Discovery of Unconventional Kinetochores in Kinetoplastids

Bungo Akiyoshi1,2,* and Keith Gull1

1Sir William Dunn School of Pathology, University of Oxford, Oxford OX1 3RE, UK
2Department of Biochemistry, University of Oxford, Oxford OX1 3QU, UK
*Correspondence: bungo.akiyoshi@bioch.ox.ac.uk
http://dx.doi.org/10.1016/j.cell.2014.01.049
This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

SUMMARY

The kinetochore is the macromolecular protein complex that directs chromosome segregation in eukaryotes. It has been widely assumed that the core kinetochore consists of proteins that are common to all eukaryotes. However, no conventional kinetochore components have been identified in any kinetoplastid genome, thus challenging this assumption of universality. Here, we report the identification of 19 kinetochore proteins (KKT1–19) in Trypanosoma brucei. The majority is conserved among kinetoplastids, but none of them has detectable homology to conventional kinetochore proteins. These proteins instead have a variety of features not found in conventional kinetochore proteins. We propose that kinetoplastids build kinetochores using a distinct set of proteins. These findings provide important insights into the longstanding problem of the position of the root of the eukaryotic tree of life.

INTRODUCTION

Faithful transmission of genetic material is essential for the survival of all organisms. Eukaryotic chromosome segregation is driven by the kinetochore, a macromolecular protein complex that assembles onto centromeric DNA and captures spindle microtubules to govern the movement of chromosomes (Cheeseman and Desai, 2008; Santaguida and Musacchio, 2009). Kinetochores consist of more than 40 different components even in the simple yeast kinetochore (Biggins, 2013) and are recognized as one of the most complex structures in the cell. A hallmark of eukaryotic kinetochores is the centromere-specific histone H3 variant (CENP-A), which specifies the site of kinetochore assembly by creating a specialized chromatin environment (Hori and Fukagawa, 2012; Maddox et al., 2012; Westhorpe and Straight, 2013). Putative CENP-A homologs can be identified in nearly all sequenced eukaryotes (Talbert et al., 2009), suggesting that most eukaryotes utilize CENP-A to assemble kinetochores. However, notable exceptions are found in the kinetoplastid species, a group of unicellular flagellated eukaryotes, including parasitic Trypanosomatidae (e.g., Trypanosoma brucei, Trypanosoma cruzi, and Leishmania species) and free-living Bodonida (e.g., Bodo saltans). Their genome sequences have so far failed to reveal any CENP-A homolog (Lowell and Cross, 2004; Berriman et al., 2005), suggesting that kinetoplastids assemble kinetochores without CENP-A, possibly using a distinct set of proteins. Consistent with this possibility, bioinformatic analyses have failed to detect any conventional kinetochore protein homolog in kinetoplastids, whereas at least a fraction of kinetochore components can be readily identified in other diverse eukaryotes (Meraldi et al., 2006; Westermann and Schleiffer, 2013) (Table S1 available online). By comparison, kinetoplastids possess the CDK/cyclin system, cohesion complex, separase, condensin complex, Aurora B, the anaphase promoting complex (APC/C), and proteasomes, suggesting that the most basic cell-cycle machinery is conserved in these distant eukaryotes (Akiyoshi and Gull, 2013).

T. brucei is the causative agent of devastating African sleeping sickness in humans and nagana in livestock. In addition to 11 homologous pairs of large chromosomes (also called megabase chromosomes, 1–6 Mb in size), T. brucei possesses ~100 small chromosomes (minichromosomes, 50–150 kb; intermediate chromosomes, 200–700 kb) (Daniels et al., 2010; Ersfeld, 2011). Previous studies suggest that megabase chromosomes contain regional centromeres, whereas intermediate or minichromosomes may not contain canonical centromeres (Obado et al., 2007). The core of minichromosomes consists of 177 bp repeats and is constructed in a palindromic manner (Wickstead et al., 2004). Although minichromosomes do not possess housekeeping genes, they are crucial for increasing the capacity of antigenic variation (Sloof et al., 1983) and individual minichromosomes appear to segregate faithfully at each cell division (Wickstead et al., 2003). T. brucei undergoes a closed mitosis and forms a mitotic spindle within the nucleus (Ogbadoyo et al., 2000), and segregation of both megabase chromosomes and minichromosomes depends on spindle microtubules (Ersfeld and Gull, 1997). Ultrastructural studies have detected kinetochore-like electron-dense plaques that appear to form end-on attachments to spindle microtubules in mitotic cells (Ogbadoyo et al., 2000). Through blocking the accurate segregation of these chromosomes, cell growth or immune evasion could be inhibited. Understanding the underlying molecular mechanism is therefore critical to developing treatment strategies against kinetoplastid...
diseases. Furthermore, there is a great interest in understanding how kinetochores can be assembled in the absence of a CENP-A homolog in kinetoplastids. Identification of kinetochore proteins is an essential step toward both of these goals. Here, we describe the identification of 19 kinetochore proteins in *T. brucei*.

