DETERMINING CENTROMERE IDENTITY: CYCLICAL STORIES AND FORKING PATHS

Beth A. Sullivan, Michael D. Blower*‡ and Gary H. Karpen**

The centromere is the genetic locus required for chromosome segregation. It is the site of spindle attachment to the chromosomes and is crucial for the transfer of genetic information between cell and organismal generations. Although the centromere was first recognized more than 120 years ago, little is known about what determines its site(s) of activity, and how it contributes to kinetochore formation and spindle attachment. Recent work in this field has supported the hypothesis that most eukaryotic centromeres are determined epigenetically rather than by primary DNA sequence. Here, we review recent studies that have elucidated the organization and functions of centromeric chromatin, and evaluate present-day models for how centromere identity and propagation are determined.

ANEUPLOIDY
The presence of extra copies, or no copies, of some chromosomes.

CENTROMERE
The genetic locus required for chromosome segregation; contains DNA and proteins on which the kinetochore is formed.

**Molecular and Cell Biology Laboratory, The Salk Institute for Biological Studies, 10010 N. Torrey Pines Road, La Jolla, California 92037, USA.*
‡Department of Biology, University of California at San Diego, La Jolla, California 92037, USA.
Correspondence to G.H.K.
e-mail: karpen@salk.edu

> "… I imagined a labyrinth of labyrinths, a maze of mazes, a twisting, turning, ever-widening labyrinth that contained both past and future…"[1]

Chromosome inheritance should be understood because it affects the lives of all cells and organisms. Without proper chromosome segregation, functional germ cells and zygotes would not form, and somatic tissues and organs would not develop and differentiate. ANEUPLOIDY results from aberrant chromosome inheritance, and is an important component of human birth defects and an underlying cause of tumour progression[2,3]. Despite the key role of chromosome inheritance in cell and organism viability and function, we know surprisingly little about the fundamental molecular mechanisms that mediate inheritance and how they go wrong in disease states. Nevertheless, this is an exciting time in chromosome biology. Recently developed optical tools, fluorescent labelling methods and real-time analyses, combined with studies of genetically tractable model organisms, have significantly advanced our understanding of chromosome structure and behaviour. However, more remains to be learned about this complex, mysterious and essential biological process.

The CENTROMERE is central to the chromosome inheritance process. It was originally defined in 1880 by Walther Flemming as a cytologically visible 'primary' constriction in the chromosome[4]. In the early 1900s, centromeres were defined genetically as chromosomal sites that were essential for normal inheritance and as regions of greatly reduced or absent meiotic recombination. We now know that the centromere is the site of kinetochore formation, the proteinaceous structure on each chromosome (FIG. 1) that is responsible for their attachment to and movement along microtubules; the centromere is therefore essential for chromosomal PLATEWARD PROMETAPHASE and POLEWARD ANAPHASE movements[5]. The terms centromere and kinetochore are commonly used interchangeably, and many functions, DNA sequences and proteins have been called centromeric, even when there has been no evidence that they have a centromeric function. As optical methods and functional analyses have improved, it has become evident that the CENTROMERE REGION is structurally and functionally more complex than was previously thought. In this review, we use the term 'centromere' to refer specifically to the chromatin (DNA and proteins) that is responsible for kinetochore formation, and the term 'centromere region' in reference to the domains and functions present in the

© 2001 Macmillan Magazines Ltd

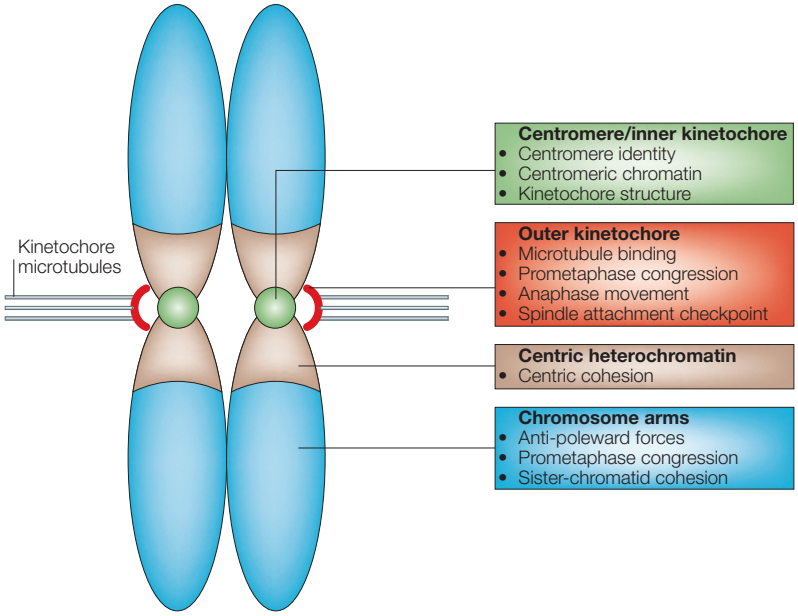


Figure 1 | **Structural and functional elements of the centromere region.** The centromere/inner kinetochore, outer kinetochore, centric heterochromatin and chromosome arms, and associated functions, are shown.

vicinity of a centromere (centric HETEROCHROMATIN plus centromere, FIG. 1). The outer kinetochore (FIG. 1) encodes specific functions and is specifically responsible for microtubule interactions. It also mediates the transition from metaphase to anaphase, serving as the site of action for the SPINDLE ASSEMBLY CHECKPOINT (SAC)[6]. Defects in proteins involved in this checkpoint occur in human tumours, which indicates an important role for specific centromere functions in disease[7]. In addition, specialized sister-chromatid cohesion functions are associated with the centromeric region and are essential for proper reductional meiosis I divisions[8,9].

Centromeres/kinetochores come in several sizes and forms[10] (FIG. 2). Most eukaryotes have MONOCENTRIC chromosomes, the size of which can vary markedly between species (FIG. 2a–d). HOLOCENTRIC chromosomes (FIG. 2e) are present in organisms such as nematodes and crayfish[11]. Organisms such as centipedes have both holocentric and monocentric chromosomes in the same nuclei[10]. How and why such an essential cellular component has evolved into different structures and whether functional mechanisms are conserved between these different structures remains unclear (we discuss this issue further towards the end of this review).

Generally, we understand the role that the kinetochore has in mediating chromosome movement; it recruits the microtubule motor proteins, DYNEINS and KINESINS, which probably determine the direction and speed of chromosome movement along the microtubules[5]. We also know that key molecules involved in the SAC (such as Mad2 (mitotic arrest deficient 2) and Bub1 (budding uninhibited by benzimidazoles 1)) are mostly concentrated in the outer kinetochore, where they are appropriately positioned to signal the start of anaphase once all kinetochores are attached to the spindle[6]. However, it is not clear how the centromere acts to form the kinetochore or how and why sites destined for centromere function are chosen (centromere identity). What is the relative importance of primary DNA sequence compared with EPIGENETIC mechanisms? Which proteins are essential to centromere identity? How is centromere identity replicated and what regulates the cycle of centromere propagation?

This review describes recent advances in the study of molecular mechanisms of centromere identity. We provide a brief overview of centromeric DNAs and present evidence that shows that they are neither necessary nor sufficient for kinetochore formation in multicellular eukaryotes. We then discuss recent studies that have described subdomains in the centromere region that mediate different inheritance functions. Finally, we evaluate present-day models that account for the determination and propagation of centromere identity. Outer kinetochore functions, chromosome movement, and sister-chromatid cohesion and separation are outside the focus of this review (see REFS 5,12 for recent reviews on these topics).

Forking paths to centromere identity

Many studies have focused on identifying 'the' sequences that confer centromere function (see REF. 13 and the references therein). The essential nature of centromeres and their stable location in the chromosome had indicated that specific primary DNA sequences might determine centromere identity and propagation. These studies identified regions involved in centromere function by several approaches, for example by screening for mutations that caused chromosome loss and that silenced centromeric marker genes, by functionally analysing chromosomal deletions and by *de novo* centromere assembly.

The best-defined centromeric DNAs are in *Saccharomyces cerevisiae*. These simple, 'point' centromeres — 16 in total (one per chromosome) — consist of an essential, conserved 125-bp sequence that comprises three functional elements that recruit centromere proteins and organize them into cytologically invisible kinetochores[14] (FIG. 2a).

Initially, it was expected that this dependence on primary sequence for centromere function in *S. cerevisiae* would exist in other organisms; however, it is notably absent in other eukaryotes, although some functional and structural characteristics of centromeric domains are conserved. Generally, centromeres are surrounded by or embedded in heterochromatin, a repeat-rich, gene-poor region of the genome that normally represses the transcription of euchromatic genes (in a process called POSITION EFFECT VARIEGATION)[15]. In *Schizosaccharomyces pombe*, non-homologous 4–5-kb central core sequences are flanked by various inverted repeats that are shared between the three chromosomes[16–18] (FIG. 2b). A minimum sequence of 25 kb, which contains the non-repetitive central core, inner repeats and a portion of the outer repeats, is an absolute requirement for centromere function and for stable chromosome transmission[19–21]. Deleting the inner

PLATEWARD METAPHASE MOVEMENTS
(prometaphase congression). The movement of condensing chromosomes towards the metaphase plate, during which they capture spindle microtubules and orientate themselves in preparation for anaphase sister-chromatid separation.

POLEWARD ANAPHASE MOVEMENTS
The movement of chromatids along spindle microtubules towards the spindle poles.

CENTROMERE REGION
Chromatin and DNA in the vicinity of the functional centromere, including the pericentric heterochromatin.

HETEROCHROMATIN
A cytologically defined genomic component that contains repetitive DNA (highly repetitive satellite DNA, transposable elements and ribosomal DNA gene clusters) and some protein-coding genes. Most eukaryotic centromeres are embedded in heterochromatin.

© 2001 Macmillan Magazines Ltd

repeats compromises meiotic sister-chromatid segregation[20], showing that centromeric regions function in processes other than kinetochore assembly, and that kinetochore and cohesion domains are closely linked and important for proper chromosome segregation (see below). Two studies have been instrumental in defining the boundaries[22] and functional subdomains[23] of centromere 1 of *S. pombe* by assaying for the expression of marker genes that have been inserted across the centromere (discussed in more detail below).

SPINDLE ASSEMBLY CHECKPOINT
(SAC). A highly conserved surveillance mechanism in mitosis and meiosis that minimizes chromosome loss by preventing chromosomes from initiating anaphase until all kinetochores have successfully captured spindle microtubules.

MONOCENTRIC
When a kinetochore forms on a specific, limited region of a chromosome.

HOLOCENTRIC
When a kinetochore forms along the entire length of a chromosome.

DYNEINS
Microtubule-based molecular motors that move towards the minus end of microtubules.

KINESINS
Microtubule-based molecular motors that, in general, move towards the fast-growing, plus end of microtubules.

EPIGENETIC
Any heritable influence (in the progeny of cells or of individuals) on chromosome or gene function that is not accompanied by a change in DNA sequence. Examples of epigenetic events include mammalian X-chromosome inactivation, imprinting, centromere inactivation and position effect variegation.

POSITION EFFECT VARIEGATION
(PEV). The variable, heritable suppression of genes by their juxtaposition to heterochromatin or telomeres, or by movement of a gene into a different nuclear domain or chromosomal context.

