 2-3 h). Protein expression was induced using 250 μM of isopropyl-beta-D-thiogalactopyranoside (IPTG), and the cells were then incubated at 18°C overnight. Cells were harvested by centrifugation (6,000×g, 12 min, 4°C). Cell pellets were resuspended in prechilled lysis buffer (20 mM Tris-HCl (pH 8.0), 100 mM NaCl, 20 mM imidazole, 10 mM beta-mercaptoethanol, 5% (v/v) glycerol) and were lysed using an EmulsiFlex-C3 homogenizer (Avestin). Cell debris was removed by centrifugation (25,000×g, 40 min, 4°C), and the supernatant was filtered through a 0.45-μm pore size filter and then loaded onto a 5-ml Ni-HiTrap column (GE Healthcare) pre-equilibrated in the same lysis buffer. After washing with 5 column volumes (cv) of lysis buffer, bound protein was eluted using a linear gradient concentration of imidazole in the lysis buffer (20 to 600 mM, 10×cv). The N-terminal His_6 tag was removed by incubating the purified protein with ~5% (w/w) thrombin (4°C, overnight). In order to achieve higher purity, all proteins were subjected to SEC using a Superdex S-200 16/60 column (GE Healthcare) pre-equilibrated with running buffer (20 mM Tris-HCl (pH 8.0), 100 mM NaCl). Purified proteins were concentrated using Amicon Ultra Centrifugal Filter Units (Millipore) with appropriate molecular weight cutoffs.

For purifying proteins used in the protein lysine methyltransferase (PKMT) assay, BL21-CodonPlus(DE3)RP strain *E. coli* (CAT#230255: Stratagene, CA) were transformed using pET15b-His6-AKMT-ΔNTD or pET15b-His6-AKMT-ΔNTD-4A, and grown and induced as described above. 50 ml of bacterial culture was pelleted (6,000×g, 12 min, 4°C) and

bioRxiv preprint doi: https://doi.org/10.1101/264291; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

stored at -80 °C until use. Bacterial pellets were resuspended and lysed in Tris-acetate lysis buffer (8mM Tris-acetate pH 7.5, 7 mM Tris base pH unadjusted, 100 mM KAcetate, 1mM MgAcetate) containing 1 μM TAME (CAT# T4626: Sigma), 1 μM PMSF (CAT# P7626: Sigma), 1 μM Pepstatin A (CAT#P5318, Sigma), 0.5% Triton-100 (TX-100) and 25 mg/ml CelLytic express (CAT# C1990: Sigma), followed by sonications (Branson Sonifier 250). His_6-AKMT bound to Talon-resin (CAT# 635503: Clontech) for 1 hr at 4°C was washed with Tris-acetate lysis buffer containing 5 mM imidazole-acetate, and eluted with Tris-acetate lysis buffer containing 200 mM imidazole. dithiothreitol was added to the eluted proteins to 4 mM. Eluted proteins were dialyzed against dialysis buffer (Tris-Acetate lysis buffer containing 4 mM dithiothreitol) at 4°C in Slide A lyzer Mini Dialysis Devices (CAT# 88401: Thermo Scientific) with 3 buffer exchanges (3 to 15 hr between each exchange). Protein concentration was measured using absorbance at λ = 280 nm (Nano Drop Technologies Llc). For long-term storage at -20°C, glycerol was added to dialyzed proteins to a final concentration of 36% (v/v).

Crystallization and data collection

Initial trials to crystallize all the three AKMT truncations (aa289-709, aa296-709, and aa301-709) did not yield any crystals. In order to facilitate the crystallization process we carried out reductive lysine methylation on the purified proteins using a published protocol (Walter et al., 2006). One of the truncations (aa289-709, denoted as AKMT-ΔNTD) with methylated surface lysine residues was successfully crystallized using the hanging drop vapor diffusion method against a reservoir solution containing 0.1 M Tris-HCl (pH 9.0), 0.2 M $MgCl_2$, 30% (w/v) polyethylene glycol (PEG) 4,000 (or 20 to 30% PEG 8,000). For harvesting, crystals were soaked in the same reservoir solution augmented with increasing concentrations of glycerol (final concentration 20% [v/v]), mounted in loops (Hampton research), and flash frozen in liquid nitrogen. X-ray diffraction data collection was carried out at the European Synchrotron Radiation Facility (ESRF) on the beamline ID29. A complete and highly redundant data set to 2.1-Å resolution was collected at the absorption edge of zinc (λ = 0.9792 Å). The crystals belonged to the space group P1.

Structure determination and analyses

bioRxiv preprint doi: https://doi.org/10.1101/264291; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Diffraction data were integrated and scaled using XDS [49]. Phases were determined *de novo* using the single-wavelength anomalous dispersion (SAD) technique based on the anomalous signal from the zinc atoms bound to AKMT. In total, eight zinc ions were located and initial experimental maps were calculated using AutoSol in the software suite Phenix [50]. A preliminary structural model was built using AutoBuild, and the resulting model was optimized by multiple rounds of manual rebuilding in COOT [51] and refinement in Phenix [50]. The final model was validated in MolProbity [52].

Detailed analysis of intermolecular interfaces was performed using PISA software [23]. Structural alignment of AKMT with SMYD1-3 proteins was done by the RAPIDO server [33].

Static light scattering (SLS)

SLS measurements were carried out by coupling SEC with mass determination. 50 µl of protein sample at 2 mg/ml was analyzed on a Superdex S-200 10/300 GL column (GE Healthcare) pre-equilibrated with a buffer containing 20 mM Tris-HCl (pH 8.0), 100 mM NaCl, 1mM dithiothreitol, and 1% (v/v) glycerol. The sample was run at a flow rate of 0.5 ml/min on a High performance liquid chromatography (HPLC) system (Agilent Technologies 1260 infinity) which was directly connected to a Mini-DAWN Treos light-scattering instrument (Wyatt Technology Corp., Santa Barbara, CA). Data analyses were carried out using ASTRA software provided by the manufacturer. Final curves were built in SigmaPlot [53].

Circular dichroism (CD)

Far-UV CD spectra of AKMT-ΔNTD and AKMT-ΔNTD-4A between 180 and 280 nm were measured on a Chirascan plus CD spectrometer (Applied Photophysics) in a cuvette with a 0.5-mm path length. Proteins were diluted to approximately 0.2 mg/ml using a buffer containing 10 mM Tris-HCl (pH 8.0) and 100 mM NaF. Data points were corrected for buffer signal and drifts. CD curves were generated using SigmaPlot by averaging data collected from two scans for each protein sample.

Differential scanning fluorimetry (DSF)

bioRxiv preprint doi: https://doi.org/10.1101/264291; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

DSF measurements were performed on a CFX Connect Real-Time PCR machine (Bio-rad). Proteins were diluted to 0.4 mg/ml using a buffer containing 10 mM Tris-HCl (pH 8.0) and 100 mM NaCl. To prepare for the measurements, 24 µl of the protein sample was mixed with 1 µl of 20 × SYBR orange dye (ThermoFisher Scientific). Each sample was prepared and measured in duplicates, and the two measurements were averaged to produce the final curve. Melting temperature (T_m) values were obtained using the software provided by the manufacturer.

Isothermal titration calorimetry (ITC)

ITC measurements were performed on a MicroCal^{TM} iTC200 microcalorimeter (GE Healthcare). Cofactors SAM and SFG (Sigma-Aldrich) were dissolved in a buffer containing 20 mM Tris-HCl (pH 8.0) and 100 mM NaCl. Protein samples were dialyzed overnight against the same buffer before the measurements. A typical ITC titration experiment consisted of 20 injections of the cofactor solution, first with 1 × 0.2 µl and then with 19 × 2 µl, into the cell filled with 200 µl of protein solution. All measurements were carried out under constant stirring at 350 rpm, and each injection lasted for 4 sec with a 180 sec interval between two injections. Titration peaks were analyzed using the Origin 7.0 Microcal software and corrected for the SAM/SFG dilution heat measured by injecting cofactors into the buffer containing no protein using the same protocol described above. Non-linear least-squares fitting using one binding site model was used to calculate the association constant (K_a) and stoichiometry values. Dissociation constants (K_d) were calculated according to the formula $K_d = 1/K_a$.

