Leading Edge
Minireview
Cell

Centromere Assembly and Propagation

Corey A. Morris[1,*] and Danesh Moazed[1,*]
[1]Department of Cell Biology, Harvard Medical School, Boston, MA 02115, USA
*Correspondence: cmorris@hms.harvard.edu (C.A.M.), danesh@hms.harvard.edu (D.M.)
DOI 10.1016/j.cell.2007.02.002

Centromere assembly provides a unique example of how elaborate protein structures can be assembled onto DNA, independent of sequence, and then stably propagated through numerous cell divisions. Here, we review the possible epigenetic strategies that organisms ranging from yeast to human use to assemble and propagate active centromeres.

The eukaryotic centromere is the specialized chromosomal region upon which kinetochores—the structures that link centromeres to spindle microtubules—assemble and direct the equal segregation of chromosomes during mitosis and meiosis. The discovery of sequence-dependent "point" centromeres in the budding yeast *Saccharomyces cerevisiae* suggested an attractive model in which analogous sequence elements in metazoans would also define the assembly of kinetochores at distinct chromosomal loci. However, studies over the past 15 years have failed to identify sequence-specific DNA elements that are sufficient to define the centromeres in other eukaryotes. Instead, what has emerged is strong evidence that chromatin-based epigenetic mechanisms establish and propagate active centromeres in most, if not all, eukaryotes.

Well-characterized centromeres range in size from the 125 bp centromere of budding yeast to the several megabase centromeres of human chromosomes. The budding yeast centromere is divided into three distinct centromere DNA elements, two of which act as sequence-specific binding sites (CDEs; Figure 1). However, directly analogous elements are absent in the fission yeast *Schizosaccharomyces pombe* and metazoans. Rather, these organisms have repetitive centromeric DNA that directs heterochromatin assembly, which appears to substitute for the role of sequence-specific binding sites.

Despite significant differences in size and sequence composition, general features of centromeres from various organisms appear to be highly conserved (Figure 1). Fission yeast centromeres are reminiscent of their human counterparts in several respects, including the fact that they contain highly repetitive (A+T)-rich sequence, are flanked by domains of heterochromatin, and are associated with similar proteins. These conserved features, combined with genetic and biochemical tractability, have made fission yeast an excellent system in which to study centromere biology. Here, we discuss results from work on fission yeast centromeres in the context of those from several other organisms.

The Epigenetic Nature of Centromeres

Epigenetic inheritance of centromeres has been noted in several species. Early evidence came from attempts to identify a minimal centromere in fission yeast. Here, it was found that particular plasmids could adopt either "stable" or "unstable" segregation states when transformed into cells (Steiner and Clarke, 1994). The sequences that conferred this epigenetic centromere function are the same sequences that are able to initiate heterochromatin formation. In addition, marker genes placed at the central core of fission yeast centromeres exhibit meta-stable inheritance of either a silenced state, expressed state, or intermediate state, indicating that epigenetic modes of inheritance are at work (reviewed in Pidoux and Allshire, 2005).

Examples of centromere epigenetics have also been found in metazoans. In *Drosophila*, centromeric activity can be imparted to noncentromeric chromatin and propagated even when dissociated from that centromere. In human dicentric chromosomes (containing two centromeres) only one satellite array is active and recruits kinetochore

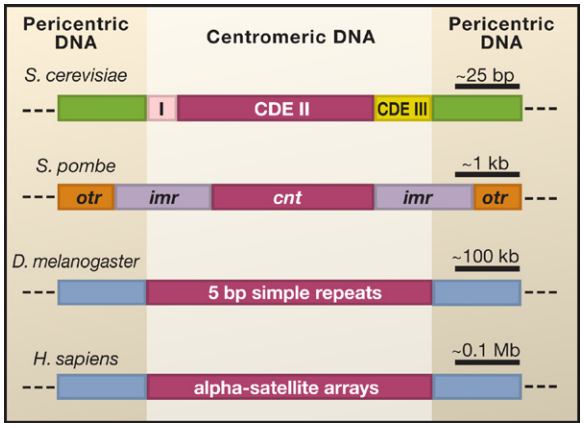

Figure 1. Organization of the Eukaryotic Centromere
The basic structure of the centromere from *S. cerevisiae*, *S. pombe*, *D. melanogaster*, and *H. sapiens* is depicted. The *S. cerevisiae* centromere is comprised of three distinct DNA elements: CDEI, CDEII, and CDEIII. *S. pombe* centromeres consist of a nonrepetitive central core region (*cnt*), which is symmetrically flanked by inverted repeat regions—the innermost repeats (*imr*) and the outer repeats (*otr*). Five base pair repeats and alpha satellites are shown for *D. melanogaster* and *H. sapiens*, respectively; however, other sequence elements also help to define these regions. A conserved feature in all of these organisms is the incorporation of the histone H3 variant CenH3 at centromeric DNA to create unique chromatin upon which the kinetochore assembles. Centromeric DNA is flanked by pericentric DNA, which consists of phased nucleosomes in budding yeast and heterochromatin in organisms from fission yeast to human.