MINICHROMOSOME
An extranumerary chromosome that contains functional elements, such as telomeres and centromeres, and is transmitted in meiosis and mitosis.

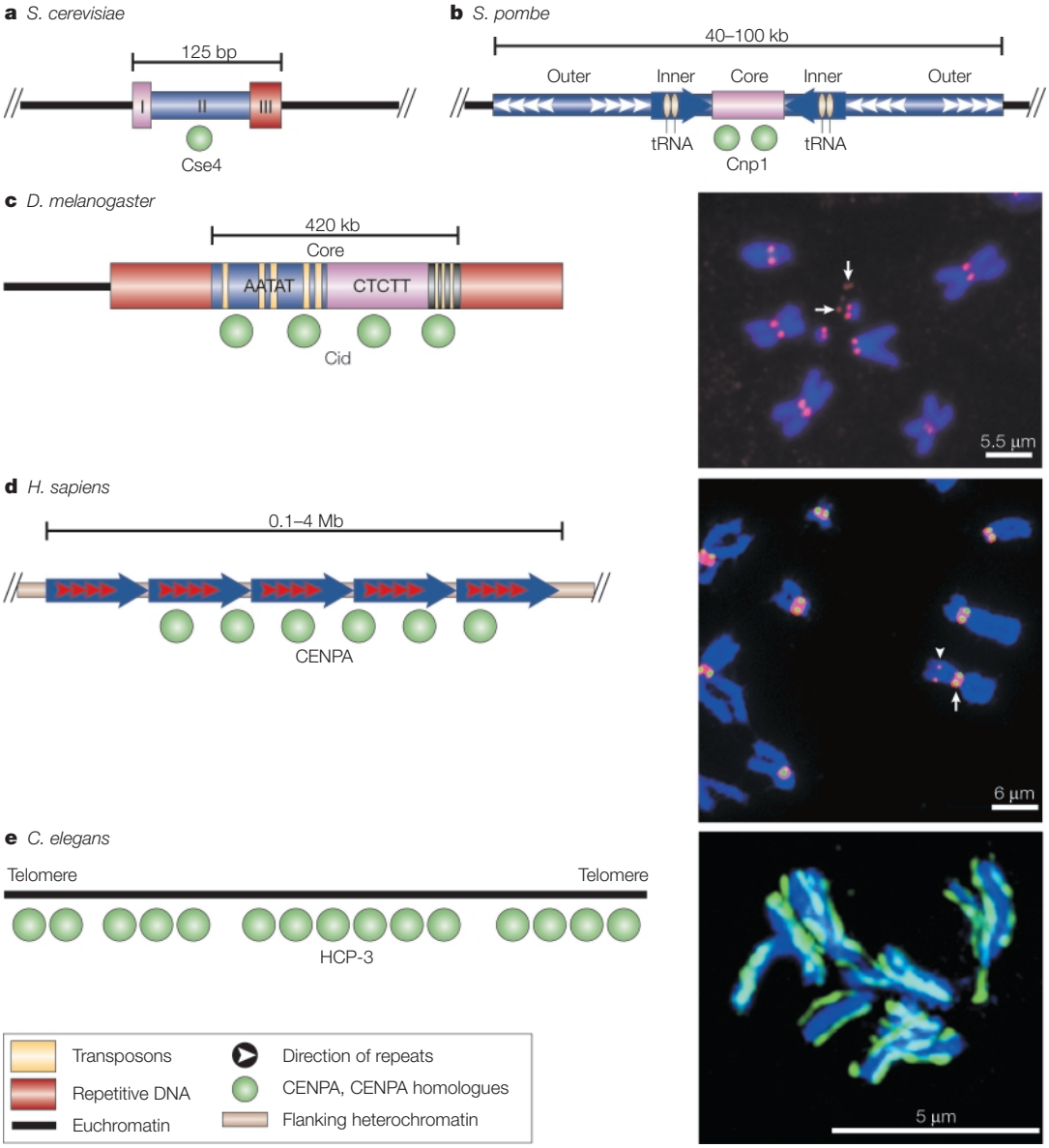


Figure 2 | **Centromeric DNAs and conservation of CENPA at structurally distinct centromeres.** The DNA sequence of centromeres differs between species, but the presence and function of CENPA and its homologues (shown in green) at kinetochores is highly conserved (**c**–**d** (right), chromosomes shown in blue). **a** | *Saccharomyces cerevisiae* centromere function depends on a region that contains three conserved elements (I, II, III), to which Cse4 localizes. **b** | *Schizosaccharomyces pombe* centromeres contain a unique central core, which Cnp1 localizes to, flanked by conserved inverted inner and outer repeats. **c** | The MINICHROMOSOME *Dp1187*, with the only defined *Drosophila melanogaster* centromere, consists of a core of 5-bp satellites and transposons, flanked by other repetitive DNA (red). Right, Cid (red) at endogenous fly centromeres and at minichromosomes (arrows). **d** | Human centromeres consist of alpha-satellite DNA (red arrows) tandemly arranged into higher-order repeats (blue arrows), which extend over megabases. CENPA localizes to a portion of these arrays. Right, a human cell line that contains a dicentric chromosome with one active (arrow) and one inactive (arrowhead) centromere. CENPB (red), a centromeric alpha-satellite-binding protein, is present at both centromeres. CENPA (green) locates to only endogenous centromeres and the active centromere of the dicentric chromosome. **e** | *Caenorhabditis elegans* kinetochores assemble along the length of each chromosome. Right, at metaphase, the centromeric histone HCP-3 (green) is present on the poleward-facing side of chromatids. CENPA, centromere-specific histone H3-like; Cid, Centromere identifier; Cnp1, *S. pombe* homologue of CENPA; Cse4, *S. cerevisiae* homologue of CENPA; and HCP-3, *C. elegans* homologue of CENPA. (Images courtesy of B. Sullivan (**c**,**d**), and L. Moore and M. Roth (**e**).)

© 2001 Macmillan Magazines Ltd

Table 1 | **Proteins involved in the functions of the centromere region***

Location	Function	*S. cerevisiae*	*S. pombe*	*C. elegans*	*D. melanogaster*	Mammals
Heterochromatin	Heterochromatin formation		Swi6 Chp1		Su(var)2-5/Hp1 Prod	HP1 (humans), M31 (mouse)
	Histone H3 methyltransferase		Clr4		Su(var)3-9	SUVAR39H1 (humans) Suvar39h1,2 (mouse)
	Sister-chromatid cohesion	Mcd1, Scc1	Rad21, Mis4,6,12		dRad21, Scc1 Mei-S332	SCC1/hRAD21
Centromeric chromatin	Centromere-specific histone H3	Cse4	Cnp1	HCP-3	Cid	CENPA (human) Cenpa (mouse)
	CENPA loading		Mis6			
Kinetochore	Inner plate	Mif2	Cnp3	HCP-4		CENPC (human) Cenpc (mouse)
	Outer plate			HCP-1,-2	Cmet Cana Zw10 Rod	CENPE (human),CENPF Cenpe (mouse)

*Proteins included in this table are only those referred to in this review. Known homologues (rows) are designated for various organisms (columns). For a more comprehensive list of proteins involved in chromosome inheritance, and in centromere and heterochromatin function, please see REF. 101. Full species names: *Saccharomyces cerevisiae*; *Schizosaccharomyces pombe*; *Caenorhabditis elegans*; *Drosophila melanogaster*. Cana, CenpE anaphase; CENPA, centromere protein A; CENPC, centromere protein C; CENPE, centromere protein E, kinesin-like; CENPF, centromere protein F; Chp1, chromodomain protein 1; Cid, Centromere identifier; Clr4, cryptic loci regulator 4; Cmet, CenpE metaphase; Cnp1,3, centromere proteins 1,3; HCP-1,-2,-3,-4, holocentric centromere proteins; Hp1, heterochromatin protein 1; M31, mammalian HP1; Mei-S332, meiotic-S332; Mif2, minichromosome fidelity 2; Mis4,6,12, mis-segregation 4,6,12; Prod, Proliferation disrupter; Rad21/Scc1/Mcd1, cohesins; Rod, Rough deal; Scc1/ SCC1, colon tumour susceptibility; Su(var)2-5,3-9/SUVAR39H1/Suvar39h1,2, suppressors of variegation (Note that the methyltransferase activity of *Drosophila* Su(var) 3-9 has not yet been shown); Swi6, switching gene 6; Zw10, zeste-white 10.

Drosophila centromeres are also located in repetitive DNA. So far, the only molecularly and functionally defined *Drosophila* centromere is on a 1.3-Mb X-derived minichromosome, *Dp1187* (REFS 24,25) (FIG. 2c), progressive deletions of which have defined a 420-kb region that genetically confers normal chromosome inheritance.

The primary repetitive DNA at human centromeres is alpha-satellite DNA, which consists of a 171-bp monomer that is tandemly arranged into higher-order arrays that extend for 100 kb to several megabases[26] (FIG. 2d). The entire alpha-satellite array is unlikely to be involved in kinetochore assembly because antibodies to kinetochore proteins, such as CENPA (see TABLE 1 and below), localize to only a portion of the alpha-satellite DNA[27]. In addition, chromosomes that are naturally or artificially deleted for much of this array can still assemble a kinetochore and segregate normally[28–30]. Furthermore, not all alpha-satellite arrays form centromeres *de novo* (when in the form of human artificial chromosomes)[31,32], indicating that other sequences or factors might be required to assemble and maintain functional human centromeres. However, *de novo* chromosome assembly seems to be most efficient when alpha-satellite DNA is introduced into human cell lines[31,32], indicating that this repetitive DNA might be the best template for new centromere formation in cultured cells.

Arabidopsis thaliana centromeres have only lately been localized to 500–1000-kb regions that contain 180-bp repeats, which are bounded by ribosomal DNA arrays and other repetitive sequences[33–36]. Genetic mapping and recombination suppression have been used to localize these centromeres, rather than kinetochore function. These studies might therefore have overestimated the size or location of the centromere, because recombination suppression is a property of the flanking heterochromatin, rather than an intrinsic function or characteristic of the site of kinetochore formation[37,38]. Nevertheless, *A. thaliana* studies have provided important descriptions of the structure and sequence composition of plant centromeric regions. Beet (genus *Beta*) centromeric regions have recently also been shown to consist of satellite DNA and transposable elements[39]. Two beet minichromosomes that contain a subset of satellites found at normal centromeres show differential rates of meiotic transmission and should help to identify the minimal DNA elements required for *Beta* kinetochore assembly.

Finally, in the nematode *Caenorhabditis elegans*, the holocentric chromosomes recruit and assemble centromeric proteins along their lengths (FIG. 2e). Specific DNA sequences are seemingly not required as concatamers of phage lambda DNA and many other types of DNA are stably transmitted[40]. Proteins are recruited into 'bundles' at prophase, then spread evenly on the poleward face of the chromosome arms by metaphase, which indicates that many areas of the *C. elegans* genome can support kinetochore assembly.