Denaturing liquid chromatography-mass spectrometry (LC-MS)

The purified sample of AKMT-ΔNTD at 1mg/ml was reduced by incubating with 100 mM dithiothreitol (30 min, RT). HPLC was performed on a Dionex Ultimate 3000 HPLC system configured with the Chromeleon 6.0 software (both Thermo Fisher Scientific). The protein sample was applied on an Aeris Widepore C4 column and run at a flow rate of 300 µl/min using a 6-min step gradient with increasing acetonitrile (ACN) concentration from 9 to 36% (w/v) in 5 min, and then from 36 to 63% (w/v) in 1 min at 50°C. The HPLC system was coupled online to a quadrupole-time of flight mass spectrometer Synapt G2-Si via a Z Spray ESI source (both Waters) operated via the MassLynx V 4.1 software package (Waters). Mass

bioRxiv preprint doi: https://doi.org/10.1101/264291; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

spectra were acquired in the m/z range from 500-2000 Th at a scan rate of 1 sec. Glu[1]-Fibrinopeptide B (Glu-Fib) was used as a lock mass and spectra were corrected during data acquisition. Acquired data were analyzed with the MaxEnt 1 algorithm to reconstruct the uncharged average protein mass.

Native mass spectrometry

The protein sample at 22 mg/ml was incubated with 2 mM SFG (1 h, 4°C). After incubation, the sample was diluted to a final protein concentration of 2 mg/ml. The original buffer (20 mM Tris-HCl, pH 8.0 and 100 mM NaCl) was substituted for 400 mM ammonium acetate at pH 8.0 in Bio-Spin 6 columns (Bio-Rad Laboratories). The protein solution was transferred into pre-opened metal-coated emitters (PicoTips™, New Objective) and measured off-line on the Synapt G2-Si using a source voltage of 1.9 kV. Mass spectra were acquired in the m/z range from 100-5000 Th at a scan rate of 1 sec. To facilitate efficient desolvation of the complex, the trap collision energy was set to 12 eV and was increased up to 100 eV to shake off the non-covalently bound SFG molecules. Mass determination was performed manually.

Small angle X-ray scattering (SAXS)

Scattering curves for AKMT-ΔNTD were collected at the beamline BM29 Bio-SAXS at the ESRF in Grenoble, France, using an online SEC setup. The protein sample (200 µl, 5 mg/ml) was applied to a Superdex-200 10/30 column (GE Healthcare) equilibrated with 20 mM Tris-HCl (pH 7.5), 150 mM NaCl, 1 mM TCEP [tris(2-carboxyethyl)phosphine] and 1% (v/v) glycerol. SAXS data frames were recorded every second during the chromatography run. The data were processed in PRIMUS [54]. Fifty buffer frames selected after the protein elution peak were averaged and used for buffer subtraction from 100 averaged protein frames from the peak region where R_g value was determined to be stable. The program GNOM [55] was used to plot the pair distribution function and to determine the D_{max} value of the protein from the scattering profile. An *ab initio* model of the AKMTΔNTD dimer was reconstructed by running 10 cycles of dummy beads model building and simulated annealing implemented in DAMMIF [56]. The most probable models were selected and averaged in DAMAVER [57]. Superposition of the *ab initio* model with the crystal structure was performed by SUPCOMB [58]. SAXS curves for A-

bioRxiv preprint doi: https://doi.org/10.1101/264291; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

B and A-C dimers of the AKMT-ΔNTD were calculated and fitted to the experimental data using CRYSOL [59].

Protein lysine methyltransferase (PKMT) assay

The PKMT assay was performed as described previously[6] with some minor modifications. For activity assays with recombinant wild-type and mutant AKMT truncations (AKMT-ΔNTD and AKMT-ΔNTD-4A, respectively), 12 µl reactions were set up with varying amounts of AKMT truncations, 0.75 µg (4.1 µM) human histone H3.3 (CAT# M2507S: New England Biolabs), 0.75 mM dithiothreitol, 1.66 µCi (9.2 µM) ^{3}H-S-adenosyl-L-methionine (^{3}H-SAM) (CAT# NET155001MC: Perkin Elmer) in Tris-acetate lysis buffer (see above). Reactions were incubated at 30°C for 20 min, and stopped by the addition of LDS sample buffer (Life Technologies-Invitrogen) and reducing agent together with a subsequent 10-min heating at 70ºC. Proteins in the reactions were separated on 4-12% gradient bis-tris NuPAGE gels (CAT# NP0332: Life Technologies-Invitrogen). For detecting the methylated histones by autoradiography, proteins resolved on NuPAGE gels were transferred to 0.2µm PVDF membranes (CAT# 88520, Thermo Scientific). To visualize total proteins, the PVDF membranes were first briefly washed twice in DPBS, stained with amido black staining solution [0.1% (w/v) amido black 10B, 45% (v/v) methanol, 10% (v/v) acetic acid] for ~5 min and rinsed briefly 4-6 times in the destaining solution [5% (v/v) methanol and 7% (v/v) acetic acid]. The membranes were then air-dried, sprayed three times with EN^{3}HANCE Spray (CAT# 6NE970C: Perkin Elmer) following the manufacturer's instructions and exposed to X-ray films for 3 to 16 hours to detect the ^{3}H signal incorporated into the histone H3.3 protein. Three independent repeats were conducted with consistent results.

***Toxoplasma*, host cell cultures, and parasite transfection**

Tachyzoite *T. gondii* Δ*akmt* parasites (RhΔhx strain) were maintained by serial passage in confluent human foreskin fibroblast (HFF) (CAT# SCRC-1041: ATCC) monolayers in Dulbecco's Modified Eagle's Medium (CAT# 10569-010: Life Technologies-Gibco), supplemented with 1% (v/v) heat-inactivated cosmic calf serum (Cat# SH30087.3: Hyclone) as previously described [60, 61]. *T. gondii* transfections were carried out as previously described [62].

bioRxiv preprint doi: https://doi.org/10.1101/264291; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Egress assay

Plasmids of pmin-eGFP-AKMT-4A or pmin-eGFP-AKMT-WT [6] were used to transfect Δ*akmt* parasites as previously described [6]. The parasites were added to 35 mm dishes (#1.5) with a 20 mm microwell (CAT# P35G-1.5-20-C: MatTek, Ashland, MA) seeded with a near confluent monolayer of HFF or A7r5 (CAT# CRL-1444: ATCC) host cell monolayer and grown for ~48 h. Before imaging, the culture medium was replaced with 1 ml of phenol red-free CO_2-independent medium (SKU#RR060058: Life Technologies-Gibco) supplemented with 1% (v/v) heat-inactivated cosmic calf serum, GlutaMAX®(SKU#35050-061: Life Technologies-Gibco), and sodium pyruvate (SKU#11360: Life Technologies-Gibco). 0.5 ml of 15 µM A23187 in CO_2-independent medium was added (for a final A23187 concentration of 5 µM) to induce egress. Egress induction and imaging collection were carried out at 37°C on a DeltaVision imaging station (GE Healthcare-Applied Precision, WA) constructed on an Olympus IX-70 inverted microscope base. Differential interference contrast (DIC) and GFP fluorescence images were recorded. Vacuoles from 6-7 independent transfections each were examined for eGFP-AKMT-4A and eGFP-AKMT-WT expressing parasites.

Accession code

Coordinates and structure factors have been deposited in the Protein Data Bank (PDB) under accession code 6FND.

Acknowledgements

The authors are grateful to the staff at the beamlines of ID29 and BM29 at the European Synchrotron Radiation Facility (ESRF) for their help with X-ray diffraction and SAXS data collections. We greatly appreciate B. Morriswood for critical reading of the manuscript. This work was supported by the grant P23440-B20 from the Austrian Science Fund (FWF) to GD, and funding from the National Institutes of Health/National Institute of Allergy and Infectious Diseases (R01-AI098686) and the March of Dimes (6-FY15-198) awarded to KH. YP is enrolled in the Integrative Structural Biology PhD program funded by the FWF.

bioRxiv preprint doi: https://doi.org/10.1101/264291; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Author Contributions

G. Dong and K. Hu conceived the project. Y. Pivovarova and G. Dong designed the experiments and performed analyses for all structural studies. J. Liu and K. Hu conducted the *in vitro* methylation and *in vivo* egress assays. J. Lesigang, O. Koldyda, and R. Rauschmeier assisted in generating preliminary data to prove the feasibility of the project. Y. Pivovarova, J. Liu, K. Hu, and G. Dong wrote the manuscript.

Conflict of interest

The authors declare that they have no conflict of interest.