**RESULTS**

**Identification of KKT1 in *T. brucei***

*T. brucei* possesses two DNA-containing organelles, the nucleus and the kinetoplast. The former contains nuclear DNA, whereas the latter contains a cluster of mitochondrial DNA. These organelles have distinct replication and segregation timings and serve as good cell-cycle markers (Woodward and Gull, 1990; Siegel et al., 2008). To identify proteins that are relevant for mitosis, we carried out a yellow fluorescent protein (YFP)-tagging screen to examine the localization of uncharacterized proteins whose transcript levels are upregulated later during the cell cycle (Archer et al., 2011). This screen identified a protein (ORF Tb927.10.6330) that has a localization pattern characteristic of kinetochore proteins (Figure 1A). There is little YFP signal in G1, and dots appear in the nucleus around S phase, align at the center of the nucleus in metaphase, and then move to opposite poles and localize near the leading edge of separating chromosomes during anaphase (Figure 1B). The protein is well conserved among kinetoplastids (Table 1), so we named it KKT1 for kinetoplastid kinetochore protein 1.

**Identification of KKT2–19***

To identify more kinetochore proteins, we affinity purified a YFP-tagged version of KKT1 (Figure S1) and identified the copurifying proteins by mass spectrometry (MS) (Table S2). Twelve uncharacterized proteins were identified that copurified with KKT1 in an apparently specific manner. We tagged these proteins with YFP and found that they all have kinetochore-like localization patterns (see below). We therefore named them KKT2–13. We then looked for more kinetochore proteins by the affinity purification/MS of YFP-tagged versions of these 12 proteins (Figure S1 and Table S2), followed by the YFP-tagging of candidate proteins, which led to the identification of six additional kinetochore proteins (KKT14–19). Affinity purification/MS of these six proteins (Figure S1 and Table S2) failed to identify any more kinetochore proteins, indicating that the approach had reached saturation. Although there may be more kinetochore proteins still unidentified, we began to characterize the 19 KKT proteins identified from this methodology.
KKT Proteins Are Conserved in Kinetoplastids

The majority of KKT proteins appear to be well conserved among kinetoplastid species (Table 1). However, homology search programs using position-specific iterated (PSI)-BLAST (Altschul et al., 1997) or hidden Markov models (Eddy, 1998) failed to identify homologous proteins in other organisms except for proteins that contain shared conserved domains (see below). Furthermore, we could not find any significant homology between KKT proteins and conventional kinetochore proteins. These results raise the possibility that the KKT proteins constitute an unconventional kinetochore unique to kinetoplastids, which is in line with the absence of CENP-A. Interestingly, conservation of KKT proteins in those kinetoplastids that do not possess intermediate or minichromosomes (Obado et al., 2007), and it remains unknown whether these chromosomes utilize the same segregation machinery (Gull et al., 1998).

To address these questions, we performed chromatin immunoprecipitation of YFP-tagged KKT proteins followed by deep-sequencing (ChIP-seq). We chose KKT2 and KKT3 because they have punctate signals throughout the cell cycle and therefore may directly bind DNA (see below). A histone H3 variant, H3v, was also analyzed for comparison. Sequencing reads were mapped to a reference genome that contains 11 megabase chromosomes, as well as a model minichromosome that mostly consists of the 177 bp repeats (see Experimental Procedures). The results were normalized based on the number of reads from each input sample, and we calculated enrichment ratios for nonoverlapping windows of 150 bp in size. Centromeres for chromosomes 9, 10, and 11 are not in the genome assembly, so we focused on chromosomes 1–8 and the model minichromosome.

We found that both YFP-KKT2 and YFP-KKT3 have a strong peak on each megabase chromosome that corresponds to the mapped centromeric region (Figures 2A and S2). YFP-H3v did not have specific enrichment at centromeric regions but was enriched at transcription termination sites as previously reported (Siegel et al., 2009) (Figure 2B). These results show that KKT2 and KKT3 are enriched at the identified centromeric regions in the megabase chromosomes and thus confirm that they are bona fide kinetochore proteins.

For seven out of eight megabase chromosomes (chromosomes 1, 2, and 4–8), the highly enriched regions correspond whether kinetochore assembly in fact occurs at the mapped regions, and if so, exactly where kinetochores are assembled. Site-specific topoisomerase II activity was not identified for the intermediate or minichromosomes (Obado et al., 2007), and it remains unknown whether these chromosomes utilize the same segregation machinery (Gull et al., 1998).