proteins, whereas the other is inactivated. Several studies have shown that centromeres can form and propagate at chromosomal loci bearing no alpha satellite DNA, emphasizing the sequence-independent and epigenetic nature of human centromere formation (reviewed in Sullivan, 2001).

Even budding yeast's point centromere behaves distinctly depending on whether it is recently activated or has been active for several cell cycles. Mutations in the core CDE element reduce the association of cohesin with naive centromeres; however, such mutations have little effect on established centromeres (Tanaka et al., 1999), indicating that although the mutations reduce the probability of forming the chromatin structure necessary for activity de novo, they have only a modest effect once these structures are formed. Furthermore, in budding yeast, naked centromere DNA introduced into the cell is unable to assemble kinetochores in the absence of certain kinetochore proteins; however, established centromeres retain their ability to segregate in the absence of these proteins (Mythreye and Bloom, 2003). Thus, once assembled, even budding yeast centromeres appear to be inherited epigenetically.

Perhaps the most striking example of epigenetic propagation of a centromere comes from a recent study in the pathogenic yeast *Candida albicans* (Baum et al., 2006). Naked centromeric DNA that is sufficient to confer centromere activity in vivo was shown to be unable to assemble functional centromeric chromatin and kinetochores de novo when reintroduced into cells. These observations indicate that *C. albicans* centromeres are entirely dependent on their preexisting chromatin state for the propagation of functional centromeric chromatin.

CenH3, a Centromere-Specific Histone H3 Variant

The basic unit of chromatin is the nucleosome, in which ~147 bp of DNA is wrapped around an octamer of four histones, H2A, H2B, H3, and H4. Distinct chromatin domains are defined by specific posttranslational modifications of each histone as well as by packaging of DNA using histone variants. The most prominent example of the use of a histone variant to define a specialized chromatin domain is found at centromeres, which are universally defined by the presence of a conserved histone H3 variant in centromeric nucleosomes (referred to hereafter as CenH3; known as *S. cerevisiae* Cse4p, *S. pombe* Cnp1, *D. melanogaster* Cid, and *H. sapiens* CENP-A). The fact that CenH3 is incorporated at the level of the nucleosome combined with its conserved and essential presence at centromeres throughout the cell cycle make it a strong candidate for specifying centromeres as the site of kinetochore assembly.

If the presence of CenH3 is the epigenetic mark that confers centromeric identity, how, then, is CenH3 deposited at the right location and epigenetically inherited? Despite the essential function of CenH3, the mechanisms that specifically deposit and maintain its centromeric localization remain poorly understood. Several proteins affect CenH3 localization, but none appear to be specific for CenH3 assembly. Identified proteins include RbAp48, a histone H3/H4 chaperone that is found in known chromatin assembly complexes (Hayashi et al., 2004) and also interacts with the *Drosophila* CenH3^{Cid} (Furuyama et al., 2006). Although RbAp48 was able to passively facilitate CenH3 loading in vitro, it has yet to be determined how this chaperone—which also facilitates H3 loading in vitro—contributes specifically to CenH3 localization in vivo.

Domains of Centromeric Chromatin

In fission yeast and metazoans, chromatin structure of centromeres is comprised of a central domain of unique chromatin flanked by heterochromatin (reviewed in Pidoux and Allshire, 2005). Central core chromatin contains CenH3, which is thought to be interspersed with canonical H3-containing nucleosomes (reviewed in Carroll and Straight, 2006). CenH3 occupies a region ranging from several kb in fission yeast to several thousand kb in metazoans; in contrast, the budding yeast CenH3^{Cse4} is thought to be present in only a single nucleosome at the CEN element. However, a recent study showed that CenH3^{Cse4} is present in an ~20 kb domain on either side of the CEN sequence in cells arrested at the G2/M transition, a stage of the cell cycle during which microtubules attach to the chromosomes (Riedel et al., 2006). Whether this represents an intermediate step in CenH3^{Cse4} loading previously undetected or is instead related to cell-cycle arrest has yet to be determined. Interestingly, it has been observed that mislocalized CenH3^{Cse4} is restricted to the centromere via proteolysis in budding yeast (Collins et al., 2004), a phenomenon recently also observed for *Drosophila* CenH3^{Cid} (Moreno-Moreno et al., 2006); thus, it is possible that this apparently conserved mechanism may be inhibited in these checkpoint-arrested cells. If a domain of CenH3^{Cse4} proves to be the physiological scenario, it may suggest that budding yeast chromosomes contain a region of centromeric chromatin more similar to other organisms than previously thought.