Inactivated centromeres and neocentromeres

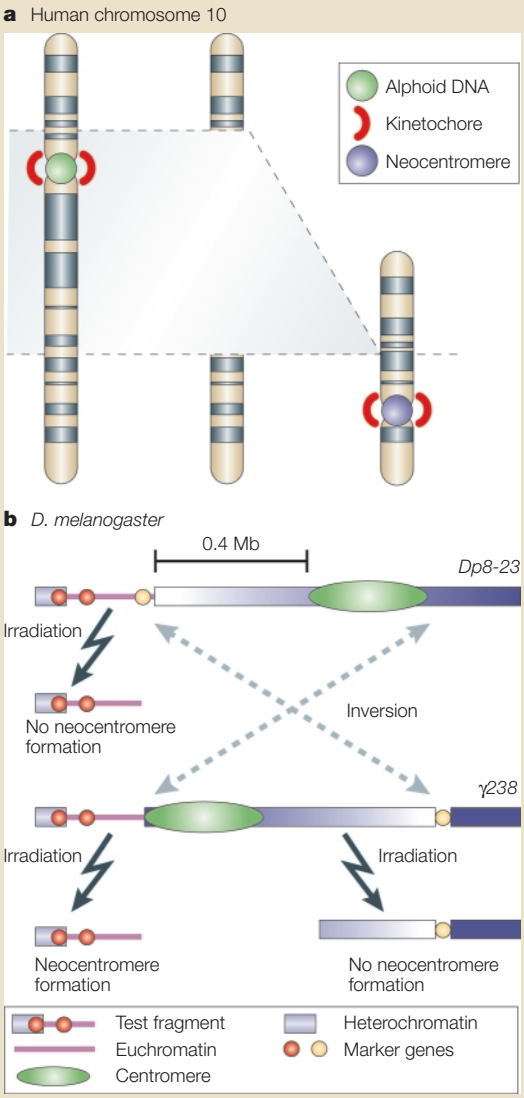
The simple presence of centromeric DNA in a cell or on a chromosome does not automatically correlate with centromere function, as shown in humans and in flies. Dicentric chromosomes are engineered or naturally occurring abnormal chromosomes that contain two regions of centromeric DNA (for example, alpha-satellite DNA in humans) that are capable of functioning as centromeres[41,42]. More than 60 years ago, Barbara McClintock described aberrant segregation of

CENPA
A centromere-specific, histone H3-like protein.

© 2001 Macmillan Magazines Ltd

Box 1 | **Neocentromere formation in humans and in flies**

A naturally occurring human neocentromere derived from chromosome 10 is shown in part a of the figure. Human neocentromeres are often associated with gross chromosomal rearrangements; in this case, a large interstitial deletion that removed the middle part of chromosome 10. The order of events in this process is unclear; chromosomal deletions might occur first, followed by neocentromere formation (as shown), or neocentromere formation might occur on an intact chromosome, forming a dicentric that would then undergo rearrangements. Experiments have shown that the formation of experimentally induced neocentromeres in *Drosophila melanogaster* requires proximity to a functional centromere[57]. Part b of the figure shows the behaviour of a 320-kb euchromatic 'test fragment' in different contexts: the *Dp1187* minichromosome derivative, *Dp8-23*; and its inversion derivative, *γ238* (dashed lines/arrows show inversion breakpoints). Neocentromeres are only produced after the irradiation-induced breakage of *γ238*. They do not form when the test fragment is located far from the endogenous centromere before chromosome breakage, as in *Dp8-23*. Neocentromeres were also not recovered from the centric heterochromatin to the right of the centromere. These results indicate that neocentromere formation in flies might occur owing to the spreading of centromeric chromatin onto adjacent euchromatic DNA, but only when the flanking heterochromatin is eliminated. These studies also showed that once normally non-centromeric chromatin acquires neocentromere function, it is faithfully inherited through mitosis and meiosis.



dicentric chromosomes in maize, characterized by ANAPHASE BRIDGING, chromosome breakage and chromosome loss[43]. However, dicentric chromosomes in humans (FIG. 2d) and in flies can be stably transmitted owing to the functional inactivation of one centromere[27,41,44–46] or to centromere cooperation, when both centromeres remain functional and presumably act coordinately for kinetochore formation, microtubule attachment and anaphase segregation[41,42]. These studies explain the stable transmission of structurally dicentric chromosomes and show that the presence of centromeric DNA on a chromosome is not sufficient for centromere function.

Centromeric DNAs are not always necessary for kinetochore formation, as shown by human and *Drosophila* chromosomes that lack typical centromeric DNA but can assemble kinetochores at sites called NEOCENTROMERES, and that are mitotically and meiotically stable[47–50]. (Normally, ACENTRIC fragments are lost during cell division because they cannot attach to the spindle.) Human neocentromeres are found on naturally occurring, small, rearranged marker chromosomes (see BOX 1), which are typically ascertained through prenatal screening or from cytogenetic analyses of individuals with developmental delay or congenital abnormalities[47]. These phenotypic abnormalities are probably caused by trisomy or tetrasomy for the genes that are encoded by the supernumerary chromosomes, rather than by the presence of a neocentromere[47,50]. More than 40 human neocentromeres that involve 15 different chromosomes have been identified so far, indicating that many genomic regions are amenable to centromere activation or that any sequence can be activated as a centromere under certain conditions. A detailed study has shown that 22 centromere/kinetochore/centromere-region proteins are present at normal human centromeres and at neocentromeres[51], which indicates that neocentromeres might mediate inheritance through the same proteins and mechanisms as normal centromeres.

ANAPHASE BRIDGES
The physical stretching of dicentric chromosomes during anaphase due to the orientation and movement of linked kinetochores towards opposite spindle poles.

NEOCENTROMERE
Chromosomal sites that do not contain typical repetitive centromeric DNA but do acquire centromeric chromatin, can assemble kinetochores, can recruit other centromeric proteins, and are transmitted faithfully in meiosis and mitosis.

ACENTRIC
A chromosome or chromosomal fragment that lacks a centromere.

© 2001 Macmillan Magazines Ltd

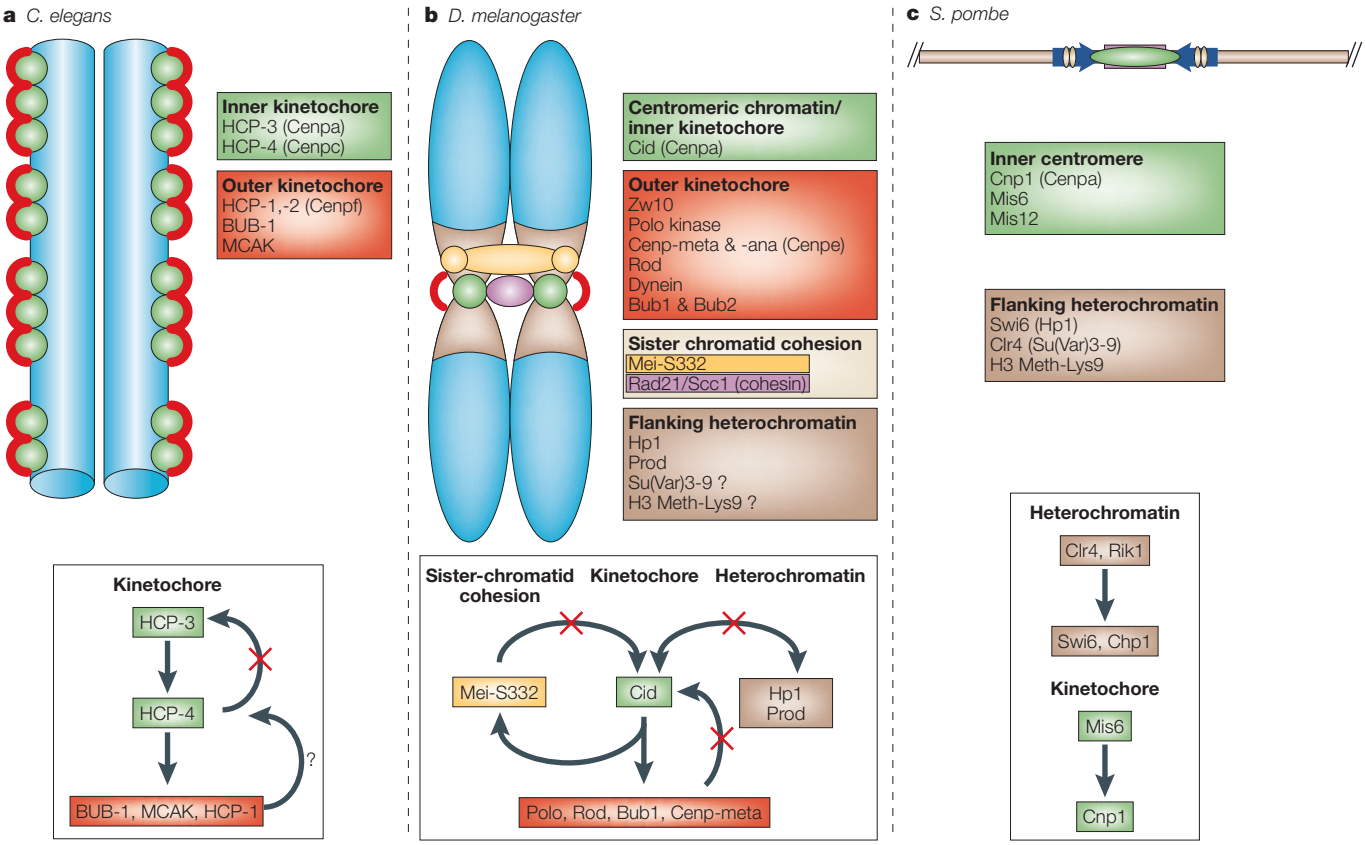


Figure 3 | **Structural and functional analyses of protein subdomains in the centromere region.** The schematics show the location of different centromere-region proteins with respect to the DNA sequence, and the boxes summarize the epistatic relationships between them (arrows show localization dependency and red crosses indicate no dependency). Cnp1, Cid and HCP-3 are CENPA homologues; other homologies are indicated in parentheses. **a** | Prometaphase *Caenorhabditis elegans* holocentric chromosomes contain inner and outer kinetochore proteins distributed in clusters. Recruitment of outer kinetochore proteins is dependent on the presence of HCP-4, which requires HCP-3 for its localization, but not vice versa. **b** | *Drosophila melanogaster* centromere proteins. Cid is required for the recruitment of all outer kinetochore proteins and a sister-chromatid cohesion protein (Mei-S332), but not for that of proteins that associate with the flanking heterochromatin. Cid localizes to centromeres in the absence of these outer-kinetochore, flanking-heterochromatin and sister-cohesion proteins. **c** | *Schizosaccharomyces pombe* centromere proteins bind either the inner centromere (the centromere core and inner repeats), or the flanking heterochromatin. Mutation studies (see text) show these domains are functionally distinct. Mis6 is required to localize Cnp1; Clr4 and Rik1 are required to localize Swi6 and Chp1. BUB-1/Bub1,2, budding uninhibited by benzimidazoles 1; Chp1, chromodomain protein 1; Cid, Centromere identifier; Clr4, cryptic loci regulator 4; Cnp1, centromere protein 1; H3 Meth-Lys9, methylated lysine 9 of histone H3, mediated by Su(var)3-9 homologues; Hp1, heterochromatin protein 1; MCAK, mitotic centromere-associated kinesin; Mei-S332, meiotic-S332; Mis6,12, mis-segregation 6,12; Prod, Proliferation disrupter; Rod, Rough deal; Su(Var)3-9, suppressor of variegation 3-9; Swi6, switching gene 6; Zw10, zeste-white 10.