<table>
<thead>
<tr>
<th>Name</th>
<th>T. brucei Accession</th>
<th>T. cruzi Accession</th>
<th>T. congolense Accession</th>
<th>T. vivax Accession</th>
<th>L. mexicana Accession</th>
<th>L. braziliensis Accession</th>
<th>Bodo saltans Accession</th>
</tr>
</thead>
<tbody>
<tr>
<td>KKT1</td>
<td>Tb927.10.6330</td>
<td>TcCLB.507641.190</td>
<td>TcIL3000.10.5400</td>
<td>TvY486_1006290a</td>
<td>LmxM.36.1900</td>
<td>LbrM.35.2090</td>
<td>BS71780</td>
</tr>
<tr>
<td>KKT2</td>
<td>Tb927.11.10520</td>
<td>TcCLB.510285.70</td>
<td>TcIL3000.11.11110</td>
<td>TvY486_1111400</td>
<td>LmxM.36.5350</td>
<td>LbrM.35.5600</td>
<td>BS50690</td>
</tr>
<tr>
<td>KKT3</td>
<td>Tb927.9.10920</td>
<td>TcCLB.508461.230</td>
<td>TcIL3000.9.4440</td>
<td>TvY486_0904950a</td>
<td>LmxM.34.4050</td>
<td>LbrM.34.4040</td>
<td>BS05255</td>
</tr>
<tr>
<td>KKT4</td>
<td>Tb927.8.3680</td>
<td>TcCLB.511575.70</td>
<td>TcIL3000.0.43630</td>
<td>TvY486_0803080a</td>
<td>LmxM.10.0300</td>
<td>LbrM.10.0320</td>
<td>BS42920</td>
</tr>
<tr>
<td>KKT5</td>
<td>Tb927.7.4850</td>
<td></td>
<td></td>
<td>TyY486_0704920</td>
<td>LmxM.06.0200</td>
<td>LbrM.06.0180</td>
<td>BS78350</td>
</tr>
<tr>
<td>KKT6</td>
<td>Tb927.6.1210</td>
<td>TcCLB.507603.70</td>
<td>TcIL3000.0.27570</td>
<td>TyY486_0600720</td>
<td>LmxM.12.0080</td>
<td>LbrM.12.0110</td>
<td>BS31505</td>
</tr>
<tr>
<td>KKT7</td>
<td>Tb927.11.1030</td>
<td>TcCLB.506925.490</td>
<td>TcIL3000.11.950</td>
<td>TyY486_1100920</td>
<td>LmxM.27.0430</td>
<td>LbrM.27.0520</td>
<td>BS46505</td>
</tr>
<tr>
<td>KKT8</td>
<td>Tb927.4.5110</td>
<td>TcCLB.510593.40</td>
<td>TcIL3000.8.7400</td>
<td>TyY486_0806790</td>
<td>LmxM.30.2750</td>
<td>LbrM.31.3100</td>
<td>BS23830</td>
</tr>
<tr>
<td>KKT9</td>
<td>Tb927.8.1150</td>
<td>TcCLB.506401.160</td>
<td>TcIL3000.0.31110</td>
<td>TyY486_080590a</td>
<td>LmxM.02.0610</td>
<td>LbrM.02.0590</td>
<td>BS91660</td>
</tr>
<tr>
<td>KKT10</td>
<td>Tb927.11.12410</td>
<td>TcCLB.511127.320</td>
<td>TcIL3000.0.20360</td>
<td>TyY486_1113350</td>
<td>LmxM.09.0400</td>
<td>LbrM.09.0410</td>
<td>BS75410</td>
</tr>
<tr>
<td>KKT11</td>
<td>Tb927.2.7110</td>
<td>TcCLB.506883.30</td>
<td>TcIL3000.7.2460</td>
<td>TyY486_0701840</td>
<td>LmxM.22.0120</td>
<td>LbrM.22.0120</td>
<td>BS26590</td>
</tr>
<tr>
<td>KKT12</td>
<td>Tb927.8.1680</td>
<td>TcCLB.505071.50</td>
<td>TcIL3000.8.1520</td>
<td>TyY486_0801090</td>
<td>LmxM.24.1400</td>
<td>LbrM.24.1480</td>
<td>BS37635</td>
</tr>
<tr>
<td>KKT13</td>
<td>Tb927.7.4860</td>
<td>TcCLB.507817.30</td>
<td>TcIL3000.7.4030</td>
<td>TyY486_0704930</td>
<td>LmxM.06.0210</td>
<td>LbrM.06.0190</td>
<td>BS78345</td>
</tr>
<tr>
<td>KKT14</td>
<td>Tb927.10.7240</td>
<td>TcCLB.509537.40</td>
<td>TcIL3000.10.6210</td>
<td>TyY486_1007090a</td>
<td>LmxM.36.2800</td>
<td>LbrM.35.3020</td>
<td>BS67465</td>
</tr>
<tr>
<td>KKT15</td>
<td>Tb927.6.3760</td>
<td>TcCLB.507029.50</td>
<td>TcIL3000.10.6210</td>
<td>TyY486_0603240</td>
<td>LmxM.29.2520</td>
<td>LbrM.30.2470</td>
<td>BS89620</td>
</tr>
<tr>
<td>KKT16</td>
<td>Tb927.11.1000</td>
<td>TcCLB.506925.460</td>
<td>TcIL3000.11.910</td>
<td>TyY486_1100870</td>
<td>LmxM.27.0400</td>
<td>LbrM.27.0490</td>
<td>BS79390</td>
</tr>
<tr>
<td>KKT17</td>
<td>Tb927.3.2330</td>
<td>TcCLB.508479.240</td>
<td>TcIL3000.3.1390</td>
<td>TyY486_0301690</td>
<td>LmxM.25.2220</td>
<td>LbrM.25.1800</td>
<td>BS26765</td>
</tr>
<tr>
<td>KKT18</td>
<td>Tb927.9.3800</td>
<td>TcCLB.511577.160</td>
<td>TcIL3000.9.1240</td>
<td>TyY486_0901270</td>
<td>LmxM.01.0350</td>
<td>LbrM.01.0400</td>
<td>BS26765</td>
</tr>
<tr>
<td>KKT19</td>
<td>Tb927.11.12420</td>
<td>TcCLB.511127.330</td>
<td>TcIL3000.11.13030</td>
<td>LmxM.09.0410</td>
<td>LbrM.09.0420</td>
<td></td>
<td>BS75425</td>
</tr>
</tbody>
</table>