It has become increasingly clear that in addition to properly loading CenH3 at the central core, the chromatin context of the centromere is important for proper chromosome segregation. The flanking sequences are assembled into heterochromatin: a form of condensed and repressed chromatin that plays a crucial role in chromosome stability and the regulation of gene expression. In fission yeast, the centromere is flanked by heterochromatin that is established and maintained in an RNAi-dependent manner (Verdel and Moazed, 2005). As in metazoans, fission yeast pericentric heterchromatin is marked by the presence of methylated histone H3K9 and several chromodomain proteins that bind to the methylated H3 tail. Here, the siRNA-containing RNA-induced transcriptional silencing (RITS) complex associates with nascent transcripts and methylated H3K9, leading to further H3K9 methylation and the recruitment of Swi6—a chromodomain protein that binds to the methylated H3K9 and is the fission yeast homolog of the *Drosophila* and human heterochromatin protein 1 (HP1) (Figure 2A). RNAi has also been shown to be important for pericentric heterochromatin formation and centromere function in vertebrate cells (Fukagawa et al., 2004; Kanellopoulou et al.,

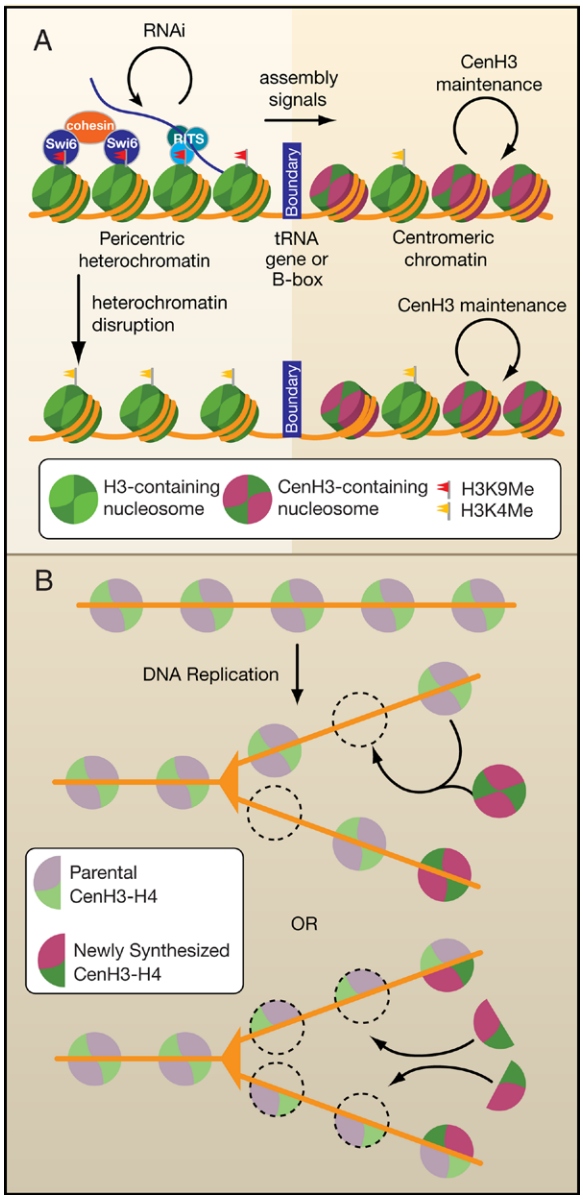

Figure 2. Centromere Establishment and Propagation

(A) Fission yeast have two self-propagating chromatin domains at centromeres: pericentric heterochromatin (with H3K9-methylated histones; H3K9Me) and CenH3-containing centromeric chromatin (thought to be interspersed with H3K4-methylated histones; H3K4Me). The flanking pericentric heterochromatin domain may provide assembly signals to a newly forming centromere. These could include RNA-induced transcriptional silencing (RITS) machinery; pericentric histone modifications such as H3K9Me; pericentric heterochromatin-associated proteins such as Swi6 or cohesin; and/or the transcription of tRNA genes. Mutations in proteins of either domain disrupt proper chromosome segregation; however, loss of heterochromatin does not affect CenH3 incorporation in already functional centromeres.