References

1. Santos-Rosa, H., et al., *Active genes are tri-methylated at K4 of histone H3.* Nature, 2002. **419**(6905): p. 407-11.
2. Schaft, D., et al., *The histone 3 lysine 36 methyltransferase, SET2, is involved in transcriptional elongation.* Nucleic Acids Res, 2003. **31**(10): p. 2475-82.
3. Krogan, N.J., et al., *Methylation of histone H3 by Set2 in Saccharomyces cerevisiae is linked to transcriptional elongation by RNA polymerase II.* Mol Cell Biol, 2003. **23**(12): p. 4207-18.
4. Huang, J. and S.L. Berger, *The emerging field of dynamic lysine methylation of non-histone proteins.* Curr Opin Genet Dev, 2008. **18**(2): p. 152-8.
5. Yang, X.D., A. Lamb, and L.F. Chen, *Methylation, a new epigenetic mark for protein stability.* Epigenetics, 2009. **4**(7): p. 429-33.
6. Heaslip, A.T., et al., *The motility of a human parasite, Toxoplasma gondii, is regulated by a novel lysine methyltransferase.* PLoS Pathog, 2011. **7**(9): p. e1002201.
7. Li, M.M., et al., *ALKBH4-dependent demethylation of actin regulates actomyosin dynamics.* Nat Commun, 2013. **4**: p. 1832.
8. Cui, L., et al., *Histone lysine methyltransferases and demethylases in Plasmodium falciparum.* Int J Parasitol, 2008. **38**(10): p. 1083-97.
9. (WHO), W.H.O., *World Malaria Report 2014*. 2014, WHO: Geneva.
10. Luft, B.J., et al., *Toxoplasmic encephalitis in patients with the acquired immunodeficiency syndrome. Members of the ACTG 077p/ANRS 009 Study Team.* N Engl J Med, 1993. **329**(14): p. 995-1000.
11. Luft, B.J. and J.S. Remington, *Toxoplasmic encephalitis in AIDS.* Clin Infect Dis, 1992. **15**(2): p. 211-22.
12. Paquet, C., et al., *Toxoplasmosis in pregnancy: prevention, screening, and treatment.* J Obstet Gynaecol Can, 2013. **35**(1): p. 78-81.

bioRxiv preprint doi: https://doi.org/10.1101/264291; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

13. Luft, B.J., et al., *Toxoplasmic encephalitis in patients with the acquired immunodeficiency syndrome. Members of the ACTG 077p/ANRS 009 Study Team.* New Eng. J. Med., 1993. **329**(14): p. 995-1000.
14. Zuckerman, A., *Recent studies on factors involved in malarial anemia.* Military Medicine, 1966. **131**(9): p. Suppl 1201-216.
15. Nagel, R.L., *Malarial anemia.* Hemoglobin, 2002. **26**(4): p. 329-43.
16. Ehrhardt, S., et al., *Malaria, anemia, and malnutrition in african children--defining intervention priorities.* Journal of Infectious Diseases, 2006. **194**(1): p. 108-14.
17. Katris, N.J., et al., *The apical complex provides a regulated gateway for secretion of invasion factors in Toxoplasma.* PLoS Pathog, 2014. **10**(4): p. e1004074.
18. Nagayasu, E., et al., *Loss of a doublecortin (DCX)-domain protein causes structural defects in a tubulin-based organelle of Toxoplasma gondii and impairs host-cell invasion.* Mol Biol Cell, 2017. **28**(3): p. 411-428.
19. Stadler, R.V., et al., *Direct measurement of cortical force generation and polarization in a living parasite.* Mol Biol Cell, 2017. **28**(14): p. 1912-1923.
20. Dillon, S.C., et al., *The SET-domain protein superfamily: protein lysine methyltransferases.* Genome biology, 2005. **6**(8): p. 227.
21. Cheng, X., R.E. Collins, and X. Zhang, *Structural and sequence motifs of protein (histone) methylation enzymes.* Annu Rev Biophys Biomol Struct, 2005. **34**: p. 267-94.
22. Sivagurunathan, S., et al., *Identification of Functional Modules in AKMT, a novel lysine methyltransferase that regulates Toxoplasma gondii motility.* Molecular & Biochemical Parasitology, 2013. **189**(1-2): p. 43-53.
23. Krissinel, E. and K. Henrick, *Inference of macromolecular assemblies from crystalline state.* J Mol Biol, 2007. **372**(3): p. 774-97.
24. Greenfield, N.J., *Using circular dichroism spectra to estimate protein secondary structure.* Nat Protoc, 2006. **1**(6): p. 2876-90.
25. Malik, K., et al., *Differential scanning fluorimetry: rapid screening of formulations that promote the stability of reference preparations.* J Pharm Biomed Anal, 2013. **77**: p. 163-6.
26. Letunic, I. and P. Bork, *20 years of the SMART protein domain annotation resource.* Nucleic Acids Res, 2017.
27. Rozovskaia, T., et al., *Self-association of the SET domains of human ALL-1 and of Drosophila TRITHORAX and ASH1 proteins.* Oncogene, 2000. **19**(3): p. 351-7.
28. Manzur, K.L., et al., *A dimeric viral SET domain methyltransferase specific to Lys27 of histone H3.* Nat Struct Biol, 2003. **10**(3): p. 187-96.
29. Wei, H. and M.M. Zhou, *Dimerization of a viral SET protein endows its function.* Proc Natl Acad Sci U S A, 2010. **107**(43): p. 18433-8.
30. Zhang, Z., et al., *The APC/C subunit Cdc16/Cut9 is a contiguous tetratricopeptide repeat superhelix with a homo-dimer interface similar to Cdc27.* EMBO J, 2010. **29**(21): p. 3733-44.
31. Yip, Y.Y., et al., *The light chains of kinesin-1 are autoinhibited.* Proc Natl Acad Sci U S A, 2016. **113**(9): p. 2418-23.
32. Jinek, M., et al., *The superhelical TPR-repeat domain of O-linked GlcNAc transferase exhibits structural similarities to importin alpha.* Nat Struct Mol Biol, 2004. **11**(10): p. 1001-7.
33. Mosca, R. and T.R. Schneider, *RAPIDO: a web server for the alignment of protein structures in the presence of conformational changes.* Nucleic Acids Res, 2008. **36**(Web Server issue): p. W42-6.
34. Liu, Y., et al., *Structural basis for recognition of SMRT/N-CoR by the MYND domain and its contribution to AML1/ETO's activity.* Cancer Cell, 2007. **11**(6): p. 483-97.

bioRxiv preprint doi: https://doi.org/10.1101/264291; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

35. Spadaccini, R., et al., *Structure and functional analysis of the MYND domain.* J Mol Biol, 2006. **358**(2): p. 498-508.
36. Spellmon, N., et al., *Structure and function of SET and MYND domain-containing proteins.* Int J Mol Sci, 2015. **16**(1): p. 1406-28.
37. Hamamoto, R., et al., *SMYD3 encodes a histone methyltransferase involved in the proliferation of cancer cells.* Nat Cell Biol, 2004. **6**(8): p. 731-40.
38. Brown, M.A., et al., *Identification and characterization of Smyd2: a split SET/MYND domain-containing histone H3 lysine 36-specific methyltransferase that interacts with the Sin3 histone deacetylase complex.* Mol Cancer, 2006. **5**: p. 26.
39. Sajjad, A., et al., *Lysine methyltransferase Smyd2 suppresses p53-dependent cardiomyocyte apoptosis.* Biochim Biophys Acta, 2014. **1843**(11): p. 2556-62.
40. Cho, H.S., et al., *RB1 methylation by SMYD2 enhances cell cycle progression through an increase of RB1 phosphorylation.* Neoplasia, 2012. **14**(6): p. 476-86.
41. Mazur, P.K., et al., *SMYD3 links lysine methylation of MAP3K2 to Ras-driven cancer.* Nature, 2014. **510**(7504): p. 283-7.
42. Kunizaki, M., et al., *The lysine 831 of vascular endothelial growth factor receptor 1 is a novel target of methylation by SMYD3.* Cancer Res, 2007. **67**(22): p. 10759-65.
43. Jiang, Y., et al., *Structural insights into estrogen receptor alpha methylation by histone methyltransferase SMYD2, a cellular event implicated in estrogen signaling regulation.* J Mol Biol, 2014. **426**(20): p. 3413-25.
44. Van Aller, G.S., et al., *Structure-Based Design of a Novel SMYD3 Inhibitor that Bridges the SAM-and MEKK2-Binding Pockets.* Structure, 2016. **24**(5): p. 774-781.
45. Fu, W., et al., *Structural Basis for Substrate Preference of SMYD3, a SET Domain-containing Protein Lysine Methyltransferase.* J Biol Chem, 2016. **291**(17): p. 9173-80.
46. Jacot, D., et al., *An Apicomplexan Actin-Binding Protein Serves as a Connector and Lipid Sensor to Coordinate Motility and Invasion.* Cell Host Microbe, 2016. **20**(6): p. 731-743.
47. Kafsack, B.F., et al., *Rapid membrane disruption by a perforin-like protein facilitates parasite exit from host cells.* Science, 2009. **323**(5913): p. 530-3.
48. Tyagi, R., R. Lai, and R.G. Duggleby, *A new approach to 'megaprimer' polymerase chain reaction mutagenesis without an intermediate gel purification step.* BMC Biotechnol, 2004. **4**: p. 2.
49. Kabsch, W., *Xds.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 2): p. 125-32.
50. Adams, P.D., et al., *PHENIX: a comprehensive Python-based system for macromolecular structure solution.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 2): p. 213-21.
51. Emsley, P. and K. Cowtan, *Coot: model-building tools for molecular graphics.* Acta Crystallogr D Biol Crystallogr, 2004. **60**(Pt 12 Pt 1): p. 2126-32.
52. Chen, V.B., et al., *MolProbity: all-atom structure validation for macromolecular crystallography.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 1): p. 12-21.
53. Kornbrot, D., *Statistical software for microcomputers: SigmaPlot 2000 and SigmaStat2.* Br J Math Stat Psychol, 2000. **53 (Pt 2)**: p. 335-7.
54. Konarev, P.V., et al., *PRIMUS: a Windows PC-based system for small-angle scattering data analysis.* Journal of Applied Crystallography, 2003. **36**: p. 1277-1282.
55. Svergun, D.I. and M.H. Koch, *Advances in structure analysis using small-angle scattering in solution.* Curr Opin Struct Biol, 2002. **12**(5): p. 654-60.
56. Svergun, D.I., *Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing.* Biophys J, 1999. **76**(6): p. 2879-86.
57. Volkov, V.V. and D.I. Svergun, *Uniqueness of ab initio shape determination in small-angle scattering.* Journal of Applied Crystallography, 2003. **36**: p. 860-864.