aGene fragment.

KKT Proteins Are Highly Conserved among Kinetoplastid Species Table 1.
to the AT-rich repetitive arrays (Figures 2A, S2A, and S2B), suggesting that kinetochores are assembled onto repetitive sequences, as in humans (Hayden et al., 2013). In contrast, chromosome 3 had a strong enrichment adjacent to the repetitive sequences (Figures 2A and S2B). It is interesting that neither the repetitive arrays nor the enriched regions are AT rich for this chromosome.

We observed some enrichment on the minichromosome 177 bp repeats for the two KKT proteins, as well as H3v (Figure S2C). The core of minichromosomes consists of the
177 bp repeats in a palindromic manner with an inversion point in the middle (Wickstead et al., 2004). Although the nature of highly homogenous sequences does not allow us to determine where along the minichromosomes KKT proteins are enriched, this result implies either that kinetochores are also assembled onto minichromosomes or that minichromosomes are in close proximity to the kinetochores assembled on megabase chromosomes, possibly hijacking them to facilitate the segregation of minichromosomes.

**KKT Proteins Are Essential for Chromosome Segregation**

We next examined the biological importance of KKT proteins by inducible RNAi-mediated knockdown. We focused on KKT2, KKT7, KKT9, KKT11, KKT12, and KKT10/KKT19 for which a reasonable level of depletion of protein was achieved at 48 hr postinduction (Figure 3A). Upon induction of RNAi, growth retardation was observed in each case, albeit at a varying degree (Figure 3B). As expected, we detected abnormal DNA content induced cells (0% in control, n = 40 each) (Figure 3G). Anaphase cells had missegregation in KKT10/KKT19 RNAi-induced cells (93% of anaphase cells had abnormal signals in RNAi-induced cells, compared to 10% in control cells (n = 30 each) (Figure 3H). These results reveal that KKT proteins are essential for the segregation of both megabase chromosomes and minichromosomes.

**Predictions of Function**

Having established that KKT proteins are kinetochore proteins that play crucial roles in chromosome segregation, we next aimed to gain insights into the potential functions of individual KKT proteins from their localization patterns and bioinformatic analyses. Studies of conventional kinetochore proteins have established that their functions are often manifested in localization patterns. For example, in humans, the CENP-A protein that directly binds DNA is constitutively localized at centromeres, whereas the microtubule-binding Ndc80 subcomplex localizes at kinetochores from the onset of mitosis until the end of anaphase (Cheeseman and Desai, 2008). We observed the following patterns for the KKT proteins (Figures 4 and S3): constitutive (therefore potential DNA-binding candidates): KKT2, KKT3, and KKT4; S phase specific: KKT13; detectable from S phase until the end of anaphase (structural role and/or microtubule-binding candidates): KKT1, KKT5, KKT6, KKT7, KKT16, KKT17, and KKT18 (it is noteworthy that KKT16, KKT17, and KKT18 additionally have diffuse nuclear signals in G1); from G2/M until the end of anaphase (microtubule-binding candidates): KKT14 and KKT15; from S phase until the anaphase onset (regulator of kinetochore function candidates): KKT8, KKT9, KKT10, KKT11, KKT12, and KKT19. We speculate that those proteins that localize until the end of anaphase likely constitute the core kinetochore (KKT1, 2, 3, 4, 5, 6, 7, 14, 15, 16, 17, and 18).

Studies from other eukaryotes also established that kinetochores often consist of stable subcomplexes that form functional units and have similar localization patterns (Cheeseman and Desai, 2008). From our affinity purification/MS results, we deduced the following subcomplexes for the T. brucei kinetochore (Table S3): KKT14-KKT15 subcomplex, KKT16-KKT17-KKT18 subcomplex, and KKT6-KKT7-KKT8-KKT9-KKT10-KKT11-KKT12-KKT19 subcomplex. The fact that these proteins have largely similar localization patterns supports the assignment of these putative subcomplexes (Figure 4).