Neocentromeres are often identified cytologically but only a few have been analysed molecularly. Recent studies[52–54] of human neocentromeric regions on three different chromosomes have defined the regions of CENPA binding to ~400 kb, remarkably similar to the size of the *Drosophila* minichromosome centromere[52–54]. None of the activated neocentromeric regions differed in sequence composition from the parental homologous loci and the regions also did not acquire alpha-satellite DNA. Furthermore, these neocentromeres did not show significant homology to each other, indicating that different sequences can acquire neocentromeric function. Interestingly, all three neocentromeres were significantly (A+T)-rich (>60%) and enriched for retroviral elements, long terminal repeats and/or short tandem repeats. It is possible that (A+T)-rich DNA or repetitive arrays can more easily achieve the conformations required for centromere function and kinetochore assembly[55].

How are these normally non-centromeric regions activated? It is difficult to analyse the mechanism(s) of neocentromere formation in humans, as such cases are quite rare and presumably occur during meiosis in a parent. However, neocentromeres that bind centromere and kinetochore proteins can be experimentally induced on normally non-centromeric, euchromatic DNA after irradiation-induced breakage of a *Drosophila* minichromosome[24,49] (BOX 1). In such experiments, neocentromere formation on a 'test fragment' that normally lies more than 40 Mb from the endogenous X centromere was assessed in different chromosomal contexts. The test fragment was only able to become a neocentromere if it was adjacent to the functional centromere of the minichromosome before irradiation[49,56,57]. Therefore, neocentromere activation in *Drosophila* is thought to occur by the spreading *in cis* of centromeric proteins into adjacent, non-centromeric regions when heterochromatic BOUNDARY ELEMENTS (see

BOUNDARY ELEMENTS
Chromatin that acts as an insulator to block changes in chromatin structure, protein binding, or the spreading of functional domains.

© 2001 Macmillan Magazines Ltd

HISTONE
A family of small, highly conserved basic proteins, found in the chromatin of all eukaryotic cells, that associate with DNA to form a nucleosome.

CREST ANTISERA
Autoantibodies that recognize centromeric antigens (CENPs) found in the sera of patients with autoimmune diseases, such as CREST (calcinosis, Raynaud syndrome, oesophageal dysmotility, scleroderma and telangiectasia).

NUCLEOSOME
The fundamental unit into which DNA and histones are packaged in eukaryotic cells.

RNA INTERFERENCE
(RNAi). A process by which double-stranded RNA specifically silences the expression of homologous genes by degrading their cognate mRNA.

CENPC
A constitutive kinetochore protein. Its localization to the inner kinetochore is dependent on CENPA.

MEI-S332
A centromere-region protein involved in sister-chromatid cohesion.

SWI6
A chromodomain-containing *Schizosaccharomyces pombe* homologue of *Drosophila* heterochromatin protein 1. The chromodomain is a protein motif — common to proteins that in some cases interact with chromatin — that is involved in binding certain methylated histones; often associated with transcriptional repression.

CHP1
A *Schizosaccharomyces pombe* chromodomain protein.

MIS6
A centromere protein that binds to the central core of the *Schizosaccharomyces pombe* centromere and is required to establish or maintain centromeric chromatin structure.

CHROMATIN IMMUNOPRECIPITATION
(ChIP). A technique that isolates sequences from soluble DNA chromatin extracts (complexes of DNA and protein) using antibodies that recognize specific chromosomal proteins.

below) are removed[57]. The mechanisms of human neocentromere formation are less clear, but most human neocentromeres are reported to occur at sites that are significantly distant from the endogenous centromeres, as judged from metaphase chromosome analyses. Activation by the spreading of centromere proteins *in cis* might not be the mechanism for neocentromere formation in humans. However, as the nucleus is dynamically compartmentalized[58–60], perhaps neocentromere activation in humans occurs *in trans*, through the inappropriate nuclear positioning of normally non-centromeric regions in centromeric nuclear domains[56,57].

Centromere plasticity is surprising; centromeric DNAs are not sufficient for centromere function and non-centromeric DNA occasionally acquires, then faithfully propagates, centromere function. And yet centromeres are essential elements that usually show a stable location and function. Neocentromere activation cannot be a frequent occurrence, otherwise monocentric chromosomes would become di- or multicentric and would be lost or become holocentric. Nevertheless, these observations indicate that centromere identity might be established and propagated at a specific and consistent chromosomal site by epigenetic mechanisms[13,55].

The epigenetic determination and self-propagation of centromere identity can account for both centromere plasticity and stability. A centromeric epigenetic mark could be specified by exclusive protein binding, HISTONE modifications and/or by the spatial and temporal organization of chromosomes or chromosomal processes, such as replication (as discussed in more detail below). However, the epigenetic determination of centromere identity and propagation, and the apparent lack of dependence on primary DNA sequence, do not rule out the possibility that sequence composition, such as enrichment for repeat sequences or an AT sequence bias, also has a role in these processes.

Subdomains of centromeric chromatin

***Specification by* CENPA.** The designation of the site of centromere formation in higher eukaryotes might be determined through an epigenetic mechanism, but the specifics of this process remain unclear[13]. CENPA — a component of the centromere/kinetochore (TABLE 1) — was originally identified as an antigen that was recognized by human CREST ANTISERA; subsequent biochemical and molecular analyses showed it to be a histone-H3-related protein[61–63]. CENPA proteins are present in yeasts (*S. cerevisiae*[64,65] and *S. pombe*[66]), *C. elegans*[67] and *Drosophila*[68], representing an evolutionary link between the seemingly divergent centromeres of these organisms (FIG. 2). NUCLEOSOMES can be assembled *in vitro* from purified CENPA and from histones H2A, H2B and H4 (REF. 69), consistent with previous observations, which indicates that CENPA nucleosomes are homotypic *in vivo* (that is, they contain two copies of CENPA, not one copy of H3 and one of CENPA)[70]. The homology of CENPA proteins to a core chromatin component (histone H3) and their presence at centromeres throughout the cell cycle make them a strong candidate for a protein that specifies and propagates the site of kinetochore assembly.

Several recent studies highlight the importance of CENPA proteins in the establishment and function of kinetochores in various organisms. Inactivating CENPA proteins in yeasts, worms, flies and mammals severely disrupts mitosis and cell-cycle progression[64,71,72]. Furthermore, inactivating or deleting CENPA causes the mislocalization of many kinetochore and centromere-region proteins: knocking it out in mice or inhibiting its homologue in *C. elegans* by RNA INTERFERENCE (RNAi) abolishes the ability of CENPC proteins to target the kinetochore[72–74]. By comparison, disrupting CENPC in *C. elegans* has no effect on CENPA targeting (FIG. 3a). So, CENPA is upstream of CENPC in the kinetochore-assembly pathway in worms and in mice. In addition, the outer-kinetochore proteins in the worm — BUB-1, MCAK and HCP-1 — are mislocalized when CENPA is inhibited by RNAi[73,74]. Similarly, many *Drosophila* centromere-region and outer-kinetochore proteins (such as Polo, Bub1, Cmet (CENP-meta), Rod (Rough deal) and MEI-S332 (Meiotic-S332)) are mislocalized in embryos and in cultured cells when the *Drosophila* homologue of CENPA (called Cid (Centromere identifier)) is inhibited by antibodies or by RNAi[71] (FIG. 3b). These data indicate that CENPA is a central component in kinetochore formation and centromere function. Understanding the biochemical mechanisms and proteins required to propagate CENPA-containing chromatin during or after DNA replication should improve our understanding of how centromere identity and function are maintained (see below).

S. pombe *centromere subdomains*. Recent studies in *S. pombe* have indicated that the centromere is not a single locus, as previously thought, but has an unexpectedly fine structure (FIG. 3c). It is repressive for transcription at the inner and outer repeats and at the central core (FIG. 2b), but transcriptional silencing at the central core and at the outer repeats are mediated by different proteins[23]. Mutations in SWI6 and CHP1 (TABLE 1) alleviate the silencing of transgenes that have been inserted into the inner and outer repeats, but not of those inserted into the central core. Conversely, a mutation in MIS6 (TABLE 1) alleviates transgene silencing in the central core but not in the flanking repeats. Consistent with the genetic analysis of centromeric silencing, CHROMATIN IMMUNOPRECIPITATION (ChIP) analysis has shown that the centromere region is composed of non-overlapping protein domains: the inner and outer repeats contain Swi6, and Chp1, whereas the central core contains Mis6 (REF. 23) (FIG. 3c). The central core, which is the site of kinetochore formation and spindle attachment[75], also contains Cnp1, the *S. pombe* homologue of CENPA (REF. 66). Proteins located in the outer repeats and in the central core are also essential for chromosome transmission, because mutations in *mis6*, *mis12* (REF. 76), and *swi6* (REF. 77) lead to increased chromosomal loss and sister-chromatid non-disjunction. The inner-centromere and flanking-heterochromatin domains seem to be separated by boundary elements, perhaps encoded by tRNA genes (FIGS 2b, 3c). The *S. pombe* centromere regions are excellent models for further investigation into the biochemical mechanisms responsible for