bioRxiv preprint doi: https://doi.org/10.1101/264291; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

58. Kozin, M.B. and D.I. Svergun, *Automated matching of high- and low-resolution structural models.* Journal of Applied Crystallography, 2001. **34**: p. 33-41.
59. Svergun, D., C. Barberato, and M.H.J. Koch, *CRYSOL - A program to evaluate x-ray solution scattering of biological macromolecules from atomic coordinates.* Journal of Applied Crystallography, 1995. **28**: p. 768-773.
60. Liu, J., et al., *An ensemble of specifically targeted proteins stabilizes cortical microtubules in the human parasite Toxoplasma gondii.* Mol Biol Cell, 2016. **27**(3): p. 549-71.
61. Roos, D.S., et al., *Molecular tools for genetic dissection of the protozoan parasite Toxoplasma gondii.* Methods Cell Biol., 1994. **45**: p. 27-78.
62. Liu, J., et al., *Novel thioredoxin-like proteins are components of a protein complex coating the cortical microtubules of Toxoplasma gondii.* Eukaryot Cell, 2013.
63. Wallace, A.C., R.A. Laskowski, and J.M. Thornton, *LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions.* Protein Eng, 1995. **8**(2): p. 127-34.

bioRxiv preprint doi: https://doi.org/10.1101/264291; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure legends

Figure 1. Crystal structure of *T. gondii* AKMT. (**A**) Schematic showing the domain arrangement of AKMT. The folded part used for structure determination (aa290-709) contains four predicted domains. The AKMT-specific insertion is denoted as "ASI". (**B**) Structure of AKMT with the same color scheme as in (A). All α-helices and β-strands are labeled, and zinc ions are shown as black spheres. (**C**) Secondary structure diagram of AKMT with the post-SET and ASI areas highlighted. (**D**) A representative region of the 2Fo-Fc map contoured at 1.0 σ. (**E**) Electron density maps (1.0 σ) around the two bound zinc ions in AKMT.

Figure 2. AKMT is a homodimer. (**A**) The structural model of AKMT contains four molecules (A, B, C & D, differently colored) per asymmetric unit. Enlarged boxes show the two types of intermolecular contacts. These were disrupted separately to identify the native interaction, either by deleting the N-terminal extension (ΔNTD-short, aa309-709) or by mutating the four residues that form an extensive hydrogen bond network in the vertical dimer (ΔNTD-4A). (**B**) Superposition of the four molecules in the asymmetric unit, which shows a RMSD of ~0.2 Å for the aligned Cα atoms. The slight rigid-body shift of the N-termini (boxed) is shown in an enlarged view. (**C**) SLS data of full-length (FL), truncated, and mutated AKMT. Except for the ΔNTD-4A mutant (red), all other constructs form dimers. The asterisk (*) marks impurities co-purified with the full-length protein. (**D**) Low-angle region of the SAXS curve of AKMT-ΔNTD (black) overlaid with theoretical scattering curves for the N-terminally mediated A-C dimer (red) and the C-terminally mediated A-B dimer (blue) calculated in CRYSOL. The green curve represents the simulated scattering from *ab initio* model computed in DAMMIF. (**E**) The *ab initio* low resolution model reconstructed in DAMAVER by selecting and averaging the most probable models generated in 10 cycles of DAMMIF. The mean normalized spatial discrepancy (NSD) is 0.538. The N-terminal fragments (aa289-309) in the A-B dimer were rebuilt after rearranging the swapped domains. The model is shown superimposed on a ribbon diagram of the A-B dimer.

Figure 3. Intermolecular interactions in the AKMT homodimer. (**A**) Homodimer of AKMT with chain A shown in ribbon diagrams and chain B in a semi-transparent surface plot. Insets show the enlarged views of the two bound zinc ions in chain A. (**B**) Sequence alignment of the

bioRxiv preprint doi: https://doi.org/10.1101/264291; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

ASI motifs in AKMT and orthologs. The four conserved cysteines that coordinate zinc binding are indicated with open arrowheads. (**C**) Ribbon diagram of AKMT monomer with individual domains colored as in (A). (**D**) Electrostatic plot showing the hand-like structure with the C-terminal end of the protein forming the "thumb", the ASI motif forming the "fingers", and the central helices of the TPR-like domain forming the "palm". (**E**) "Handshake"-like dimer of AKMT. (**F**) Intermolecular interactions in the AKMT homodimer. Residues contributing to dimer formation are plotted with DIMPLOT in the LigPlot plus suite [63]. Side chains of residues involved in hydrogen bond formation are shown as ball-and-stick. Oxygen, nitrogen and carbon atoms are colored in red, blue, and black, respectively. Water molecules mediating inter-molecular hydrogen bond formation are shown as cyan-colored spheres. Green dotted lines indicate hydrogen bonds, with the length of the bonds shown on top of the lines. Non-bonded residues involved in hydrophobic interactions are shown as spoked arcs.

Figure 4. Cofactor binding and biophysical characterizations of AKMT. (**A**) ITC binding curves of wild-type full-length and AKMT-ΔNTD with the cofactor SAM or the analog SFG. (**B**) ITC binding curves of AKMT-ΔNTD-4A with SAM or SFG. (**C**) Pymol B-factor putty representation of the AKMT dimer. Width of the tubes, which is proportional to the values of the B-factors of each residue, indicates the dynamics of the local structure. Shown in the inset is an enlarged view of the highly rigid region around residue H447 that bridges the putative cofactor binding site to the zinc-bound ASI motif. The cofactor SAM (shown in yellow sticks) is based on superposition of SAM bound SMYD2 structure (PDB code: 5ARG) onto the AKMT structure. (**D**) Deconvoluted MS spectrum of AKMT-ΔNTD in the denatured state. The measured molecular mass (48202.54 Da) is in excellent agreement with the theoretical value calculated from the primary sequence of the protein (48202.76 Da). (**E**) Mass determination of SFG-bound AKMT-ΔNTD at the native state. Based on the calculated masses, each of the three peaks could be mapped to a subspecies of the complex, with the majority being the homodimer loaded with four zinc ions and two SFG molecules (2×SFG). (**F**) Collision-induced dissociation of native SFG-bound AKMT-ΔNTD analyzed by tandem mass spectrometry. Increasing trap collision energy (CE) led to dissociation of SFG from the dimer, but did not affect stability of the zinc-bound dimer. (**G**) CD spectra of AKMT-ΔNTD and AKMT-ΔNTD-4A. (**H**) Melting curves of various AKMT constructs by the DSF assay. AKMT-FL and -ΔNTD have a similar Tm, whereas that of the monomeric mutant is substantially lower.

bioRxiv preprint doi: https://doi.org/10.1101/264291; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 5. Functional defect of AKMT-ΔNTD-4A and AKMT-4A revealed by *in vitro* methylation and *in vivo* egress assays. (**A**) Methylation assay: The top panels show total protein (amido black staining), and the bottom panels show tritium incorporation into histone H3.3. A 1:2 serial dilutions of enzyme (left panels: 2^{0} -2^{-6}; right panels: 2^{-7} -2^{-12} dilutions) from ~7.6-7.8 μM were used. Each reaction contains 4.1 μM H3.3 as the artificial substrate. Note that the methylation level of AKMT-ΔNTD-4A at 7.8×2^{-8} μM (red arrow) is lower than that in the AKMT-ΔNTD reaction with 7.6×2^{-11} μM enzyme (orange arrow) (**B**) Images selected from time-lapse imaging of calcium ionophore (A23187)-induced egress of Δ*akmt* parasites expressing eGFP-AKMT-WT (top panels) or eGFP-AKMT-4A (bottom panels). Insets (2×) include the apical region from which the eGFP-AKMT-WT or eGFP-AKMT-4A translocates after the addition of A23187. The elapsed time (min:sec) since A23187 addition is indicated on each panel. The parasites shown here initiated egress around 3:59 for eGFP-AKMT-WT- and 7:39 for eGFP-AKMT-4A-expressing parasites. Fluorescence and DIC/fluorescence overlay images are shown. The red dotted circle in the overlay indicates untransfected Δ*akmt* parasites. (**C**) Percentage of vacuoles that egressed within ~11 min of A23187 addition. WT: Δ*akmt* parasites expressing eGFP-AKMT-WT. 4A: Δ*akmt* parasites expressing eGFP-AKMT-4A. (**D**) Box plot of the egress response time for vacuoles containing eGFP-AKMT-WT or eGFP-AKMT-4A transfected parasites that did egress. For C and D, the vacuoles analyzed were from 6 independent transfections for eGFP-AKMT-WT and 7 for eGFP-AKMT-4A expression.

