

# Centromeres Take Flight: Alpha Satellite and the Quest for the Human Centromere

## Minireview

Terence D. Murphy and Gary H. Karpen  
Molecular Biology and Virology Laboratory  
The Salk Institute  
La Jolla, California 92037

Chromosome inheritance must be amazingly efficient to ensure that each of the 100 trillion ( $10^{14}$ ) cells in the human body contains the full complement of 46 chromosomes required for normal viability and development. Mitotic chromosome inheritance requires successful completion of three basic functions—replication, sister chromatid cohesion and separation, and attachment to and movement along the spindle. Chromosomes contain many origins of DNA replication and sites of cohesion, which helps to ensure that these tasks are accomplished even when individual sites fail to function. However, in many organisms spindle attachment occurs at a single, localized region of each chromosome, the centromere (Figure 1A), whose chromosome location does not appear to change from one division to the next. What specifies the site of centromere function, and how does the centromere assemble the components necessary for spindle attachment and movement? Despite the essential role of centromeres in cell division, answers to these questions have baffled biologists for over a century.

The best-characterized centromere is found in the yeast *Saccharomyces cerevisiae*. In this unicellular eukaryote, centromere function is conferred by a small (125 bp) DNA sequence whose overall organization and sequence composition are highly conserved among the 16 different centromeres. The apparent dependence on primary DNA sequence is reminiscent of transcriptional regulation—specific sequences bind specific proteins, which in this case mediate chromosome movement rather than gene expression. In contrast, *Schizosaccharomyces pombe* and *Drosophila melanogaster* do not appear to contain such a “magic sequence” within their more complex and larger centromeres and may instead rely on higher order structure and epigenetic regulation rather than primary DNA sequence (reviewed in Karpen and Allshire, 1997; Wiens and Sorger, 1998 [this issue of *Cell*]).

Studies of human centromeres over the last decade have led to controversy concerning the role of a prominent candidate DNA sequence, the centromere-associated alpha satellite, also known as alphoid DNA. Here, we evaluate recent, exciting advances that address the following question—is alpha satellite necessary and sufficient for centromere function? We discuss growing evidence that centromere behavior in humans and other eukaryotes is paradoxical—centromeres are stably propagated at the same chromosomal location through replication and division, but also display remarkable plasticity (activation and inactivation). Finally, we explore the implications of epigenetic regulation of centromere activity to chromosome function and evolution.

### If It Looks like a Duck...

There is ample evidence implicating a role for alpha satellite in centromere function. The first supportive evidence is “guilt by association”; alpha satellite is the only known human DNA that is exclusively localized to the centromeric regions of all normal human chromosomes (Figure 1A). The arrays are composed of tandemly arranged 171 bp monomers organized into higher-order repeats and range in size from 200 to 9000 kb (Figure 1B). Secondly, the human X and Y centromeres have been localized using chromosome rearrangements found in natural populations or created artificially by fragmenting chromosomes with cloned telomeric DNA. In one study, two derivative chromosomes deleted for either most of the Y long or short arms shared only 70 kb of overlapping sequences that are predominantly composed of alpha satellite. All of the chromosome rearrangements analyzed to date retain at least 140 kb of alpha satellite, further implicating this sequence as having a role in centromere function (Heller et al., 1996, and references therein).

If guilt by association were sufficient, we would conclude that alpha satellite is the “magic sequence” that determines human centromere identity and function, analogous to the conserved elements present in the *S. cerevisiae* centromere. However, structural analyses are currently not detailed enough to exclude the importance of nonalphoid DNA. Long-range restriction mapping

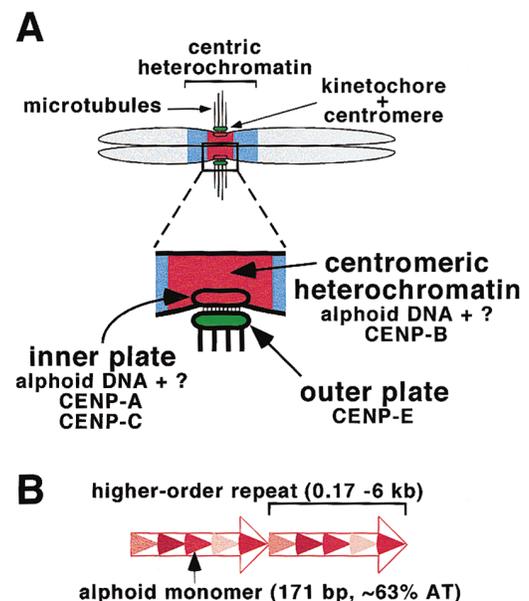


Figure 1. Structure and Organization of the Human Kinetochores and Centromeric DNA