© 2001 Macmillan Magazines Ltd

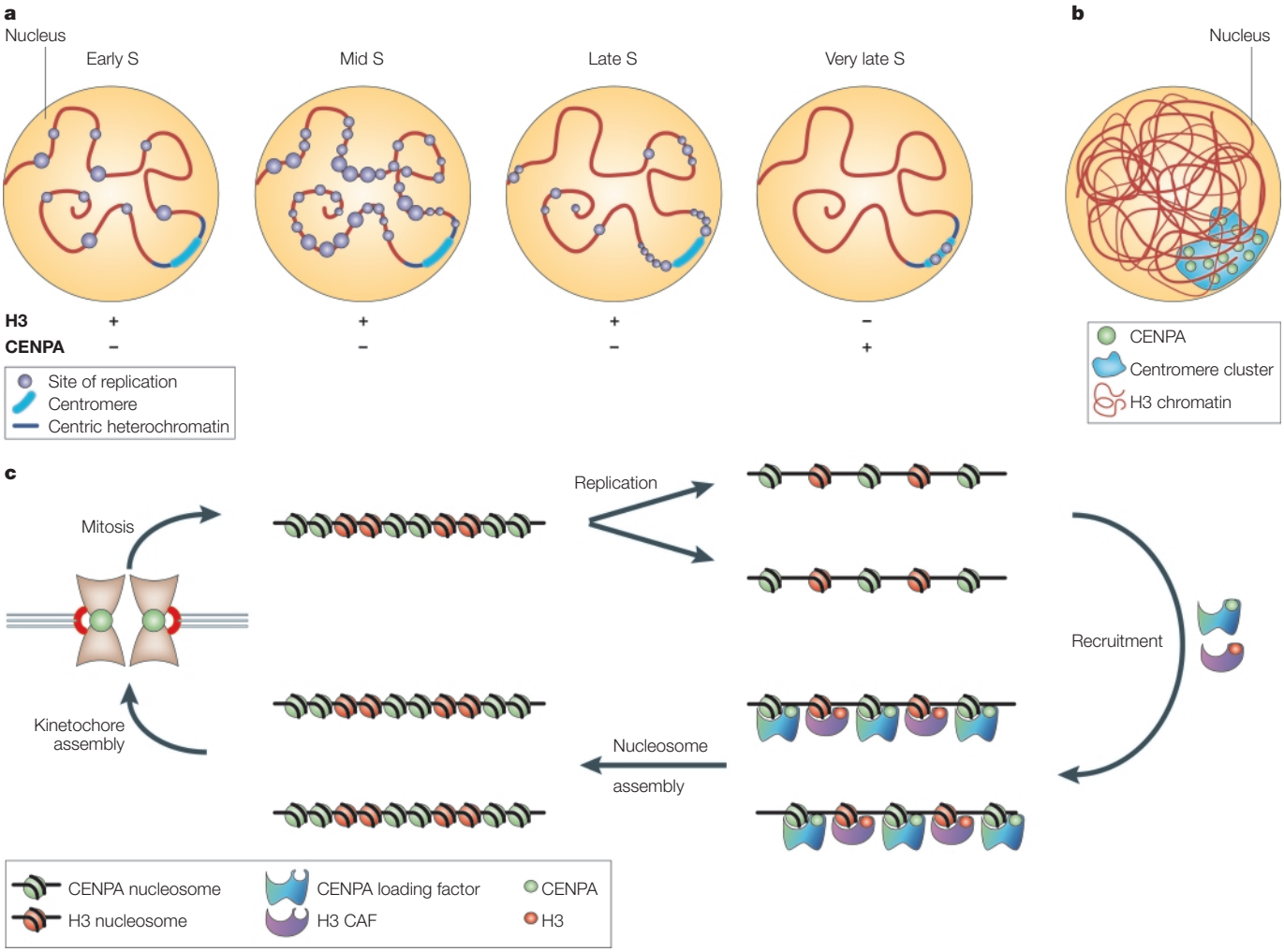


Figure 4 | **Models for the propagation of centromere identity.** CENPA deposition determines centromere identity and propagation in these models, but other molecules or epigenetic marks could be responsible for these functions. **a** | The 'last to replicate' model. Centromeres are the only sequences replicated very late in S phase, when H3 is absent and CENPA is present. The opposite occurs in the 'earliest to replicate' model: CENPA is only present when centromeres replicate early in S (not shown). **b** | Nuclear-organization model. Centromeres are clustered and sequestered into one or more centromeric 'domains', proposed to be the only site(s) of CENPA assembly in the nucleus. **c** | Cyclical chromatin assembly model. During replication, CENPA- and H3-containing nucleosomes segregate to daughter chromatids. Chromatin assembly factors (CAFs) or other loading factors then recruit CENPA and H3 to sites that already contain the appropriate histone to replenish nucleosome numbers. CENPA and H3 recruitment is unlikely to occur simultaneously as H3 assembly is coupled to replication, and CENPA assembly is not (see text). Nucleosome assembly and subsequent kinetochore assembly transmit centromeric chromatin through mitosis, where the replication and replenishment cycle begins again.

establishing and maintaining centromere region subdomains. In fact, recent studies have begun to elucidate the biochemical mechanisms involved in implementing and preserving the protein composition of the centromeric region[78–81]. For example, the selective methylation of histone H3 lysine 9 by SU(VAR)3-9 homologues (TABLE 1) has recently been found to create specialized binding sites for heterochromatin proteins, such as HP1 (heterochromatin protein 1, also called Su(Var)2-5), that are involved in marking chromatin states and organizing chromosomal domains[82].

Drosophila *centromere subdomains.* *D. melanogaster* centromeres are similarly organized. The positioning of centromeres in centric heterochromatin in most higher eukaryotes indicates that centromere function might require some property of heterochromatin. However, cytological studies in flies have shown that Cid-containing chromatin is flanked by, but does not overlap with, chromatin that contains proteins involved in centric heterochromatin function (Hp1) and centromere-region condensation (Prod (Proliferation disrupter)) (FIG. 3b and TABLE 1). Furthermore, mutations in *Hp1* and *prod* do not affect Cid localization and vice versa[71]. So, Cid deposition at centromeres is independent of at least two key heterochromatin proteins. The presence of distinct centromere region domains is also supported by the observation that inhibiting the INNER CENTROMERE PROTEIN, Incenp, does not affect Cid deposition in *Drosophila*[83]. Similarly, in worms, INCENP is not dependent on HCP3 or HCP4 for its kinetochore localization and vice versa[74] (FIG. 3a). In *Drosophila*, Mei-S332 (REF. 8) (TABLE 1) is

HETEROCHROMATIN PROTEIN 1 (Hp1). A *Drosophila* heterochromatin protein that contains chromodomains.

SU(VAR)3-9
A chromatin-binding translation initiation factor that suppresses position effect variegation in flies. Mammalian and *Schizosaccharomyces pombe* Su(Var)3-9 homologues are present in the centromere region of mitotic chromosomes, and encode a histone H3 methyltransferase.

© 2001 Macmillan Magazines Ltd

INNER CENTROMERE PROTEIN (Incenp). A family of proteins that transiently localize to the region between sister-chromatid kinetochores during mitosis.

located near but not in the Cid-containing chromatin[71] (FIG. 3b), which provides a physical basis for a previous demonstration that kinetochore function and Mei-S332-mediated cohesion can be separated using minichromosome derivatives[84]. Interestingly, Mei-S332 localization depends on the presence of functional Cid, but the absence of functional Mei-S332 has no effect on Cid centromeric localization[71] (FIG. 3b). These analyses also showed that Cid has a key role in recruiting all outer-kinetochore proteins (see above) and a centric sister-chromatid cohesion protein. Likewise, in *S. cerevisiae* the localization of the cohesion protein, Scc1, also depends on an active kinetochore[85]. So, the centromere region consists of many, functionally distinct subdomains and additional studies are necessary to further elucidate their functional interactions.

Centromere identity and DNA replication

Do the apparent distinctions between yeast and multicellular eukaryotes, and between monocentric and holocentric chromosomes, reflect differences in the mechanisms that determine centromere identity? Possibly. However, we can begin to explain these differences in a unified fashion by thinking about the propagation of centromere identity as a cyclical, epigenetic process. For instance, a region might be specified to form a centromere because it was packaged as centromeric chromatin in the previous cell cycle[13,55]. To understand centromere identity and its propagation cycle, we first must understand how centromeric chromatin, defined by the presence of CENPA, is replicated. Typically, new histones are deposited along DNA as it is synthesized during the S phase, owing to the activity of chromatin assembly factors (CAFs)[86,87]. Is centromere propagation dependent on the incorporation of CENPA into nucleosomes in place of H3? If so, then centromere propagation would be undermined by CENPA being incorporated into non-centromeric regions or by it being replaced by H3 at centromeres.

Several models have indicated that centromere identity might be propagated because CENPA-containing chromatin is temporally, spatially or physically separated from sites of bulk DNA replication and/or histone deposition (FIG. 4a, b). In the 'last to replicate' model, CENPA is exclusively incorporated into centromeres as they are replicated in very late S phase, at a time when H3-containing chromatin is no longer replicating[70,88] (FIG. 4a). Another replication-timing model proposes that centromeres could be the first to replicate, before the replication of bulk H3-containing chromatin[89]. Nuclear-organization models have also been proposed, in which the specificity of CENPA incorporation into nucleosomes is accomplished by the physical sequestration of centromere-assembly factors into isolated 'domains' inhabited by clusters of centromeres[89] (FIG. 4b), which prevent the inclusion of H3 into centromeric nucleosomes. Replication-timing and nuclear-organization models are not mutually exclusive, and another model has been proposed in which centromere sequestration and replication timing combine to propagate centromere identity[89].

In *S. cerevisiae*, centromeric DNA has been reported to replicate in early S phase[90]. Cse4 (the *S. cerevisiae* homologue of CENPA) is expressed at low levels throughout the cell cycle[65], but it is unclear at what point it is loaded onto chromatin and if centromere assembly is uncoupled from DNA replication. It is noteworthy that centromeres are not the earliest sequences to replicate in *S. cerevisiae*. In fact, they have been found to replicate approximately one-third of the way through the S phase, at a time when many non-centromeric DNAs are also replicated[90]. In *S. pombe*, Cnp1 expression occurs from M to the G1/S boundary, before the peak of H3 synthesis at the onset of S phase[66]. Mis6 is required for proper localization of Cnp1 to centromeres (FIG. 3c) and acts during G1/S or at the G1/S boundary, therefore Cnp1 is probably incorporated into nucleosomes before S phase commences. The precise replication timing of centromeric DNA in *S. pombe* has not been reported; however, the implication from both yeast studies is that CENPA is incorporated into centromeric nucleosomes independently of DNA replication.

Models that link centromere identity and propagation with very early or very late replication timing[88,89] have been directly tested using DNA-labelling methods based on the incorporation of thymidine analogues into replicating DNA. Studies of CENPA expression and the timing of centromeric chromatin replication in cultured human cells[91] have shown that histone H3 expression peaks in early to mid-S phase, and CENPA mRNA and protein levels are highest in late S phase and G2. Inappropriate expression of CENPA in early S phase in human cells abolishes its specificity for centromeres, indicating that its synthesis and/or deposition might be regulated differently from bulk histones to allow for its exclusive incorporation into centromeric chromatin[70]. These results initially provided evidence in favour of the replication-timing model for CENPA incorporation (FIG. 4a). However, human CENPA-associated chromatin has recently been shown to be neither the first nor the last region to be replicated; human centromeres replicate asynchronously in mid- to late S phase, simultaneously with non-centromeric DNA[91]. Furthermore, CENPA is actively assembled at centromeres in G2 (REFS 70,91). So, human centromere replication occurs broadly and asynchronously in S phase, before new CENPA is assembled into chromatin. As confirmation that centromeric chromatin assembly is uncoupled from replication, new CENPA was shown to be assembled into centromeres even when DNA replication was blocked[91]. Similarly, a recent study in *Drosophila* showed that Cid is incorporated into centromeres in the absence of DNA replication[89]. This study also indicated that centromeres in *Drosophila* cells cultured *in vitro* might replicate in early S phase and that H3 deposition is inhibited at this time. However, *Drosophila* centromeres have recently been found to replicate asynchronously *in vivo*, in mid- to late S phase, simultaneous with H3-containing chromatin[92].

If centromeric chromatin assembly in eukaryotes occurs independently of DNA replication timing, does physical sequestration of centromeres into unique nuclear domains allow exclusive incorporation of

© 2001 Macmillan Magazines Ltd

CENPA (FIG. 4b)? Studies in human and *Drosophila* interphase cells have shown that centromeres are not sequestered into specific replication domains or nuclear domains[91–93]. Data from yeast, *Drosophila* and human studies have also ruled out models of replication-dependent (FIG. 4a) or spatially dependent (FIG. 4b) deposition of CENPA and propagation of centromere identity. An alternative model, in which incorporation of CENPA into centromeric nucleosomes is achieved through the activity of certain factors, is discussed below.