Figure 6. AKMT forms a novel homodimer mediated by both its ASI motif and TPR-like domain. (**A**) Ribbon diagram of the vSET protein color-ramped from blue at the N-terminus to red at the C-terminus. (**B**) Side-by-side arrangement of the SET domains in the vSET homodimer. (**C**) Crystal structure of the AKMT homodimer. (**D**) An extracted view of the dimeric interface formed by the ASI motif and the TPR-like domain of AKMT. (**E**) The same structure as in (D) but with both ASI motifs deleted. The two TPR-like domains are shown in rainbow coloration to demonstrate the extensive contacts across the two antiparallelly arranged structures. Notably, without ASI motifs, the buried interface area is reduced by ~50%, and in the meantime $\Delta^{i}G$ increases significantly, suggesting a substantial drop in structural stability.

bioRxiv preprint doi: https://doi.org/10.1101/264291; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 7. Structural comparison of AKMT with SMYD proteins. (**A**) Schematics of the domain organization of AKMT and SMYD1-3. Lengths of the boxes are proportional to the sizes of the domains. (**B**) Ribbon diagrams of AKMT, SMYD1 (3N71.pdb), SMYD2 (5ARG.pdb), and SMYD3 (5EX0.pdb), in the same color scheme as in (A). All bound zinc ions are shown as black spheres. The putative cofactor and substrate binding sites in AKMT are marked by a red star and a black arrow, respectively. (**C**) Direct structural comparison between AKMT and SMYD3, with the "stirring" loop in AKMT colored in blue and the counterpart in SMYD3 in yellow. Superposition of AKMT on SMYD3 is based on the rigid body around their post-SET domains (see Fig. S7B-E). The "stirring" loops are better visualized in the enlarged view where potential clashes between the loop of AKMT and the MYND domain of SMYD3 are indicated by a dashed circle. The loops are also shown in isolation next to the inset to clearly illustrate the dramatic flip (black arrow). (**D**) Close-up view of the "stirring" loops in AKMT and SMYD1-3 structures, with regular β strands shown in flat sheets and irregular structures in loops. SMYD1, 2, 3, and AKMT are colored green, magenta, yellow, and blue, respectively. (**E**) Sequence alignment of the "stirring" loops of AKMT orthologs (boxed) and SMYD1-3. (**F**) Overlay of the AKMT and SMYD1-3 structures to demonstrate the variation of their inter-lobe grooves. The structures are superimposed on the C-terminal rigid body (see Fig. S7F-H).

Figure 8. Comparison of dimeric AKMT with monomeric SMYD1-3. (**A**) Structure of the AKMT dimer with both subunits shown as rainbow-colored ribbons. (**B-E**) Ribbon diagrams with color ramp along the sequences (left), and two orthogonal views (right) of the electrostatic plots of AKMT monomer (B), SMYD1 (C), SMYD2 (D), and SMYD3 (E). In contrast to the structurally and electrostatically complementary hand-like structure in AKMT, the corresponding surfaces in all SMYD proteins are unfavorable for intermolecular interactions.

bioRxiv preprint doi: https://doi.org/10.1101/264291; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

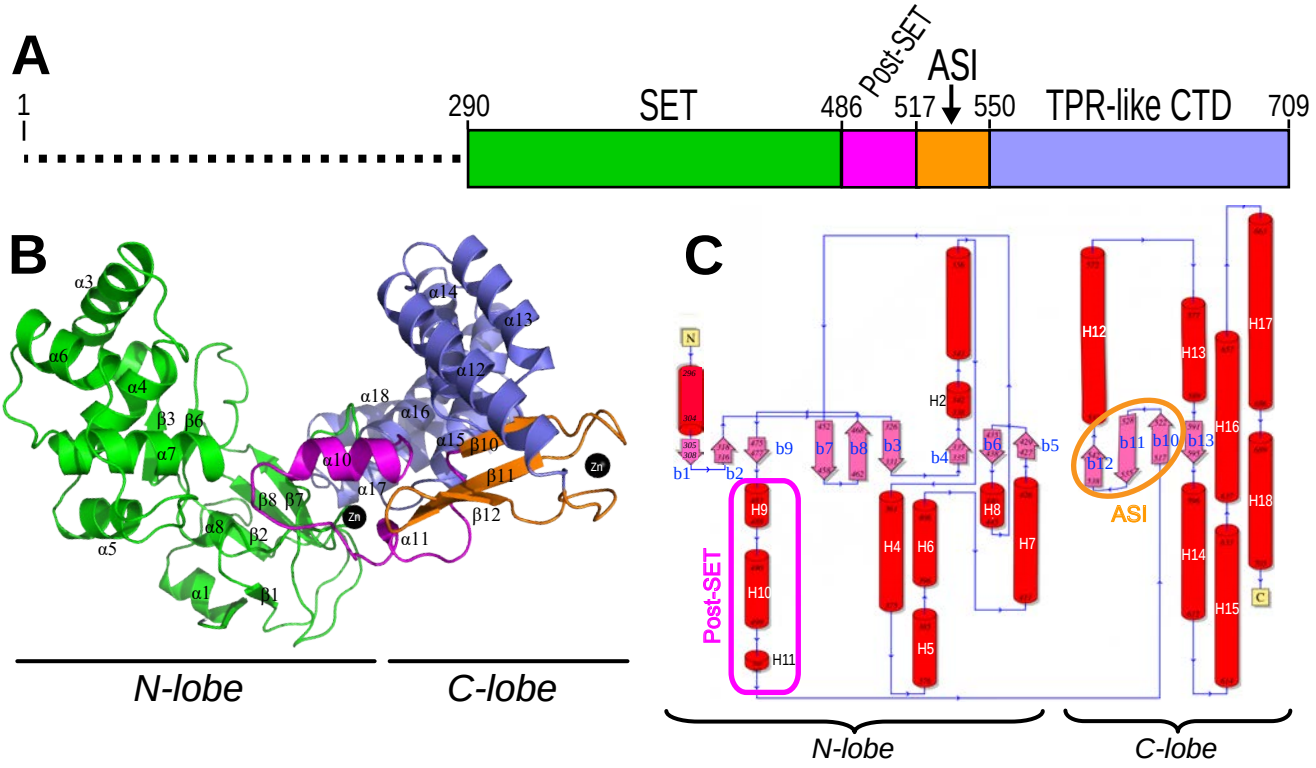
Table 1 Data collection and refinement statistics

Data collection	
Space group	P1
Wavelength (Å)	1.282
Cell dimensions	
a, *b*, *c* (Å)	61.94, 89.37, 91.72
α, β, γ (°)	108.69, 101.32, 103.41
Resolution (Å)	20-2.1 (2.23-2.10) *
No. unique reflections	167,741 (14,053)
R_{meas}	0.138(0.967)
CC(1/2)	99.9 (79.0)
I / σI	15.5 (2.2)
Overall completeness (%)	82.8 (43.2)
Overall redundancy	2.8 (2.7)
Refinement	
Resolution (Å)	20-2.1
Number of reflections	167,722
R_{work} / R_{free} (%)	19.1/22.6
No. atoms	
Protein	14,182
Ligand or ion	139
Water	434
R.m.s. deviations	
Bond lengths (Å)	0.004
Bond angles (°)	0.740
Ramachandran plot	
Favored (%)	96.94
Allowed (%)	3.06
Outlier (%)	0

*Values in parentheses are for highest-resolution shell.