Kinetochore refers to the trilaminar proteinaceous structure that mediates centromere–spindle associations and chromosome movement; and centromere describes a functional rather than a structural entity, specifically, the DNA required for kinetochore formation and spindle attachment. Red indicates alpha satellite, blue is flanking heterochromatin that contains other repeats.

suggests that the alpha satellite arrays are mostly homogeneous, but some interspersed sequences are present. These include LINES, Alu repeats, and other satellites (Lee et al., 1997), which are predominantly present in other parts of the genome that never associate with kinetochores, suggesting that they are not by themselves sufficient for centromere function. However, complete sequencing of alpha satellite arrays may identify functionally important interspersed sequences.

Critical centromeric sequences may be recognized by proteins located in the kinetochore inner plate, where DNA interfaces with the kinetochore (Figure 1A). CENP-A and CENP-C are only present at functional centromeres, and have been localized to the inner plate by coimmunofluorescence and immunoelectron microscopy, respectively (Warbuton et al., 1997, and references therein). CENP-A is a highly divergent H3-like histone that may form a specialized chromatin structure in the kinetochore, and CENP-C contains a novel DNA-binding motif. Vafa and Sullivan (1997) used chromatin immunoprecipitation to isolate the DNA associated with CENP-A *in vivo* and found a 10- to 20-fold enrichment of alpha satellite in comparison to its representation in control genomic libraries. These results demonstrate that alpha satellite is present in the kinetochore inner plate and associates with critical centromere proteins; therefore, alpha satellite is at the right place to determine centromere identity and function.

#### ...and It Waddles like a Duck...

If alpha satellite specifies the site of kinetochore formation, then it should be possible to create functional centromeres from cloned alphoid DNA, as demonstrated previously for *S. pombe* and *S. cerevisiae* centromeric DNAs. Harrington et al. (1997) devised an elegant *in vitro* concatamerization strategy to construct synthetic alpha satellite arrays up to 1 Mb in size, starting from a 2.7 kb alphoid clone. These arrays were joined to a selectable marker and lipofected into human cells along with telomeric and total genomic DNA. Some of the recovered clones contained 6–10 Mb minichromosomes bearing synthetic alpha satellite arrays, which are significantly larger than the starting material. Of four minichromosomes that were further characterized, at least two contained *de novo* centromeres that bound CENP-C and CENP-E.

Using a different approach, Ikeno et al. (1998) retrofitted a YAC containing ~100 kb of alpha satellite with nested human telomeres and introduced it into human cells. Most of the recovered clones contained small minichromosomes (~1–5 Mb) derived from the YAC DNA. Further analyses of three minichromosomes indicated that they had functional centromeres and were composed of ~30 copies of the original YAC. FISH with chromosome paint or satellite probes suggested that they had not acquired large amounts of additional host sequences; however, nonalphoid DNA could be present in the original YAC and be required for centromere formation.

What can we conclude from these reconstitution experiments? The production of minichromosomes from synthetic alphoid arrays suggests that alpha satellite may be sufficient for centromere function. In addition, not all alpha satellite clones were competent to produce

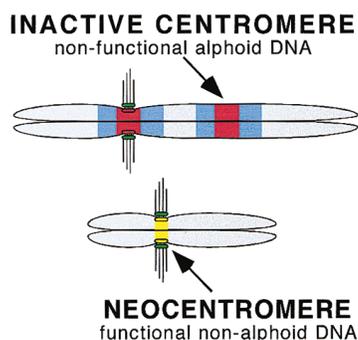


Figure 2. Examples of Human Centromere Plasticity  
Yellow denotes a euchromatic region that has acquired neocentromere activity.