Sequence analysis of the KKT proteins revealed that the following domains and motifs are conserved among kinetoplastids (Figure 5A): the BRCA1 C terminus (BRCT) domain (KKT4), the forkhead-associated (FHA) domain (KKT13), the WD40-like domain (KKT15), protein kinase domains (KKT2, KKT3, KKT10, and KKT19), cysteine-rich domains (KKT2, KKT3), and a putative PP1-binding motif (KKT7). The BRCT and FHA domains typically function as phospho-Ser/Thr and phospho-Thr binding domains, respectively, and are found in many DNA damage response proteins (Reinhardt and Yaffe, 2013). It is noteworthy that BRCT and FHA domains are not found in any known conventional kinetochore proteins. The WD40 domain is one of the most abundant domains in eukaryotic genomes and is found in proteins involved in a large variety of cellular processes (Reinhardt and Yaffe, 2013).

KKT10 and KKT19 exhibit a high degree of similarity at the protein level, as do KKT2 and KKT3 and KKT17 and KKT18, suggesting that these pairs likely arose from gene duplication events. Although KKT10 and KKT19 have previously been classified as members of the CLK/Lammer subfamily in the CMGC family (Parsons et al., 2005), there are significant differences between KKT10/KKT19 and the human or Arabidopsis CLK/Lamm mer kinases (Figure S4), implying that KKT10/KKT19 may have adapted to carry out kinetochore functions in kinetoplastids. Interestingly these kinases (named as TbCLK1 and TbCLK2 in that study) were recently identified as targets of a fungal natural product Hypothemycin (Nishino et al., 2013), demonstrating the potential of trypanosome kinetochore kinases as drug targets.

KKT2 and KKT3 possess residues characteristic of active eukaryotic protein kinases but do not have a clear affiliation to any known group or family (Parsons et al., 2005), suggesting that these proteins are likely to be kinetoplastid specific. Interestingly, these proteins also possess cysteine-rich domains (Figure 5A), in which classic zinc-finger motifs can be recognized (Figure 5B). Furthermore, several DNA-binding motifs (SPKK [Suzuki, 1989] and AT-hook [Aravind and Landsman, 1998]) are found in KKT2 and KKT3 in some kinetoplastids, although not strictly conserved across kinetoplastids (Figure 5C).
Together with their constitutive localization pattern (Figure 4), we speculate that KKT2 and KKT3 are loaded onto centromeric DNA via the cysteine-rich domains, SPKK and/or AT-hook, to modify other proteins via the unique kinase domains, contributing to the establishment of kinetochores.

Taken together, our bioinformatic analysis failed to find any evidence that KKT proteins are similar to conventional kinetochore proteins at the primary sequence level. Although this by itself does not mean that kinetoplastid kinetochores are completely different, it is striking that all of the features we detect imply...
difference, not similarity. The simplest interpretation is that kinetoplastids contain unconventional kinetochores composed of distinct kinetochore proteins.

**DISCUSSION**

Accurate transmission of genetic material in eukaryotes depends on the attachment of dynamic spindle microtubules to chromosomes via the macromolecular kinetochore complexes. Available evidence suggests that spindle microtubules composed of $\alpha/\beta$ tubulins are ubiquitously used in all eukaryotes studied thus far (Wickstead and Gull, 2011; Drechsler and McAinsh, 2012). In contrast, it was previously not clear whether all eukaryotes utilize similar kinetochore proteins because none of the conventional kinetochore components were identifiable in any kinetoplastid genome. Our identification of 19 kinetochore proteins in *T. brucei* has revealed that kinetoplastid kinetochores are composed of proteins that are distinct from conventional kinetochore proteins in other eukaryotes. This new group of proteins may therefore constitute an attractive drug target for kinetoplastid diseases, such as sleeping sickness and nagana caused by *T. brucei*, Chagas disease caused by *T. cruzi*, and leishmaniasis caused by *Leishmania* species (Stuart et al., 2008). Further studies will be required to explore the unconventional kinetochores as a means to combat these diseases.

The goal of the eukaryotic kinetochore is to mediate the interaction between DNA and microtubules. Therefore, understanding how KKT proteins achieve these tasks in kinetoplastids will contribute to a better understanding of how conventional kinetochores function. For example, we still do not know why CENP-A is so widely used in eukaryotes, despite the fact that CENP-A is not strictly essential for building functional kinetochores (Hori et al., 2013). The current prevailing idea is that CENP-A forms a centromere-specific chromatin environment that somehow acts as an epigenetic marker for kinetochore assembly (Nechemia-Arbely et al., 2012; Müller and Almouzni,
Because kinetoplastids do not possess CENP-A by nature, it is not clear whether their centromere identity is epigenetically defined by a distinct mechanism or how the kinetochore assembly site is determined. By understanding the CENP-A-independent kinetochores in kinetoplastids, we may obtain insights into the specialty of CENP-A.