(B) Two models of self-propagating CenH3 chromatin assembly at centromeres are considered: (1) random distribution of CenH3-containing nucleosomes upon DNA replication, followed by CenH3-directed incorporation of newly synthesized CenH3-containing nucleosomes; or (2) random distribution of CenH3/H4 dimers followed by CenH3-directed incorporation of newly synthesized CenH3/H4 dimers (for simplicity, only tetramers of CenH3/H4 are shown in nucleosomes).

2005). A known function of pericentric heterochromatin is the recruitment of the cohesin complex, which is required for proper chromosome segregation (see Pidoux and Allshire, 2005), and mutations that disrupt heterochromatin assembly, such as those in the RNAi pathway, increase the rate of chromosome loss.

Beyond cohesin recruitment, one role of pericentric heterochromatin may be to exclude CenH3 incorporation. Overexpressed *Drosophila* CenH3^{Cid} has been shown to mislocalize preferentially to euchromatin while being notably absent from pericentric heterochromatin (Heun et al., 2006). Another intriguing possibility is that this RNAi-dependent heterochromatin provides a favorable environment, or even specific signals, for the establishment of a functional centromere.

In fission yeast, a portion of the outer repeat has been shown to significantly enhance the formation of a stable centromere (Steiner and Clarke, 1994). This outer repeat contains a transcriptional unit that is sufficient to initiate RNAi-dependent heterochromatin formation, indicating that the requirements for heterochromatin assembly and proper chromosome segregation reside within the same DNA region. These and other studies have shown that although the inclusion of inverted repeat sequences improved function in some constructs, it was not necessary for centromere assembly. Thus, it is unlikely that the higher-order pairing of symmetrical sequences flanking the central core is the essential function of these required pericentric domains, as has been proposed.

It is clear that specific domains of chromatin/heterochromatin are established to create a fully functional centromere. What is less clear is how these domains are delineated. Recent work has demonstrated that transfer (t)RNA genes and B box motifs function as boundary elements in fission yeast (Noma et al., 2006; Scott et al., 2006). Indeed, tRNA genes and B box-containing elements are found at the boundaries between centromeric chromatin and pericentric heterochromatin in fission yeast. Disruption of the tRNA genes allows invasion of pericentric heterochromatin to the central core and disrupts chromosome segregation, suggesting that the delineation of the centromeric chromatin and flanking heterochromatin domains is essential for proper centromere function.

Conclusions and Perspective

As discussed above, binary centromeric states–displaying either mitotically stable or unstable segregation–have been reported in a number of organisms. In addition to fission yeast and animal systems discussed here, plants appear to exhibit many of the same sequence-independent centromeric features. Because plants diverged from fungi and animals well before fungi and animals diverged from one another, epigenetic centromeres are likely to have arisen early in eukaryotic evolution. In species lacking a sequence-specific centromere, formation of active centromeres in the absence of epigenetic signals appears to be a stochastic event, occurring only rarely. However, once established, active centromeres are propagated epigenetically.

Although clues have begun to emerge, the questions still remain, what are the signals that specify the location of centromere assembly and how is this state epigenetically inherited? In general, there appear to be two independent epigenetic mechanisms that may be important for centromere function: (1) RNAi-dependent heterochromatin formation at pericentric domains and (2) incorporation of CenH3 at central core chromatin (Figure 2A). Although the former mechanism is absent in budding yeasts, the latter appears to be universal. The precise signal in pericentric heterochromatin that mediates CenH3 loading remains to be defined. However, once an active centromere is established, heterochromatin is clearly dispensable for CenH3 maintenance as mutations that abrogate heterochromatin are viable. Studies exploring de novo centromere formation and propagation have not been explored in heterochromatin-defective cells. Such studies are required to determine whether heterochromatin, with or without the participation of RNAi, can indeed provide the necessary assembly signals for CenH3 deposition to the central core.

Clues to how CenH3 can be epigenetically inherited come from studies of chromatin assembly (reviewed in Annunziato, 2005). For canonical nucleosomes, two possible inheritance mechanisms have been considered. One proposes that histone octomers, or more likely H3-H4 tetramers, are distributed randomly between daughter DNA strands during DNA replication. In an alternative model, the H3-H4 tetramer splits during DNA replication and segregates to each of the daughter DNA strands. Such half nucleosomes are then converted into full nucleosomes by the deposition of newly synthesized H3-H4. Either mechanism could apply in the case of CenH3 during replication of the centromere, establishing a cyclical, chromatin-directed epigenetic propagation of CenH3 (Figure 2B). Indeed, CenH3 has been shown to be distributed between the two daughter strands during DNA replication (reviewed in Sullivan et al., 2001). Furthermore, because CenH3 nucleosomes are thought to be homotypic (containing two copies of CenH3, not one copy of CenH3 and one of H3), analogous distributive mechanisms could function to propagate CenH3 deposition at the centromere.