artificial minichromosomes, suggesting that the type of substrate may be important. However, in both studies all *de novo* chromosomes contained substantial rearrangements of the introduced arrays. Rearrangements may be required to incorporate nonalphoid genomic sequences necessary for *de novo* centromere function or to achieve a minimum overall size necessary for a noncentromeric component of human chromosome function. Furthermore, the kinetics of *de novo* centromere formation could not be determined, and it is possible that a rare activation step is required in addition to the presence of alpha satellite. Nevertheless, these studies constitute important breakthroughs that will help answer many extant questions about the substrate “rules” for *de novo* centromere formation in humans. It is exciting that the relative importance of alpha satellite can now be assessed directly by comparing the frequency of *de novo* minichromosome assembly with different, defined DNA substrates, such as alphoid repeats versus other repeats versus nonrepetitive DNA.

#### ...It Could Still Be a Duck-Billed Platypus, and Other Birds Can Fly

The localization of alpha satellite to kinetochores and the successes in generating *de novo* centromeres from alphoid clones suggest that alpha satellite encodes human centromere function. However, colocalization experiments indicate that only a subset of the alpha satellite array at each centromere is associated with CENP-A, CENP-C, and the kinetochore inner plate (Warbuton et al., 1997), suggesting that identical alphoid sequences can differ in their ability to bind critical kinetochore proteins. In addition, spatially separated alpha satellite arrays on a single chromosome can differ in function. Normally chromosomes that contain two active centromeres (dicentric) are unstable because they frequently attach to opposite spindle poles and produce chromatin bridges in anaphase. However, numerous stable dicentric chromosomes have been described in which one of the two alpha satellite blocks is inactivated and no longer binds critical kinetochore proteins (Figure 2) (Sullivan and Schwartz, 1995, and references therein). Similarly, several groups have analyzed the centromere function of arrays of alphoid DNA constructs integrated into human, simian, or hamster chromosomes. For example, Warbuton and Cooke (1997, and references

therein) demonstrated that arrays of human alphoid DNA integrated into hamster chromosomes did not bind CENP-C or CENP-E, and did not associate with spindle microtubules during mitosis. Anaphase bridges were observed, but they were caused by delayed sister chromatid separation in the region of the amplified arrays rather than centromere activity of the integrated alphoid DNA. Thus, alphoid DNA in stable dicentrics or integrated into ectopic sites is not sufficient for kinetochore formation.

Centromeres can also function without alphoid DNA. At least 17 different human marker chromosomes have been described that lack detectable amounts of alphoid DNA by FISH. Somehow, these chromosomes have compensated for loss of the normal centromere by forming a "neocentromere" on nonalphoid sequences (Figure 2), indicated by the presence of CENP-A, -C, or -E. du Sart et al. (1997, and references therein) recently localized one neocentromere to an 80 kb region of chromosome 10 euchromatin. Extensive restriction analyses indicated that the structure of this region had not changed upon neocentromere acquisition, suggesting that the same sequences can have two different functional states. Thus, alphoid DNA is not necessary for (neo)centromere function. To our knowledge, human neocentromere activation has not been observed in chromosome rearrangements induced experimentally by irradiation or telomere-associated fragmentation, suggesting that it requires a rare activation step.

#### **How Do Birds Fly? Models for the DNA Determinants of Human Centromere Identity and Function**

Although a role for unidentified sequences in centromere function cannot be ruled out, the available evidence suggests that alpha satellite plays an important role in human centromere function. Kinetochores contain alpha satellite DNA, and transformation with alphoid clones can produce functional artificial chromosomes. However, neocentromere activation and centromere inactivation demonstrate that alpha satellite is not absolutely necessary or sufficient for centromere function.

These observations present a paradox—how can centromeres consistently and efficiently form at the same sites, without an absolute requirement for specific primary DNA sequences, and also exhibit remarkable plasticity? One possibility is that centromere activity is normally specified by the primary sequence of alphoid DNA, and rare regulatory mechanisms or independent pathways are responsible for inactivation and neocentromere activation. However, inactive or active states are efficiently maintained despite the presence or absence of alphoid DNA, respectively, suggesting that these events reflect the normal mechanism for human centromere function.