There is an intimate relationship between repetitive sequences, the endogenous RNAi machinery, and the kinetochore assembly in many species (Buscaino et al., 2010). The endogenous RNAi system is important for faithful chromosome segregation in T. brucei (Durand-Dubief et al., 2007), and noncoding RNAs from some, but not all, centromeric repeats have been detected (Tschudi et al., 2012). It will be important to determine whether (and how) the RNAi system contributes to the deposition of kinetochore proteins at centromeres in T. brucei. In contrast, T. cruzi does not possess an endogenous RNAi system, and its centromeres are devoid of repetitive sequences (Obado et al., 2005). It will also be interesting to reveal how T. cruzi determines the kinetochore assembly sites in the absence of an RNAi system.

T. brucei does not appear to possess a functional spindle checkpoint system that monitors the kinetochore-microtubule attachment and regulates the activation of the anaphase-promoting complex (Ploubidou et al., 1999; Akiyoshi and Gull, 2013). However, we found that KKT4 copurifies with several APC/C subunits (Table S2), raising a possibility that this kinetochore protein may directly communicate with the APC/C. It is interesting that KKT4 signal is not mitosis specific but is found throughout the cell cycle (and thus may locate close to DNA rather than microtubules). Gaining insights into the functions of KKT4 may lead to a better understanding of the APC/C regulatory mechanism, as well as the nature of signals transmitted from kinetochores to regulate the APC/C.

Determining the position of the root of the eukaryotic tree of life remains an unresolved problem (Embley and Martin, 2006; Walker et al., 2011). Among several competing hypotheses (e.g., Stechmann and Cavalier-Smith, 2002; Rogozin et al., 2009; Katz et al., 2012), it has been proposed, based on unique mitochondrial cytochromes c/c1 and the absence of a recognizable biogenesis apparatus for these proteins (Allen et al., 2008), that Euglenozoa (a phylum that includes kinetoplastids) may represent extremely early or the earliest-branching eukaryotes (Cavalier-Smith, 2010). Therefore, it is possible that kinetoplastids evolved the KKT-based kinetochore system early in the eukaryotic history, whereas other eukaryotes evolved a system utilizing conventional kinetochore proteins. A corollary is that this controversy hypothesis that roots the base of the eukaryotic tree between Euglenozoa (or deep within the Euglenozoa tree) and all other eukaryotes now receives support from two very distinct properties (mitochondrial cytochromes and kinetochores), as well as many others (Cavalier-Smith, 2010, 2013). More work on kinetoplastids and other Euglenozoa species is very much warranted to further test the validity of this hypothesis. However, even if this rooting is correct, we would not be able to tell what kind of kinetochores the last eukaryotic common ancestor (LECA) possessed. It might be that the LECA possessed a conventional kinetochore system, which was later replaced by the KKT-based system in kinetoplastids (Figure S5A). Or it might be that the LECA utilized the KKT system, but only kinetoplastids retained it, whereas other eukaryotes lost it and developed a conventional kinetochore system (Figure S5B). Alternatively, the LECA might have possessed a hitherto-unknown type of kinetochores (Figure S5C). Whatever the evolutionary history might be, understanding the KKT-based kinetochores in kinetoplastids should lead to a better understanding of the chromosome segregation machinery in eukaryotes.

**EXPERIMENTAL PROCEDURES**

**Cells**

All cell lines used in this study were derived from T. brucei SmOxP927 procyclic form cells (TREU 927/4 expressing t7 RNA polymerase and the tetracycline repressor to allow inducible expression) (Yuon et al., 2012) and are listed in Table S4. Plasmids and primers used in this study are also listed in Table S4. Cells were grown at 28°C in SDM-79 medium supplemented with 10% (v/v) heat-inactivated fetal calf serum (Brun and Schönenberger, 1979). Cell growth was monitored using a CASY cell counter and analyzer system (Roche). For induction of RNAi, doxycycline was added to the medium to a final concentration of 1 μg/ml. Endogenous YFP and tdTomato tagging was performed using the pEnTS-Y vector (Kelly et al., 2007) and pBA148, respectively. For generation of inducible RNAi cell lines, ~500 bp fragments were amplified from genomic DNA and cloned into the p2T7-177 vector (Wickstead et al., 2003). Details on plasmid construction are described in the Extended Experimental Procedures. Plasmids linearized by Nell site were transfected to trypanosomes by electroporation into an endogenous locus (pEnTS-Y and pBA148 derivatives) or 177 bp repeats on minichromosomes (p2T7-177 derivatives). Transfected cells were selected by the addition of 25 μg/ml hygromycin (pEnTS-Y derivatives), 10 μg/ml blasticidin (pBA148 derivatives), or 5 μg/ml phleomycin (p2T7-177 derivatives).