Several other models for the propagation of centromere identity have been proposed, from the successful engagement of microtubules (see Pidoux and Allshire, 2005) to those based on timing of replication and three-dimensional nuclear organization (reviewed in Sullivan et al., 2001). However, in many species, CenH3 can be properly incorporated independent of mitosis or DNA synthesis, or even when proteins that disrupt kinetochore function are mutated. Furthermore, studies of interphase human and *Drosophila* cells have shown that centromeres do not occupy predictable nuclear space. Thus, although none of these models are mutually exclusive, they are unlikely to constitute a sole conserved mechanism for CenH3 localization. A chromatin-based assembly model whereby CenH3 directs its own loading remains attractive as it explains the epigenetic inheritance of centromeric chromatin while permitting the postreplication replenishment of CenH3 independent of both cell-cycle and DNA sequence (Figure 2B).

With the possible exception of budding yeast, the epigenetic nature of centromeres has made it challenging to identify the molecules that establish and maintain active centromeres, but the tools to approach this central problem are now available in other model systems. Although exact molecular mechanisms remain unclear, two distinct epigenetic events act on centromeres: one involving a mechanism that can specify where the centromere is assembled and a second that involves the epigenetic propagation of active centromeres once they are assembled. Initiation may be dependent on structural components of pericentric heterochromatin, including noncoding centromeric RNAs and the RNAi machinery, whereas propagation may require CenH3 chromatin "replication." This is clearly an area where many exciting discoveries are forthcoming.

REFERENCES

Annunziato, A.T. (2005). J. Biol. Chem. *280*, 12065–12068.

Baum, M., Sanyal, K., Mishra, P.K., Thaler, N., and Carbon, J. (2006). Proc. Natl. Acad. Sci. USA *103*, 14877–14882.

Carroll, C.W., and Straight, A.F. (2006). Trends Cell Biol. *16*, 70–78.

Collins, K.A., Furuyama, S., and Biggins, S. (2004). Curr. Biol. *14*, 1968–1972.

Fukagawa, T., Nogami, M., Yoshikawa, M., Ikeno, M., Okazaki, T., Takami, Y., Nakayama, T., and Oshimura, M. (2004). Nat. Cell Biol. *6*, 784–791.

Furuyama, T., Dalal, Y., and Henikoff, S. (2006). Proc. Natl. Acad. Sci. USA *103*, 6172–6177.

Hayashi, T., Fujita, Y., Iwasaki, O., Adachi, Y., Takahashi, K., and Yanagida, M. (2004). Cell *118*, 715–729.

Heun, P., Erhardt, S., Blower, M.D., Weiss, S., Skora, A.D., and Karpen, G.H. (2006). Dev. Cell *10*, 303–315.

Kanellopoulou, C., Muljo, S.A., Kung, A.L., Ganesan, S., Drapkin, R., Jenuwein, T., Livingston, D.M., and Rajewsky, K. (2005). Genes Dev. *19*, 489–501.

Moreno-Moreno, O., Torras-Llort, M., and Azorin, F. (2006). Nucleic Acids Res. *34*, 6247–6255.

Mythreye, K., and Bloom, K.S. (2003). J. Cell Biol. *160*, 833–843.

Noma, K., Cam, H.P., Maraia, R.J., and Grewal, S.I. (2006). Cell *125*, 859–872.

Pidoux, A.L., and Allshire, R.C. (2005). Philos. Trans. R. Soc. Lond. B Biol. Sci. *360*, 569–579.

Riedel, C.G., Katis, V.L., Katou, Y., Mori, S., Itoh, T., Helmhart, W., Galova, M., Petronczki, M., Gregan, J., Cetin, B., et al. (2006). Nature *441*, 53–61.

Scott, K.C., Merrett, S.L., and Willard, H.F. (2006). Curr. Biol. *16*, 119–129.

Steiner, N.C., and Clarke, L. (1994). Cell *79*, 865–874.

Sullivan, B.A., Blower, M.D., and Karpen, G.H. (2001). Nat. Rev. Genet. *2*, 584–596.

Sullivan, K.F. (2001). Curr. Opin. Genet. Dev. *11*, 182–188.

Tanaka, T., Cosma, M.P., Wirth, K., and Nasmyth, K. (1999). Cell *98*, 847–858.

Verdel, A., and Moazed, D. (2005). FEBS Lett. *579*, 5872–5878.