We prefer a more parsimonious model, which requires the fewest elements to account for both centromere plasticity and stability. Human centromere function could be determined by an epigenetic system that prefers specific substrates. In this model (Figure 3), de novo centromere formation requires an activation step that "marks" or "imprints" the DNA. Initial activation may be biased towards some characteristic of alpha satellite,

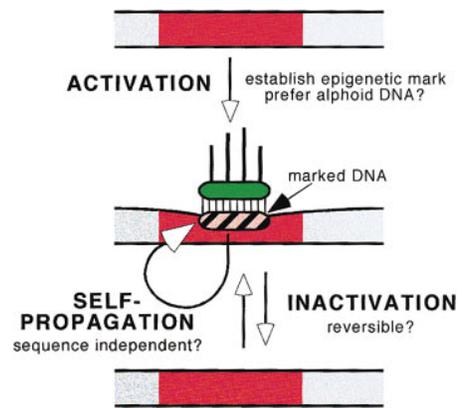


Figure 3. Model for Epigenetic Specification of Centromere Identity  
Small white arrows indicate less frequent events.

such as high A+T composition, repetitiveness, or secondary structure, but other sequences can also be activated, perhaps at a lower frequency (neocentromere formation). Once marked, any sequence can self-propagate the mark and function as a centromere in the next cell division (centromere and neocentromere propagation/maintenance). Centromere inactivation would occur by random reversal of the activation/marking or propagation steps.

Epigenetic specification of centromere identity may be a universal feature of eukaryotic regional centromeres. Recent studies have indicated that epigenetic mechanisms influence centromere function in *S. pombe* and *Drosophila* (reviewed in Karpen and Allshire, 1997; Wiens and Sorger, 1998). Evolutionary comparisons suggest that primary sequences are not important for centromere activity. Centromere-associated satellite DNAs show no obvious sequence conservation among fungi, plants, insects, mammals, and other vertebrates, and differ considerably even in closely related species. The sequences tend to be A+T-rich, have monomer subunits of nucleosome size, and possess a natural DNA curvature, but these are general characteristics of many satellite sequences. These observations can be explained if centromere function is specified by an epigenetic mechanism that prefers the overall DNA composition or repetitiveness of satellite DNA.

How could the epigenetic mark be specified and maintained during DNA replication? Active *S. pombe* centromeres contain underacetylated histones. Aberrant centromeric acetylation patterns are correlated with abnormal centromere function; these states are epigenetically maintained for many cell divisions (Ekwall et al., 1997). Another attractive paradigm is suggested by the fact that late replication of human alpha satellite coincides with the G2 peak of CENP-A expression, and misexpression of CENP-A throughout S phase results in nonspecific localization of CENP-A (Shelby et al., 1997). Perhaps a chromatin duplication mechanism acts in late S phase to copy the centromere-specific pattern of acetylation or CENP-A incorporation (reviewed in Karpen and Allshire, 1997).

#### **Implications**

At first glance, there are significant advantages for the individual cell and organism to contain centromeres that

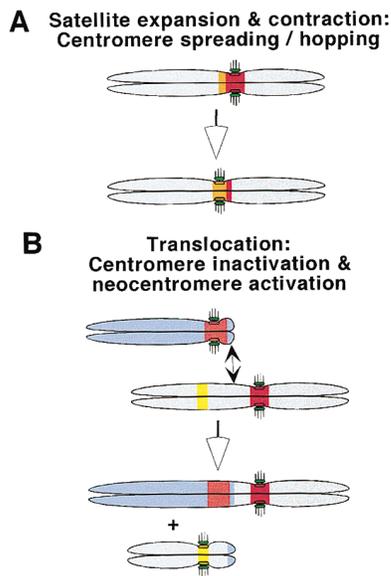


Figure 4. Evolutionary Changes in Centromeric Sequences and Function  
Orange and red denote different satellite arrays.

are dependent entirely on primary DNA sequence; given the importance of centromeres to cell and organismal viability, there should be no room for gain or loss of centromere function. Then why would centromeres utilize epigenetic mechanisms of regulation? Perhaps because it is adaptive and advantageous during evolution.