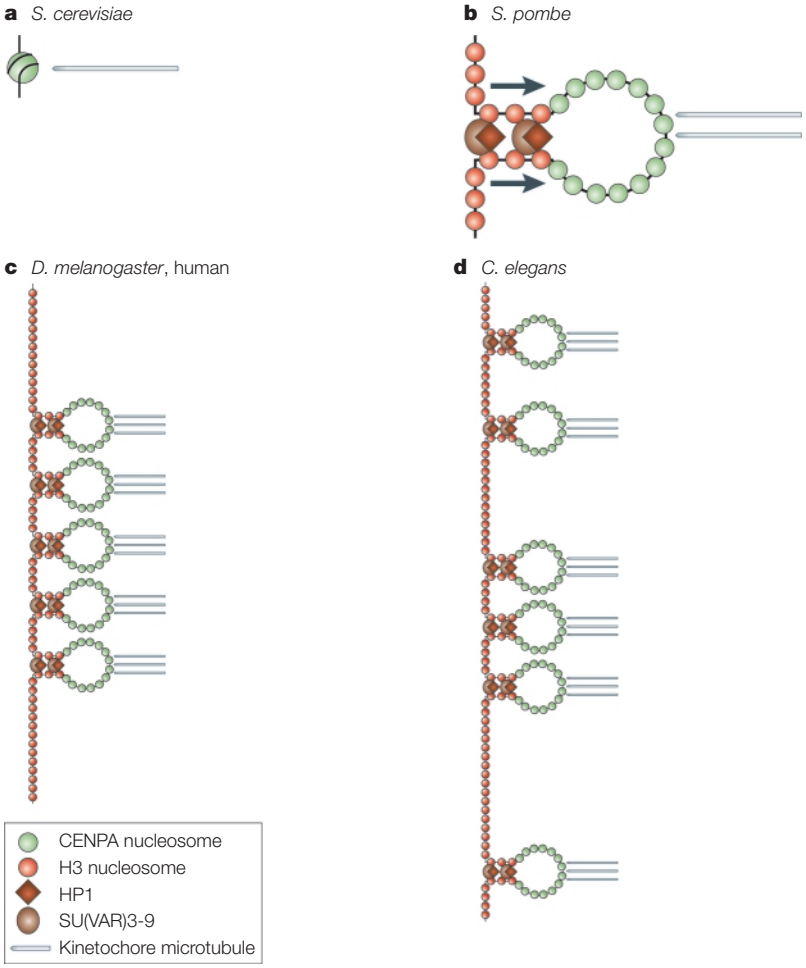


Figure 5 | **A repeat-subunit and presentation model for centromeric chromatin in different species. a** | The *Saccharomyces cerevisiae* centromere probably consists of a single, Cse4-containing nucleosome that gives rise to the kinetochore. **b** | The *Schizosaccharomyces pombe* centromere consists of multiple Cnp1-containing nucleosomes that are flanked by heterochromatin. The inverted-repeat structure of the centromere-flanking regions indicate that the centromere region might form a stem–loop structure[97], perhaps stabilized by heterochromatic proteins, such as homologues of HP1 and SU(VAR)3-9. **c** | The larger and more complex centromeres of *Drosophila melanogaster* and humans possibly contain repeats of similar 'loops', or related higher-order structures, that are concentrated at specific chromosomal regions. Because tandem, rather than inverted, repeats exist in fly and human centromeric DNA, they might not form stem–loop structures; the DNA might, for example, spiral through a centromeric higher-order structure. Flanking heterochromatin might form a boundary that limits the spreading of centromeric chromatin in monocentric chromosomes[57] (see BOX 1). Such boundaries might distinguish monocentric from holocentric centromeres. **d** | *Caenorhabditis elegans* holocentric chromosomes might contain more repeat units at multiple sites, covering most of the chromosome. In these models, the higher-order structure of centromeric chromatin is conserved, and might be required to 'present' the centromeric chromatin to the outer face of condensed chromosomes (see text for more details). CENPA, centromere-specific, histone H3-like protein; Cnp1, *S. pombe* homologue of CENPA; Cse4, *S. cerevisiae* homologue of CENPA; HP1, heterochromatin protein 1; SU(VAR)3-9, suppressor of variegation 3-9.

The centromere identity cycle

> "The only way I could surmise was that it be a cyclical, or circular, volume, a volume whose last page would be identical to the first, so that one might go on infinitely."[1]

Centromere identity studies started with ground-breaking experiments on the small centromeres of *S. cerevisiae*. The seemingly simple dependence of this centromere on specific DNA sequences for its identity, and the conservation of molecules such as CENPA, led us to believe that the centromeres of higher eukaryotes would consist of larger numbers of *S. cerevisiae*-like centromeres. We now know that higher eukaryotic centromeres are considerably larger than those in *S. cerevisiae*, and that the regulation of centromere identity in organisms as distant as *S. pombe*, flies and humans is dependent on epigenetic marking and epigenetic propagation rather than on primary DNA sequence.

The findings that human centromeres replicate asynchronously and that mammalian CENPA is synthesized in G2 have led to the proposal that the propagation of centromere identity might occur through the activity of a post-replication, CENPA-specific loading factor, perhaps equivalent to H3 CAFs. Such a centromere-identity loading factor might recognize CENPA-containing nucleosomes that were segregated to sister chromatids during replication and then deposit new CENPA onto each chromatid in the same region, replenishing the CENPA content of the chromatin[12] (FIG. 4c). Such a cyclical mechanism could account for the faithful propagation of both monocentric and holocentric centromeres, and the fact that once CENPA and neocentromere identity are acquired, they are propagated faithfully from one generation to the next[49,51,56]. The recruitment of flanking DNA into CENPA-containing chromatin can also account for the ability of centromeric chromatin to 'spread' or encroach onto non-centromeric DNA and form neocentromeres[56,57].

The key questions that remain to be answered are: whether a CAF or similar loading molecule exists that can deposit CENPA into centromeric nucleosomes; whether this function promotes the propagation of centromere identity; and how these processes relate to the establishment and propagation of centromere region subdomains. The mislocalization of Cse4 in *S. cerevisiae* in *ndc10* mutants[94] shows that the presence of CENPA at centromeres can be affected by other centromere proteins; however, controversy remains about the order in which *S. cerevisiae* kinetochore components are assembled[95–97]. *S. pombe* Mis6 (FIG. 3a and TABLE 1) is required for the proper localization of Cnp1 to the central core[66], and thus might be directly or indirectly part of the epigenetic mechanism that ensures the proper marking of the centromere. Identifying proteins that interact with CENPA should lead to a better understanding of the propagation of CENPA-containing chromatin and centromere identity.

© 2001 Macmillan Magazines Ltd

The inverted-repeat structure of the heterochromatin that flanks *S. pombe* centromeres and the behaviour of deletion constructs led to the proposal that the *S. pombe* centromere region might assume a stem–loop structure in which the core is in the loop, and the outer and inner repeats are in the stem[98] (FIG. 5b). The *S. pombe* centromere provides an attractive model for centromere organization on a small scale, but how can this model be extrapolated to the centromeres of higher eukaryotes that are much larger and how can this organization be reconciled with the holocentric chromosomes observed in some organisms? The repeat-subunit model of centromere organization proposed by Zinkowski *et al.*[99] provides an explanation for the evolutionary conservation of centromere proteins between the different centromere/kinetochore structures seen in many organisms. This model was originally proposed to explain the fact that CREST antibody staining appeared punctate and discontinuous in stretched chromosomes, whereas the underlying alpha satellite array was uniform, indicating that the centromere might not be a single entity but a series of basic, repeated subunits.

The repetitive and interspersed form of the centromere is especially obvious in the holocentric chromosomes of *C. elegans* (FIG. 5d) where HCP-3 (the *C. elegans* homologue of CENPA) is present in a few discrete foci in interphase cells, which then coalesce to cover the poleward face of the chromosomes during mitosis, forming a linear kinetochore ribbon[67] (FIG. 2e). Monocentric centromeres might be organized in a similar fashion (FIG. 5b, c), but the size and number of repeats might be limited by the presence of heterochromatic boundary elements. At *S. pombe* and *Drosophila* centromeres, each Cenpa-containing chromatin domain is flanked by heterochromatin that contains proteins such as Hp1 and Su(var)3-9 (FIG. 3). Because neocentromere formation in *Drosophila* requires proximity to a functional centromere and the absence of intervening heterochromatin (BOX 1), it is possible that the heterochromatin itself provides a boundary between centromeric and adjacent chromatin[57]. In addition, the heterochromatin could be responsible for looping the centromeric DNA[100] into a higher-order structure in such a way that the CENPA-containing chromatin is located on the poleward face of the chromosome and beneath the outer kinetochore proteins, 'presenting' them to the nucleoplasm to mediate microtubule interactions (FIG. 5c). This model can account for the differences in the localization of proteins and their different gene-silencing effects, the requirement for both heterochromatin and CENPA-containing chromatin for complete centromere function, the generation of neocentromeres and the evolution of holocentric chromosomes. Holocentric chromosomes could represent the first centromeres; kinetochore formation probably evolved first with random sequence specificity, followed by the evolution of monocentric chromosomes that arose owing to transposon invasion, to satellite DNA expansion and to the formation of flanking heterochromatin. However, it is also possible that holocentric chromosomes evolved from monocentric chromosomes, owing to the loss of heterochromatic boundary elements and by the *in cis* spreading of centromeric chromatin, in a manner analogous to the generation of *Drosophila* neocentromeres[57] (BOX 1).

Why are *S. cerevisiae* centromeres different from other well-studied centromeres, despite the presence of the CENPA orthologue Cse4? The apparent absence of epigenetic effects at *S. cerevisiae* centromeres might relate to the fundamental composition of the centromeric chromatin; unlike higher eukaryotes, it is believed that the *S. cerevisiae* centromere contains a single, Cse4-containing nucleosome[64] (FIG. 5a). Replication of a single-nucleosome centromere would produce daughter strands, of which only one sister chromatid is likely to contain a Cse4-containing nucleosome. A cyclical mechanism that requires the recognition of existing centromeric chromatin to deposit new CENPA nucleosomes (FIG. 4c) would not work if one chromatid completely lacked CENPA nucleosomes. So, a primary-sequence-dependent mechanism might be necessary to ensure that both products of replication can recruit Cse4 and other proteins. Nevertheless, it is important to note that no direct tests for epigenetic mechanisms have been reported for *S. cerevisiae* centromeres, so it is still possible that these simple centromeres are also subject to some kind of epigenetic regulation.

Conclusion

This is an exciting time for centromere biologists. The tools are now assembled to do in-depth analyses of the biochemical mechanisms that determine centromere identity and kinetochore formation. Studies aimed at identifying CENPA-interacting proteins, their biochemical activities, and the role of centromeric DNA are essential. In the next few years, we should have a significantly greater understanding of these fundamental biological processes, bringing this cycle of stories and forking paths to a more satisfying conclusion. Understanding the basic biology of centromere structure and function is also of relevance to human health: by identifying key molecules involved in birth defects and cancer[2,7], this research could lead to new diagnostic tools and treatments. Furthermore, a more complete understanding of centromeres and chromosome inheritance in multicellular eukaryotes will lead to the efficient, controlled use of artificial chromosome cloning vectors[26,31,32]; the ability to transfer large domains and genes into animal and plant cells as intact, functional chromosomes would be extremely useful for engineering livestock and crops, and for developing gene therapy vectors for use in humans.