**Fluorescence Microscopy**

For the analysis of fluorescently tagged proteins or DNA contents, cells were washed once with PBS, settled onto glass slides, and fixed with 4% paraformaldehyde in PBS for 5 min. Cells were then permeabilized with 0.1% NP-40 in PBS for 5 min and embedded in mounting media (1% 1,4-Diazabicyclo[2.2.2]octane (DBACO), 90% glycerol, and 50 mM sodium phosphate [pH 8.0] containing 100 ng/ml DAPI). FISH was carried out as described (Ersfeld and Gull, 1997) using digoxigenin-labeled probes against 177 bp repeats (Ersfeld and Gull, 1997) for minichromosomes (p2T7-177 derivatives). Transfected arrays were captured on a fluorescence microscope (Leica Microsystems) equipped with an Orca cooled charge-coupled device (CCD) camera (Hamamatsu Photonics) and processed in ImageJ (Schneider et al., 2012).

**Immunoprecipitation**

We developed an immunoprecipitation method for trypanosome kinetochore proteins based on our kinetochore purification protocol in budding yeast.
(Akiyoshi et al., 2010). Typically, 400 ml cultures of asynchronously growing cells were harvested at 1.2 $\times$ 10^7 cells/ml. Cells were pelleted by centrifugation (900 g, 10 min), washed once with PBS, and extracted in PEME (100 mM PIPES-NaOH [pH 6.9], 2 mM EGTA, 1 mM MgSO_4, and 0.1 mM EDTA) with 1% NP-40 and protease inhibitors (Leupeptin, Pepstatin, E-64, 10 $\mu$g/ml each, and 0.2 mM PMSF) and phosphatase inhibitors (1 mM sodium pyrophosphate, 2 mM Na-beta-glycerophosphate, 0.1 mM Na_3VO_4, 5 mM NaF, and 100 mM microcinyst-LR) for 5 min at room temperature, followed by centrifugation (1,800 g, 15 min). Samples were kept on ice from now on.

The pellet fractions that contain kinetochrome proteins were resuspended in modified buffer B (BH): 0.1% Triton-X-100 (25 mM HEPES [pH 8.0], 2 mM MgCl_2, 0.1 mM EDTA [pH 8.0], 0.5 mM EGTA [pH 8.0], 1% NP-40, 150 mM KCl, and 15% glycerol) containing protease inhibitors and phosphatase inhibitors. Samples were sonicated to solubilize kinetochrome proteins (12 s, 3 times with 1 min interval on ice), producing “input” extract samples. 12 $\mu$g of mouse monoclonal anti-GFP antibodies (Roche, 1181446001) that had been pre-conjugated with 60 $\mu$l slurry of Protein-G magnetic beads (Dyna, with dimethyl pimelimidate (as described in Unnikrishnan et al. [2012]) were incubated with input extracts for 3 hr with constant rotation, followed by four washes with modified BH/0.15 containing protease inhibitors, phosphatase inhibitors, and 2 mM dithiothreitol (DTT). Beads were further washed three times with pre-elution buffer (50 mM Tris-HCl [pH 8.3], 75 mM KCl, and 1 mM EGTA). Associated proteins were gently eluted from the beads by agitation in 60 $\mu$l of elution buffer (0.1% Rapigest and 50 mM Tris-HCl [pH 8.3]) for 25 min at room temperature. 10 $\mu$l of samples were analyzed by immunoblots (using monoclonal anti-GFP or anti-Ty antibodies) and Sypro-Ruby staining.

The rest of samples (50 $\mu$l) were used to identify copurifying proteins by MS (as described below). SDS-PAGE and immunoblots were performed by standard methods using the following mouse monoclonal antibodies: anti-GFP (Roche, 1181446001) or BB2 (anti-Ty) (Bastin et al., 1996) for TY-YFP-tagged KKT proteins and LC4 (anti-PFR2) (Kohl et al., 1999) for a loading control.

**Mass Spectrometry**

The samples were incubated at 100°C for 5 min. Proteins were reduced with 5 mM DTT at 37°C for 30 min and alkylated with 10 mM iodoacetamide at 37°C for 30 min. The reaction was quenched by adding 10 mM DTT at 37°C for 30 min, and 100 $\mu$l of 20 mM Tris-HCl (pH 8.3) was added. Proteins were digested overnight at 37°C with 0.4 $\mu$g of trypsin (Promega). Formic acid was added to 2%, and the samples were incubated at 37°C for 30 min to cleave the detergent Rapigest, followed by centrifugation for 10 min. The supernatant was desalted over a C18 column and analyzed by electrospray tandem mass spectrometry over a 40 min gradient using an LTQ XL-Orbitrap (Thermo Scientific) at the Central Proteomics Facility (http://www.proteomics.ox.ac.uk; Sir William Dunn School of Pathology, University of Oxford).