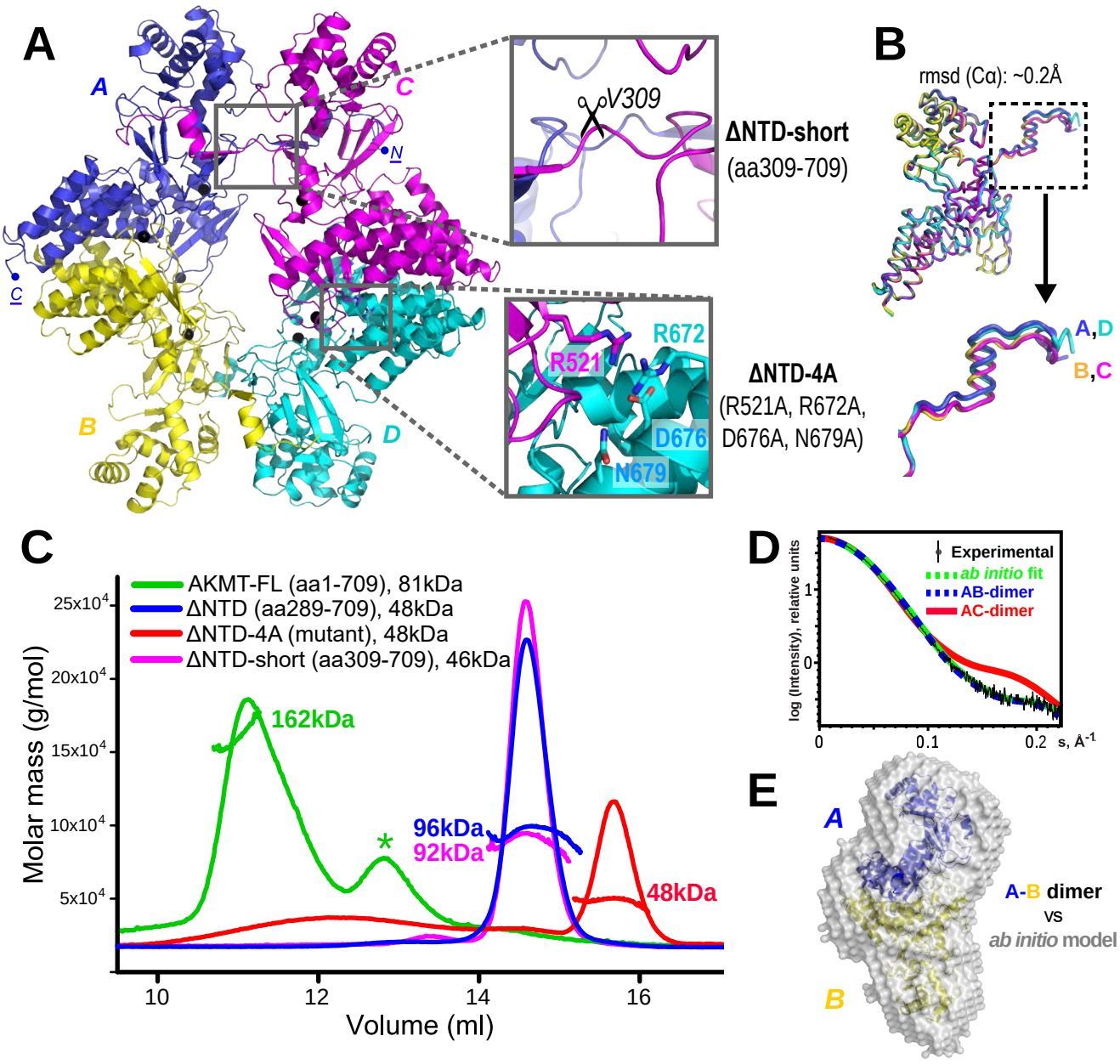
bioRxiv preprint doi: https://doi.org/10.1101/264291; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Fig 1



bioRxiv preprint doi: https://doi.org/10.1101/264291; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Fig 2



bioRxiv preprint doi: https://doi.org/10.1101/264291; this version posted February 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Fig 3

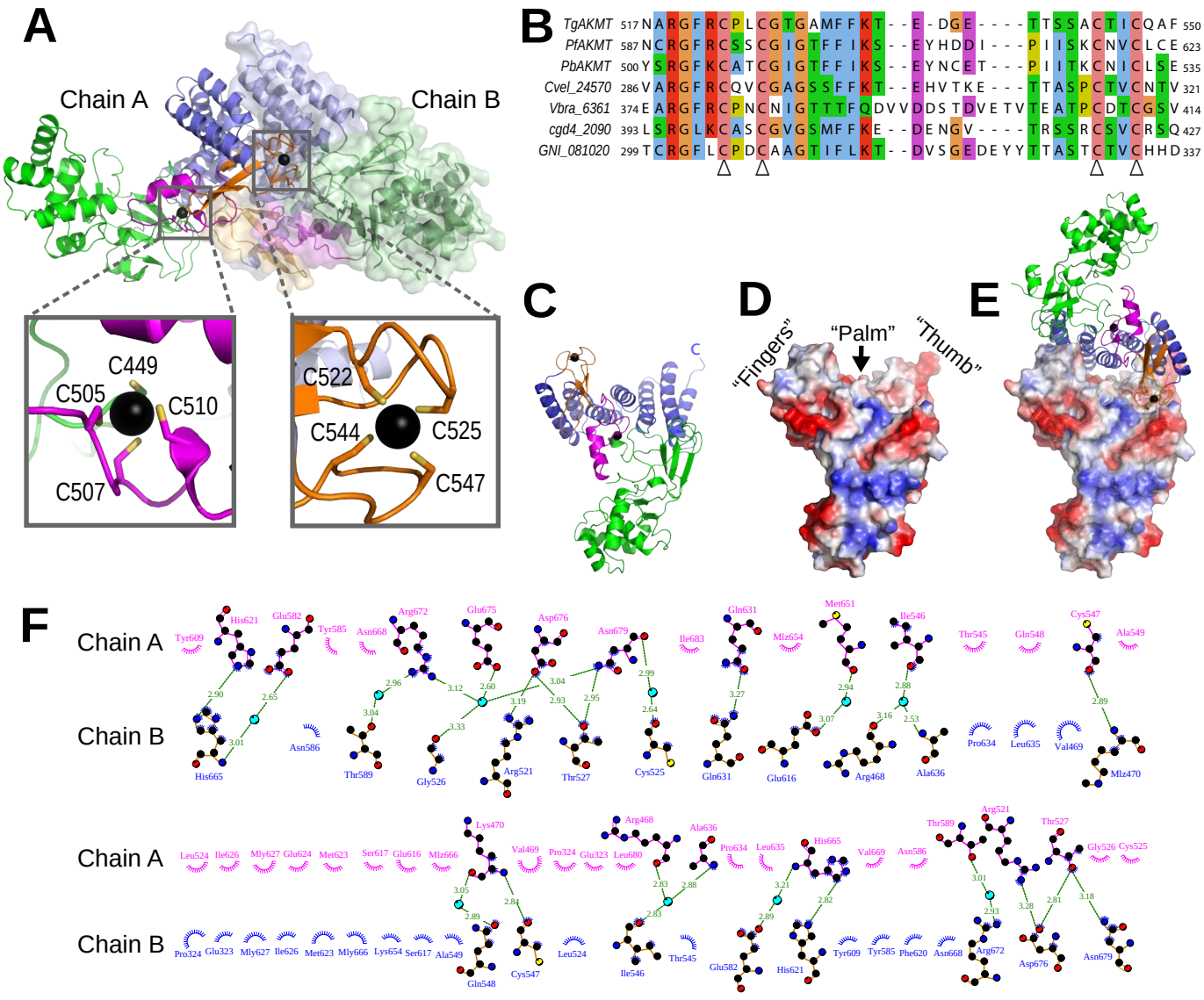
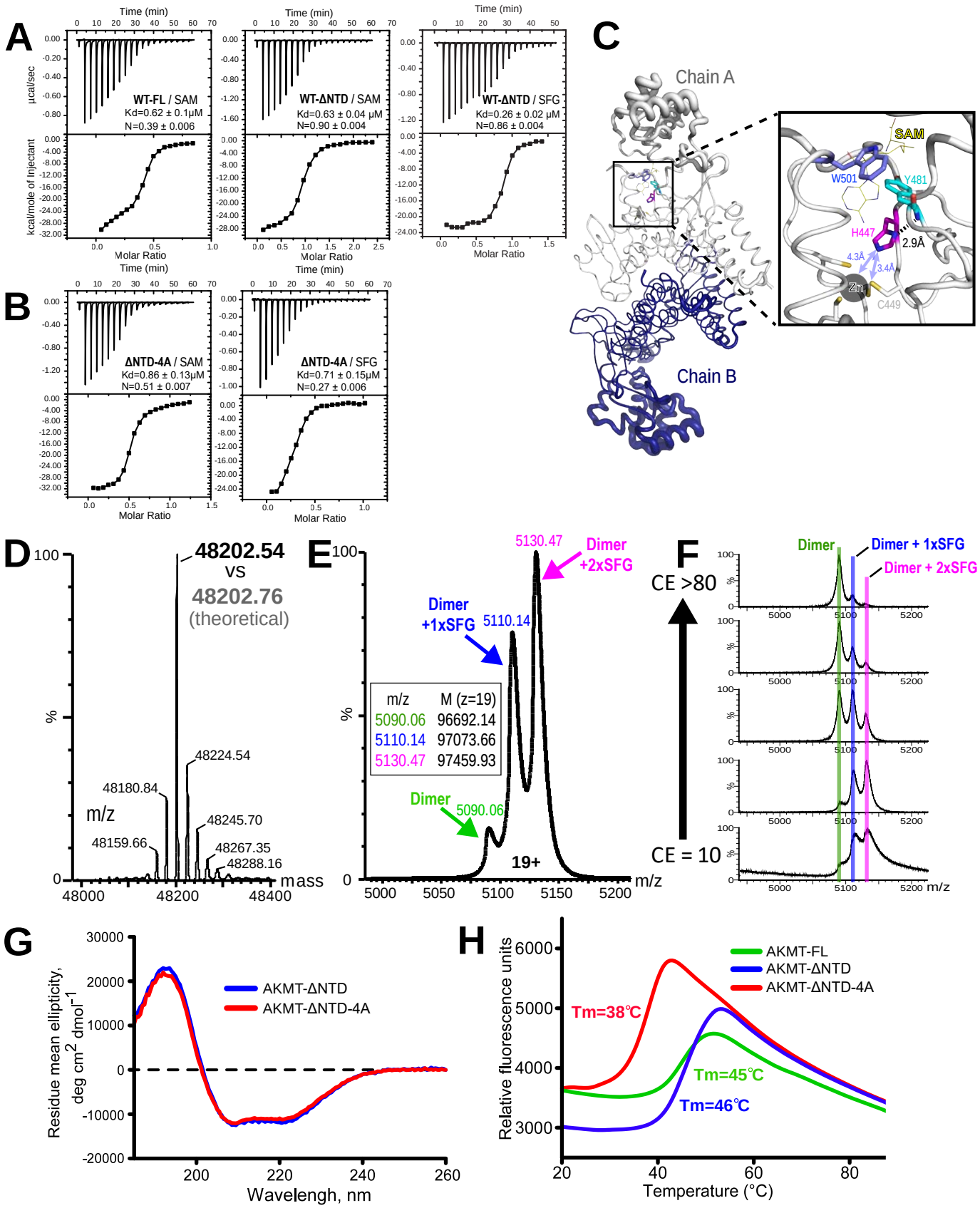