Links

DATABASE LINKS Mad2 | Bub1 | CENPA | BUB-1 | MCAK | HCP-1 | Polo | Cmet | Rod | Mei-S332 | Cid | Swi6 | Chp1 | Mis6 | Cnp1 | *mis12* | Hp1 | Prod | INCENP | Scc1 | Cse4 | *ndc10* | HCP-3 | Clr4 | Rik1

FURTHER INFORMATION Walther Flemming | Barbara McClintock | Centromere research labs | Gary Karpen's lab

© 2001 Macmillan Magazines Ltd

1. Borges, J. L. in *Jorge Luis Borges: Collected Fictions* 125 (Penguin Books, New York City, 1998).
An example of metafiction dealing with cycles of life, written by a master of the genre.
2. Hassold, T. & Hunt, P. To err (meiotically) is human: the genesis of human aneuploidy. *Nature Rev. Genet.* **2**, 280–291 (2001).
3. Mitelman, F. *Catalog of Chromosome Aberrations in Cancer* (Wiley, New York, 1994).
4. Flemming, W. Beitrag zur Kenntnis der Zelle und ihrer Lebenserscheinungen, Teil II. *Archiv. Mikrosk. Anat.* **18**, 151–259 (1880).
The first clear cytological description of the centromere.
5. Rieder, C. L. & Salmon, E. D. The vertebrate cell kinetochore and its roles during mitosis. *Trends Cell Biol.* **8**, 310–318 (1998).
6. Shah, J. V. & Cleveland, D. W. Waiting for anaphase: Mad2 and the spindle assembly checkpoint. *Cell* **103**, 997–1000 (2000).
7. Cahill, D. P. *et al.* Mutations of mitotic checkpoint genes in human cancers. *Nature* **392**, 300–303 (1998).
8. Dej, K. J. & Orr-Weaver, T. L. Separation anxiety at the centromere. *Trends Cell Biol.* **10**, 392–399 (2000).
9. Toth, A. *et al.* Functional genomics identifies monopolin: a kinetochore protein required for segregation of homologs during meiosis I. *Cell* **103**, 1155–1168 (2000).
10. White, M. J. D. *Animal Cytology and Evolution* (Cambridge Univ. Press, Cambridge, UK, 1973).
11. Pimpinelli, S. & Goday, C. Unusual kinetochores and chromatin diminution in *Parascaris*. *Trends Genet.* **5**, 310–315 (1989).
12. Sullivan, K. F. A solid foundation: functional specialization of centromeric chromatin. *Curr. Opin. Genet. Dev.* **11**, 182–188 (2001).
13. Karpen, G. H. & Allshire, R. C. The case for epigenetic effects on centromere identity and function. *Trends Genet.* **13**, 489–496 (1997).
14. Hyman, A. A. & Sorger, P. K. Structure and function of kinetochores in budding yeast. *Annu. Rev. Cell Dev. Biol.* **11**, 471–495 (1995).
15. Weiler, K. S. & Wakimoto, B. T. Heterochromatin and gene expression in *Drosophila*. *Annu. Rev. Genet.* **29**, 577–605 (1995).
16. Clarke, L., Amstutz, H., Fishel, B. & Carbon, J. Analysis of centromeric DNA in the fission yeast *Schizosaccharomyces pombe*. *Proc. Natl Acad. Sci. USA* **83**, 8253–8257 (1986).
17. Nakaseko, Y., Adachi, Y., Funahashi, S., Niwa, O. & Yanagida, M. Chromosome walking shows a highly repetitive sequence present in all the centromere regions of fission yeast. *EMBO J.* **5**, 1011–1021 (1986).
18. Nakaseko, Y., Kinoshita, N. & Yanagida, M. A novel sequence common to the centromere regions of *Schizosaccharomyces pombe* chromosomes. *Nucleic Acids Res.* **15**, 4705–4715 (1987).
19. Baum, M., Ngan, V. K. & Clarke, L. The centromeric K-type repeat and the central core are together sufficient to establish a functional *Schizosaccharomyces pombe* centromere. *Mol. Biol. Cell* **5**, 747–761 (1994).
20. Hahnenberger, K. M., Carbon, J. & Clarke, L. Identification of DNA regions required for mitotic and meiotic functions within the centromere of *Schizosaccharomyces pombe* chromosome I. *Mol. Cell. Biol.* **11**, 2206–2215 (1991).
21. Murakami, S., Matsumoto, T., Niwa, O. & Yanagida, M. Structure of the fission yeast centromere cen3: direct analysis of the reiterated inverted region. *Chromosoma* **101**, 214–221 (1991).
22. Allshire, R. C., Nimmo, E. R., Ekwall, K., Javerzat, J. P. & Cranston, G. Mutations derepressing silent centromeric domains in fission yeast disrupt chromosome segregation. *Genes Dev.* **9**, 218–233 (1995).
23. Partridge, J. F., Borgstrom, B. & Allshire, R. C. Distinct protein interaction domains and protein spreading in a complex centromere. *Genes Dev.* **14**, 783–791 (2000).
By analysing chromatin content, this study showed that the *Schizosaccharomyces pombe* centromere is composed of spatially and functionally separate protein domains.
24. Murphy, T. D. & Karpen, G. H. Localization of centromere function in a *Drosophila* minichromosome. *Cell* **82**, 599–609 (1995).
25. Sun, X., Wahlstrom, J. & Karpen, G. Molecular structure of a functional *Drosophila* centromere. *Cell* **91**, 1007–1019 (1997).
References 24 and 25 report the first genetic and molecular definition of a functional centromere in a metazoan.
26. Willard, H. F. Centromeres: the missing link in the development of human artificial chromosomes. *Curr. Opin. Genet. Dev.* **8**, 219–225 (1998).
27. Warburton, P. E. *et al.* Immunolocalization of CENP-A suggests a distinct nucleosome structure at the inner kinetochore plate of active centromeres. *Curr. Biol.* **7**, 901–904 (1997).
28. Mills, W., Critcher, R., Lee, C. & Farr, C. J. Generation of an approximately 2.4 Mb human X centromere-based minichromosome by targeted telomere-associated chromosome fragmentation in DT40. *Hum. Mol. Genet.* **8**, 751–761 (1999).
29. Wevrick, R., Earnshaw, W. C., Howard-Peebles, P. N. & Willard, H. F. Partial deletion of alpha satellite DNA associated with reduced amounts of the centromere protein CENP-B in a mitotically stable human chromosome rearrangement. *Mol. Cell. Biol.* **10**, 6374–6380 (1990).
30. Yang, J. W. *et al.* Human mini-chromosomes with minimal centromeres. *Hum. Mol. Genet.* **9**, 1891–1902 (2000).
31. Harrington, J. J., Van Bokkelen, G., Mays, R. W., Gustashaw, K. & Willard, H. F. Formation of *de novo* centromeres and construction of first-generation human artificial microchromosomes. *Nature Genet.* **15**, 345–355 (1997).
32. Ikeno, M. *et al.* Construction of YAC-based mammalian artificial chromosomes. *Nature Biotechnol.* **16**, 431–439 (1998).
References 31 and 32 are landmark papers that describe the construction and mitotic stablility of human artificial chromosomes using synthetic (31) or cloned (32) alpha satellite arrays to seed kinetochore formation in tissue culture cells.
33. Murata, M., Ogura, Y. & Motoyoshi, F. Centromeric repetitive sequences in *Arabidopsis thaliana*. *Jpn. J. Genet.* **69**, 361–370 (1994).
34. Round, E. K., Flowers, S. K. & Richards, E. J. *Arabidopsis thaliana* centromere regions: genetic map positions and repetitive DNA structure. *Genome Res.* **7**, 1045–1053 (1997).
35. Copenhaver, G. P. *et al.* Genetic definition and sequence analysis of *Arabidopsis* centromeres. *Science* **286**, 2468–2474 (1999).
36. Heslop-Harrison, J. S., Murata, M., Ogura, Y., Schwarzacher, T. & Motoyoshi, F. Polymorphisms and genomic organization of repetitive DNA from centromeric regions of *Arabidopsis* chromosomes. *Plant Cell* **11**, 31–42 (1999).
37. Merriam, J. R. & Frost, J. N. Exchange and nondisjunction of the X chromosomes in female *Drosophila melanogaster*. *Genetics* **49**, 109–122 (1964).
38. Szauter, P. An analysis of regional constraints on exchange in *Drosophila melanogaster* using recombination-defective meiotic mutants. *Genetics* **106**, 45–71 (1984).
39. Gindullis, F., Desel, C., Galasso, I. & Schmidt, T. The large-scale organization of the centromeric region in beta species. *Genome Res.* **11**, 253–265 (2001).
40. Stinchcomb, D. T., Shaw, J. E., Carr, S. H. & Hirsh, D. Extrachromosomal DNA transformation of *Caenorhabditis elegans*. *Mol. Cell. Biol.* **5**, 3484–3496 (1985).
41. Sullivan, B. A. & Willard, H. F. Stable dicentric X chromosomes with two functional centromeres. *Nature Genet.* **20**, 227–228 (1998).
42. Higgins, A. W., Schueler, M. G. & Willard, H. F. Chromosome engineering: generation of mono- and dicentric isochromosomes in a somatic cell hybrid system. *Chromosoma* **108**, 256–265 (1999).
43. McClintock, B. The behaviour of successive nuclear divisions of a chromosome broken at meiosis. *Proc. Natl Acad. Sci. USA* **25**, 405–416 (1939).
44. Agudo, M. *et al.* A dicentric chromosome of *Drosophila melanogaster* showing alternate centromere inactivation. *Chromosoma* **109**, 190–196 (2000).
45. Faulkner, N. E., Vig, B., Echeverri, C. J., Wordeman, L. & Vallee, R. B. Localization of motor-related proteins and associated complexes to active, but not inactive, centromeres. *Hum. Mol. Genet.* **7**, 671–677 (1998).
46. Sullivan, B. A. & Schwartz, S. Identification of centromeric antigens in dicentric Robertsonian translocations: CENP-C and CENP-E are necessary components of functional centromeres. *Hum. Mol. Genet.* **4**, 2189–2197 (1995).
47. Depinet, T. W. *et al.* Characterization of neo-centromeres in marker chromosomes lacking detectable alpha-satellite DNA. *Hum. Mol. Genet.* **6**, 1195–1204 (1997).
48. Du Sart, D. *et al.* A functional neo-centromere formed through activation of a latent human centromere and consisting of non-alpha-satellite DNA. *Nature Genet.* **16**, 144–153 (1997).
The first structural characterization of a human neocentromere.
49. Williams, B. C., Murphy, T. D., Goldberg, M. L. & Karpen, G. H. Neocentromere activity of structurally acentric mini-chromosomes in *Drosophila*. *Nature Genet.* **18**, 30–37 (1998).
50. Warburton, P. E. *et al.* Molecular cytogenetic analysis of eight inversion duplications of human chromosome 13q that each contain a neocentromere. *Am. J. Hum. Genet.* **66**, 1794–1806 (2000).
51. Saffery, R. *et al.* Human centromeres and neocentromeres show identical distribution patterns of >20 functionally important kinetochore-associated proteins. *Hum. Mol. Genet.* **9**, 175–185 (2000).
52. Lo, A. W. I. *et al.* A novel chromatin immunoprecipitation and array (CIA) analysis identifies a 460-kb CENP-A-binding neocentromere DNA. *Genome Res.* **11**, 448–457 (2001).
53. Lo, A. W. I. *et al.* A 330kb CENP-A binding domain and altered replication timing at a human neocentromere. *EMBO J.* **20**, 1–10 (2001).
Study that describes a neocentromeric sequence that associates with CENPA and a shift in replication timing of the CENPA-associated region compared with the endogenous locus from which it is derived.
54. Satinover, D. L., Vance, G. H., Van Dyke, D. L. & Schwartz, S. Cytogenetic analysis and construction of a BAC contig across a common neocentromeric region from 9p. *Chromosoma* (in the press).
55. Murphy, T. D. & Karpen, G. H. Centromeres take flight: α satellite and the quest for the human centromere. *Cell* **93**, 317–320 (1998).
56. Maggert, K. A. & Karpen, G. H. Acquisition and metastability of centromere identity and function: sequence analysis of a human neocentromere. *Genome Res.* **10**, 725–728 (2000).
57. Maggert, K. & Karpen, G. H. Neocentromere formation occurs by an activation mechanism that requires proximity to a functional centromere. *Genetics* **158** (in the press).
First study of the mechanism of neocentromere formation, which shows that neocentromere formation in *Drosophila* requires proximity to a functional centromere and probably occurs by spreading of centromeric proteins onto non-centromeric DNA.
58. Bridger, J. M. & Bickmore, W. A. Putting the genome on the map. *Trends Genet.* **14**, 403–409 (1998).
59. Brown, K. E. *et al.* Association of transcriptionally silent genes with Ikaros complexes at centromeric heterochromatin. *Cell* **91**, 845–854 (1997).
60. Brown, K. E., Baxter, J., Graf, D., Merkenschlager, M. & Fisher, A. G. Dynamic repositioning of genes in the nucleus of lymphocytes preparing for cell division. *Mol. Cell* **3**, 207–217 (1999).
61. Palmer, D. K., O'Day, K., Wener, M. H., Andrews, B. S. & Margolis, R. L. A 17-kD centromere protein (CENP-A) copurifies with nucleosome core particles and with histones. *J. Cell Biol.* **104**, 805–815 (1987).
62. Palmer, D. K., O'Day, K., Trong, H. L., Charbonneau, H. & Margolis, R. L. Purification of the centromere-specific protein CENP-A and demonstration that it is a distinctive histone. *Proc. Natl Acad. Sci. USA* **88**, 3734–3738 (1991).
This paper and reference 65 describe the ground-breaking biochemical characterization of CENPA and the demonstration that it is a centromere-specific core histone.
63. Sullivan, K. F., Hechenberger, M. & Masri, K. Human CENP-A contains a histone H3 related histone fold domain that is required for targeting to the centromere. *J. Cell Biol.* **127**, 581–592 (1994).
64. Meluh, P. B., Yang, P., Glowczewski, L., Koshland, D. & Smith, M. M. Cse4p is a component of the core centromere of *Saccharomyces cerevisiae*. *Cell* **94**, 607–613 (1998).
This paper describes chromatin immunoprecipitation studies that show that Cse4 is present in the *Saccharomyces cerevisiae* centromere.
65. Stoler, S., Keith, K. C., Curnick, K. E. & Fitzgerald-Hayes, M. A mutation in *CSE4*, an essential gene encoding a novel chromatin-associated protein in yeast, causes chromosome nondisjunction and cell cycle arrest at mitosis. *Genes Dev.* **9**, 573–586 (1995).
66. Takahashi, K., Chen, E. S. & Yanagida, M. Requirement of Mis6 centromere connector for localizing a CENP-A-like protein in fission yeast. *Science* **288**, 2215–2219 (2000).
The initial characterization of the *Schizosaccharomyces pombe* homologue of CENPA and the first demonstration of a protein required for CENPA localization, Mis6.
67. Buchwitz, B. J., Ahmad, K., Moore, L. L., Roth, M. B. & Henikoff, S. A histone-H3-like protein in *C. elegans*. *Nature* **401**, 547–548 (1999).
68. Henikoff, S., Ahmad, K., Platero, J. S. & Van Steensel, B. Heterochromatic deposition of centromeric histone H3-like proteins. *Proc. Natl Acad. Sci. USA* **97**, 716–721 (2000).
69. Yoda, K. *et al.* Human centromere protein A (CENP-A) can replace histone H3 in nucleosome reconstitution *in vitro*. *Proc. Natl Acad. Sci. USA* **97**, 7266–7271 (2000).
70. Shelby, R. D., Vafa, O. & Sullivan, K. F. Assembly of CENP-