Epigenetic mechanisms could help buffer organisms from frequent and possibly detrimental changes to the DNA sequence. In addition to frequent base changes, satellite arrays readily expand and contract, changing the relative representation of neighbors (Figure 4A). A primary sequence-independent, self-propagation mechanism would tolerate changes in DNA sequence, because the centromere would simply follow the rule "love the one you're with...." Alternatively, an epigenetic mechanism may even be adaptive. For example, if centromere function per se results in recombination suppression, the resulting accumulation of ephemeral satellites in the centromeric region would then favor evolution of sequence-independent mechanisms of centromere specification.

Centromere plasticity could also facilitate chromosome evolution. Related species frequently differ in the arrangement and association of chromosome arms, even when the DNA sequences are nearly identical. Chromosome parts are exchanged by fusion and fission events—the requirement for one, and only one, centromere would render many of the resulting dicentric and acentric chromosome rearrangements useless unless centromeres could be inactivated and neocentromeres activated (Figure 4B). Thus, the ability to move the centromere from one DNA sequence to another, by spreading, hopping, or activation (Figures 4A and 4B) may expedite chromosome evolution. Do centromeres move within a chromosome, in evolutionary or even cellular timescales? Perhaps our perception of the absolute stability of centromere location is false. Most centromeres

have been localized by light microscopy, which does not provide the resolution necessary to answer this question. Functional centromeric DNAs need to be identified and mapped precisely within populations, and within extensive phylogenies that contain both closely related and distant species that are accessible to experimental manipulation.

Lessons learned from centromere studies may be relevant to understanding other chromosomal functions, such as homolog pairing, long-distance regulation of gene expression, and replication. For example, replication origins in *S. cerevisiae* are competent to function when cloned and reintegrated into chromosomes; this does not appear to be true for mammalian origins. Perhaps other chromosomal functions in higher eukaryotes have, like centromeres, been resistant to comprehension because these biological processes are determined by epigenetic mechanisms; if so, new tools and ways of thinking need to be developed before we can completely understand chromosome behavior and inheritance. The determinants of centromere identity and function are being investigated intensively in humans and other eukaryotes, and in the coming years we expect to gain a deeper understanding of this essential and complex biological function.

#### Selected Reading

- du Sart, D., Cancilla, M.R., Earle, E., Mao, J.I., Saffery, R., Tainton, K.M., Kalitsis, P., Martyn, J., Barry, A.E., and Choo, K.H. (1997). *Nat. Genet.* **16**, 144–153.
- Ekwall, K., Olsson, T., Turner, B.M., Cranston, G., and Allshire, R.C. (1997). *Cell* **91**, 1021–1032.
- Harrington, J.J., Van Bokkelen, G., Mays, R.W., Gustashaw, K., and Willard, H.F. (1997). *Nat. Genet.* **15**, 345–355.
- Heller, R., Brown, K.E., Burgtorf, C., and Brown, W.R. (1996). *Proc. Natl. Acad. Sci. USA* **93**, 7125–7130.
- Ikeno, M., Grimes, B., Okazaki, T., Nakano, M., Saitoh, K., Hoshino, H., McGill, N.I., Cooke, H., and Masumoto, H. (1998). *Nat. Biotech.*, in press.
- Karpen, G.H., and Allshire, R. (1997). *Trends Genet.* **13**, 489–496.
- Lee, C., Wevrick, R., Fisher, R.B., Ferguson-Smith, M.A., and Lin, C.C. (1997). *Hum. Genet.* **100**, 291–304.
- Shelby, R.D., Vafa, O., and Sullivan, K.F. (1997). *J. Cell Biol.* **136**, 501–513.
- Sullivan, B.A., and Schwartz, S. (1995). *Hum. Mol. Genet.* **4**, 2189–2197.
- Vafa, O., and Sullivan, K.F. (1997). *Curr. Biol.* **7**, 897–891.
- Warburton, P.E., and Cooke, H.J. (1997). *Chromosoma* **106**, 149–159.
- Warburton, P.E., Cooke, C.A., Bourassa, S., Vafa, O., Sullivan, B.A., Stetten, G., Gimelli, G., Warburton, D., Tyler-Smith, C., Sullivan, K.F., Poirier, G.G., and Earnshaw, W.C. (1997). *Curr. Biol.* **7**, 901–904.
- Wiens, G.R., and Sorger, P.K. (1998). *Cell* **93**, this issue, 313–316.