Fig 4


A
Time (min)
µcal/sec
kcal/mole of Injectant
WT-FL / SAM
Kd=0.62 ± 0.1µM
N=0.39 ± 0.006
WT-ΔNTD / SAM
Kd=0.63 ± 0.04 µM
N=0.90 ± 0.004
WT-ΔNTD / SFG
Kd=0.26 ± 0.02 µM
N=0.86 ± 0.004
Molar Ratio
B
ΔNTD-4A / SAM
Kd=0.86 ± 0.13µM
N=0.51 ± 0.007
ΔNTD-4A / SFG
Kd=0.71 ± 0.15µM
N=0.27 ± 0.006
C
Chain A
Chain B
SAM
W501
Y481
H447
2.9Å
4.3Å
3.4Å
Zn
C449
D
48202.54
VS
48202.76
(theoretical)
48224.54
48180.84
48245.70
48159.66
48267.35
48288.16
m/z
mass
E
5130.47
Dimer
+2xSFG
Dimer
+1xSFG
5110.14
Dimer
5090.06
19+
m/z
M (z=19)
5090.06 96692.14
5110.14 97073.66
5130.47 97459.93
F
CE >80
CE = 10
Dimer
Dimer + 1xSFG
Dimer + 2xSFG
m/z
G
Residue mean ellipticity, deg cm2 dmol-1
AKMT-ΔNTD
AKMT-ΔNTD-4A
Wavelengh, nm
H
Relative fluorescence units
AKMT-FL
AKMT-ΔNTD
AKMT-ΔNTD-4A
Tm=38℃
Tm=45℃
Tm=46℃
Temperature (°C)

available under aCC-BY-NC-ND 4.0 International license.