Data analysis was performed by using the central proteomics facilities pipeline (CPPP) (Tradgin et al., 2010). Peptides were identified by searching MS/MS spectra against the T. brucei protein database with Mascot (Matrix Science), OMSA (Geer et al., 2004), and X!Tandem (Craig and Beavis, 2004) with carboxyamidomethyl cysteine as fixed modification. Up to two missed cleavages were allowed. Oxidized-methionine and phosphorylation were searched as variable modifications. Mass tolerances for MS and MS/MS peak identifications were 20 ppm and 0.5 Da, respectively. Peptides identified with at least two peptides were considered and shown in Table S2. Raw MS data are available upon request.

**Bioinformatics**

Sources of predicted protein database used for the search of conventional kinetochrome proteins in various organisms from the six eukaryotic superfamilies (Walker et al., 2011) are listed in Table S4. Wherever possible, protein names searchable in the NCBI database are listed in Table S1. Putative CENP-A homologs were identified as reported previously (Talbert et al., 2009). Putative homologs for conventional kinetochrome proteins (CENP-C, Ndc80, NuF2, Spc24 and Spc25) were identified using HMMER (version 3.0) (Eddy, 1998; Finn et al., 2011). Pairwise sequence alignment and motif search were performed by EMBL-EBI tools (McWilliam et al., 2013) and Pfam (Punta et al., 2012), as well as manual inspection. Multiple sequence alignment was performed with MAFFT (version 7) (Katoh and Standley, 2013) and visualized with Clustalx coloring scheme in Jalview (version 2.8) (Waterhouse et al., 2009). Genome sequences for nonkinetoplastid Euglenozoa are currently not available, and we therefore do not know whether KKT-based kinetochores are conserved across Euglenozoa.

**ChIP-Seq**

ChIP was carried out essentially as described (Siegel et al., 2009) using 10 times more cells (1 $\times$ 10^7 cells for each experiment). Briefly, cells expressing either YFP-KKT2, YFP-KKT3, or YFP-H3v were fixed with 1% formaldehyde for 20 min at room temperature and sonicated to prepare input chromatin fragments, and YFP-tagged proteins were immunoprecipitated with rabbit polyclonal anti-GFP antibodies (Invitrogen, A11212) that were preconjugated with Protein-A magnetic beads (Dyna), followed by DNA purification. Single-end sequencing (49 bp sequence tag) was carried out on a HiSeq2000 Illumina platform at BGI Hong Kong. Both input DNA and ChIP DNA were sequenced in each case. Reads with adaptors, reads with unknown nucleotides larger than 5%, and reads with low quality (more than 20% of the bases’ qualities are less than 10 in a read) were removed to provide clean reads (see Table S5 for statistics). Sequence tags were mapped using Burrows-Wheeler Aligner (version 0.7.4) (Li and Durbin, 2009), allowing up to two mismatches to the T. brucei 927 genome (version 5.0; note that centromeres of chromosomes 9, 10, and 11 are not in the genome assembly in this version) supplemented with a contig tryp. X-28409.p1c (42,529 bp) that consists mostly of the 177 bp repeat sequences (and thus is very likely to be derived from a minichromosome) to examine the enrichment ratio on 177 bp sequences. When reads map to multiple locations in the reference genome (due to the presence of identical sequences); the program randomly chooses the hits. The following centromere repeat unit pairs have a similarity level higher than 96% (Obado et al., 2007), and we therefore cannot distinguish them in our analysis: between chromosome 10 and chromosomes 4/11 and between chromosome 4 and chromosome 9. Other pairs are less than 93% identical.

The SAMtools program (Li et al., 2009) was used to generate bam files, and tag counting was done by BEDTools (Quinlan and Hall, 2010) (coverageBed) using 150 bp nonoverlapping windows. The ChIPinput ratio was then calculated for each window, and the results were visualized in Excel. To reduce noise, we ignored a few windows with less than 10 reads in the input. Background levels for each protein were calculated for 550,000–750,000 of chromosome 4 and are shown in Figure S2C (KKT2: 0.59; KKT3: 0.40; H3v: 0.41).

**ACCESSION NUMBERS**

NGS sequence data are deposited in the SRA database under accession number SRP031518.

**SUPPLEMENTAL INFORMATION**

Supplemental Information includes Extended Experimental Procedures, five figures, and five tables and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2014.01.049.

**ACKNOWLEDGMENTS**

We thank Kim Nasmyth, Stuart Ferguson, Sue Biggins, and Hannah Long for comments on the manuscript. We also thank Vladimir Varga for discussion; Jonathan Higgins for advice on kinases; Takashi Sutani, Nicolai Siegel, and Mike Fiebig for advice on ChIP-seq; and the Central Proteomics Facility, Benjamin Thomas, and Gabriela Ridlova for mass spectrometry analysis.

B.A. was supported by postdoctoral fellowships from the EMBO and Human Frontier Science Program and is currently supported by a Sir Henry Dale Fellowship jointly funded by the Wellcome Trust and the Royal Society (grant 098403/Z/12/Z), as well as a Wellcome-Beit Prize Fellowship (grant 098403/Z/12/A). Research in the Gull laboratory is supported by the Wellcome Trust. The authors declare no competing financial interests.
REFERENCES