© 2001 Macmillan Magazines Ltd

A into centromeric chromatin requires a cooperative array of nucleosomal DNA contact sites. *J. Cell Biol.* **136**, 501–513 (1997).
71. Blower, M. D. & Karpen, G. H. The role of *Drosophila* CENP-A/CID in kinetochore formation, cell-cycle progression and interactions with heterochromatin. *Nature Cell Biol.* (in the press).
Comprehensive study showing that *Drosophila* centromeres are composed of Cid-dependent multiple structural and functional domains and that Cid is required for many mitotic processes.
72. Howman, E. V. *et al.* Early disruption of centromeric chromatin organization in centromere protein A (Cenpa) null mice. *Proc. Natl Acad. Sci. USA* **97**, 1148–1153 (2000).
73. Moore, L. L. & Roth, M. B. HCP-4, a CENP-C-like protein in *Caenorhabditis elegans*, is required for resolution of sister centromeres. *J. Cell Biol.* **153**, 1199–1207 (2001).
74. Oegema, K., Desai, A., Rybina, S., Kirkham, M. & Hyman, A. Functional analysis of kinetochore assembly in *Caenorhabditis elegans*. *J. Cell Biol.* **153**, 1209–1226 (2001).
References 73 and 74 describe, for the first time, the organization and functional interactions between kinetochore proteins on holocentric chromosomes.
75. Nakaseko, Y., Goshima, G., Morishita, J. & Yanagida, M. M phase-specific kinetochore proteins in fission yeast. Microtubule-associating Dis1 and Mtc1 display rapid separation and segregation during anaphase. *Curr. Biol.* **11**, 537–549 (2001).
76. Goshima, G., Saitoh, S. & Yanagida, M. Proper metaphase spindle length is determined by centromere proteins Mis12 and Mis6 required for faithful chromosome segregation. *Genes Dev.* **13**, 1664–1677 (1999).
77. Ekwall, K. *et al.* The chromodomain protein Swi6: a key component at fission yeast centromeres. *Science* **269**, 1429–1431 (1995).
78. Rea, S. *et al.* Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* **406**, 593–599 (2000).
79. Lachner, M., O'Carroll, D., Rea, S., Mechtler, K. & Jenuwein, T. Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature* **410**, 116–120 (2001).
80. Bannister, A. J. *et al.* Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature* **410**, 120–124 (2001).
81. Nakayama, J., Rice, J. C., Strahl, B. D., Allis, C. D. & Grewal, S. I. Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly. *Science* **292**, 110–113 (2001).
82. Jenuwein, T. Re-SET-ting heterochromatin by histone methyltransferases. *Trends Cell Biol.* **11**, 266–273 (2001).
A recent review that describes newly defined histone modifications that are required to recruit proteins for establishment of heterochromatin structure. Also see references 78–81.
83. Adams, R. R., Maiato, H., Earnshaw, W. C. & Carmena, M. Essential roles of *Drosophila* inner centromere protein (incenp) and aurora b in histone h3 phosphorylation, metaphase chromosome alignment, kinetochore disjunction, and chromosome segregation. *J. Cell Biol.* **153**, 865–880 (2001).
84. Lopez, J. M., Karpen, G. H. & Orr-Weaver, T. L. Sister-chromatid cohesion via MEI-S332 and kinetochore assembly are separable functions of the *Drosophila* centromere. *Curr. Biol.* **10**, 997–1000 (2000).
85. Tanaka, T., Cosma, M. P., Wirth, K. & Nasmyth, K. Identification of cohesin association sites at centromeres and along chromosome arms. *Cell* **98**, 847–858 (1999).
86. Adams, C. R. & Kamakaka, R. T. Chromatin assembly: biochemical identities and genetic redundancy. *Curr. Opin. Genet. Dev.* **9**, 185–190 (1999).
87. Ridgway, P. & Almouzni, G. CAF-1 and the inheritance of chromatin states: at the crossroads of DNA replication and repair. *J. Cell Sci.* **113**, 2647–2658 (2000).
88. Csink, A. K. & Henikoff, S. Something from nothing: the evolution and utility of satellite repeats. *Trends Genet.* **14**, 200–204 (1998).
89. Ahmad, K. & Henikoff, S. Centromeres are specialized replication domains in heterochromatin. *J. Cell Biol.* **153**, 101–109 (2001).
90. McCarroll, R. M. & Fangman, W. L. Time of replication of yeast centromeres and telomeres. *Cell* **54**, 505–513 (1988).
91. Shelby, R. D., Monier, K. & Sullivan, K. F. Chromatin assembly at kinetochores is uncoupled from DNA replication. *J. Cell Biol.* **151**, 1113–1118 (2000).
Excellent study showing that human centromeres are replicated asynchronously in mid- to late S phase and that CENPA assembly into centromeric nucleosomes is not dependent on DNA replication.
92. Sullivan, B. A. & Karpen, G. H. Centromere identity in *Drosophila* is not determined *in vivo* by replication timing. *J. Cell Biol.* (in the press).
93. Sullivan, K. F. & Shelby, R. D. Using time-lapse confocal microscopy for analysis of centromere dynamics in human cells. *Methods Cell Biol.* **58**, 183–202 (1999).
94. Ortiz, J., Stemmann, O., Rank, S. & Lechner, J. A putative protein complex consisting of Ctf19, Mcm21, and Okp1 represents a missing link in the budding yeast kinetochore. *Genes Dev.* **13**, 1140–1155 (1999).
95. Glowczewski, L., Yang, P., Kalashnikova, T., Santisteban, M. S. & Smith, M. M. Histone–histone interactions and centromere function. *Mol. Cell. Biol.* **20**, 5700–5711 (2000).
96. Keith, K. C. & Fitzgerald-Hayes, M. CSE4 genetically interacts with the *Saccharomyces cerevisiae* centromere DNA elements CDE I and CDE II but not CDE III. Implications for the path of the centromere DNA around a Cse4p variant nucleosome. *Genetics* **156**, 973–981 (2000).
97. Chen, Y. *et al.* The N terminus of the centromere H3-like protein Cse4p performs an essential function distinct from that of the histone fold domain. *Mol. Cell. Biol.* **20**, 7037–7048 (2000).
98. Steiner, N. C. & Clarke, L. A novel epigenetic effect can alter centromere function in fission yeast. *Cell* **79**, 865–874 (1994).
The first clear demonstration of epigenetic effects on the function of the *Schizosaccharomyces pombe* centromere.
99. Zinkowski, R. P., Meyne, J. & Brinkley, B. R. The centromere–kinetochore complex: a repeat subunit model. *J. Cell Biol.* **113**, 1091–1110 (1991).
A classic paper on centromere structure that proposes that the centromere/kinetochore is composed of repeating subunits.
100. Seum, C., Delattre, M., Spierer, A. & Spierer, P. Ectopic HP1 promotes chromosome loops and variegated silencing in *Drosophila*. *EMBO J.* **20**, 812–818 (2001).
101. Dobie, K., Hari, K., Maggert, K. & Karpen, G. H. Centromere proteins and chromosome inheritance: a complex affair. *Curr. Opin. Genet. Dev.* **9**, 206–217 (1999).

Acknowledgements
We thank K. Sullivan for critical comments on the manuscript, and L. Moore and M. Roth (Fred Hutchison Cancer Research Center, Seattle) for the image in figure 2e. Our research on centromeres is supported by grants from the National Institutes of Health.

© 2001 Macmillan Magazines Ltd