Fig 5

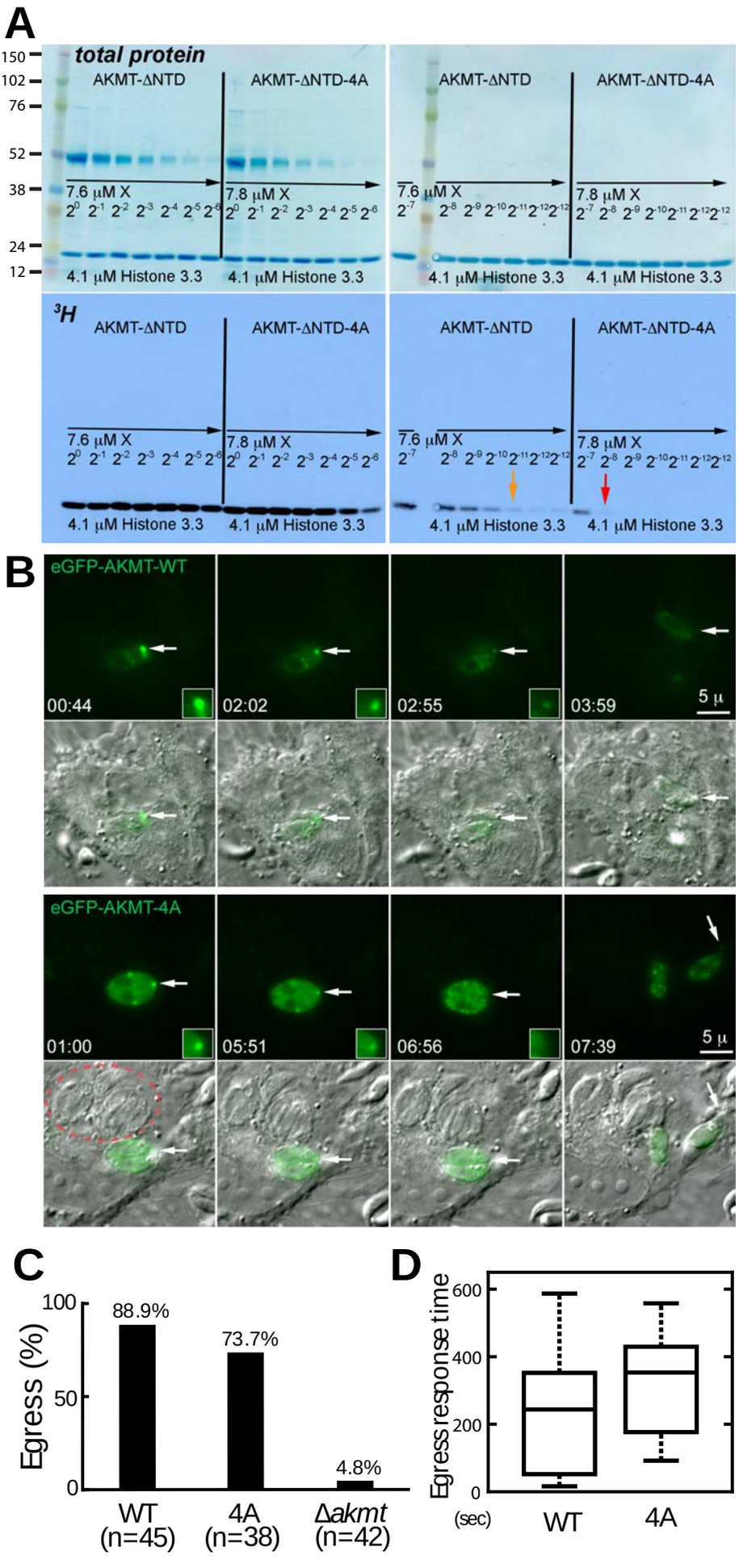


Fig 6

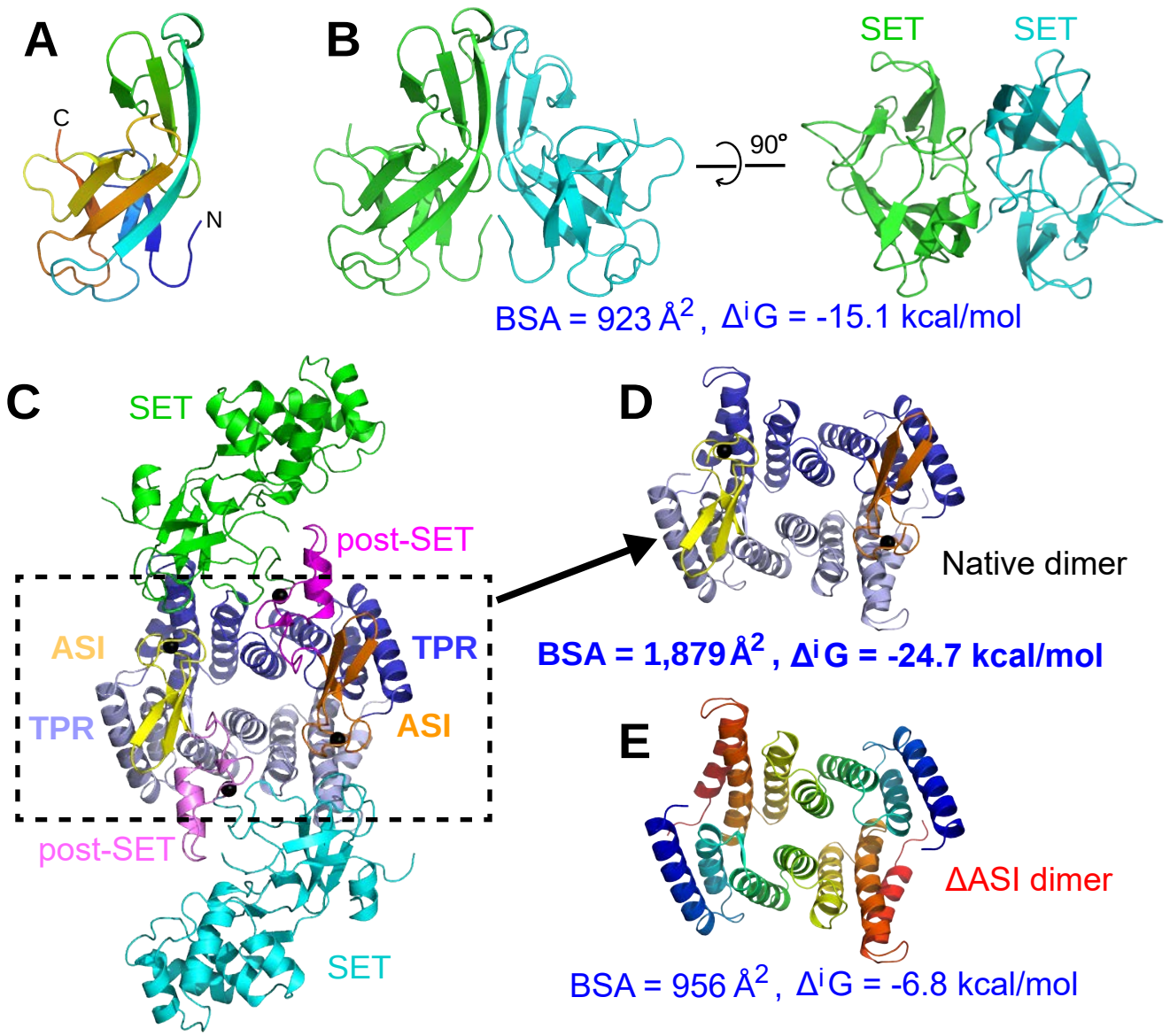

A
C
N
B
SET
SET
90°
BSA = 923 Å^2 , ΔiG = -15.1 kcal/mol
C
SET
post-SET
ASI
TPR
TPR
ASI
post-SET
SET
D
Native dimer
BSA = 1,879 Å^2 , ΔiG = -24.7 kcal/mol
E
ΔASI dimer
BSA = 956 Å^2 , ΔiG = -6.8 kcal/mol

Fig 7

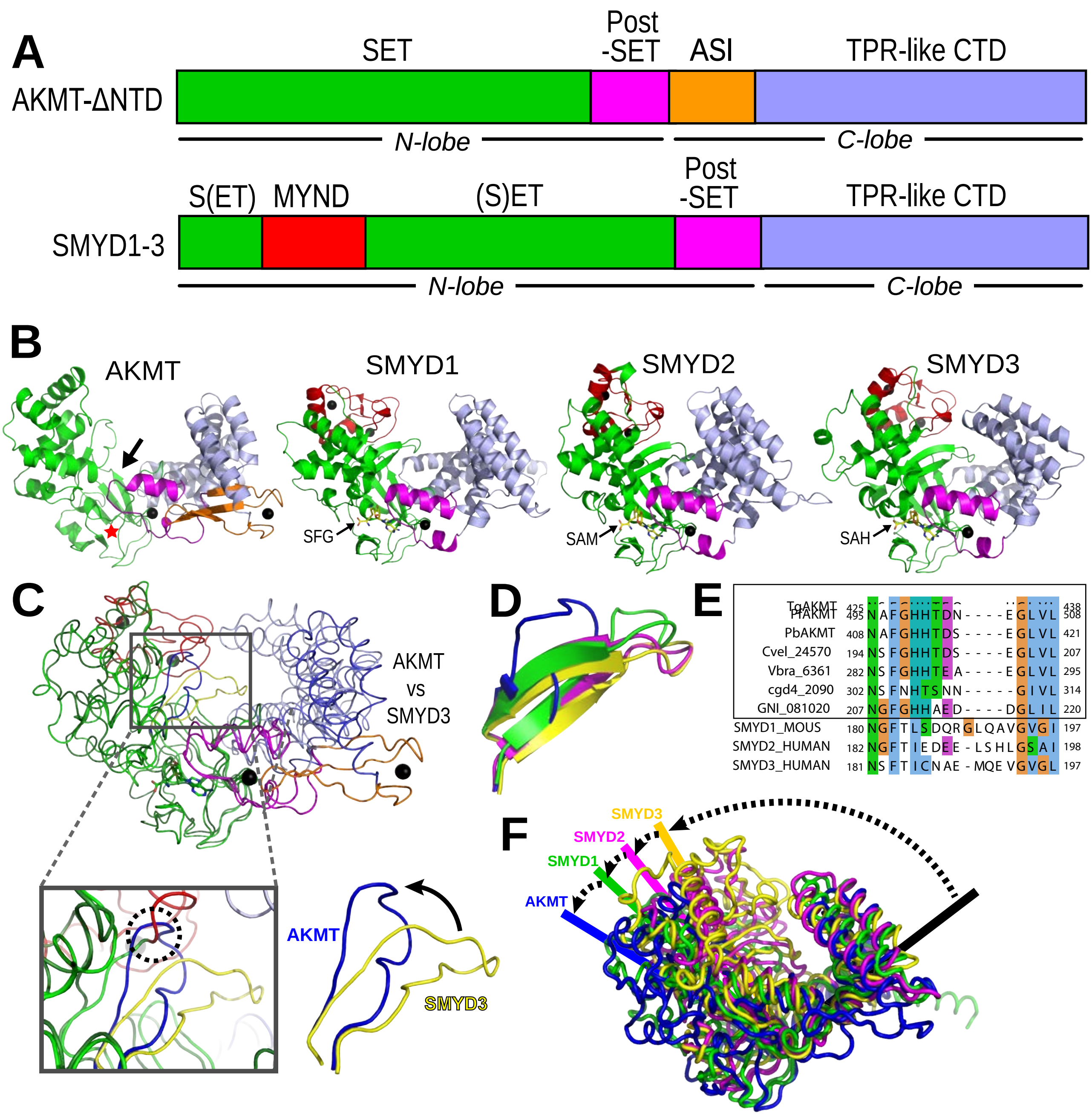

A
SET
Post -SET
ASI
TPR-like CTD
AKMT-ΔNTD
N-lobe
C-lobe
S(ET)
MYND
(S)ET
Post -SET
TPR-like CTD
SMYD1-3
N-lobe
C-lobe
B
AKMT
SMYD1
SMYD2
SMYD3
SFG
SAM
SAH
C
AKMT vs SMYD3
AKMT
SMYD3
D
E
PfAKMT 495 NAFGHHTDN----EGLVL 508
PbAKMT 408 NAFGHHTDS----EGLVL 421
Cvel_24570 194 NSFGHHTDS----EGLVL 207
Vbra_6361 282 NSFGHHTEA----EGLVL 295
cgd4_2090 302 NSFNHTSNN-----GIVL 314
GNI_081020 207 NGFGHHAED----DGLIL 220
SMYD1_MOUS 180 NGFTLSDQRGLQAVGVGI 197
SMYD2_HUMAN 182 NGFTIEDEE-LSHLGSAI 198
SMYD3_HUMAN 181 NSFTICNAE-MQEVGVGL 197
F
SMYD3
SMYD2
SMYD1
AKMT

Fig 8