The N Terminus of the Centromere H3-Like Protein Cse4p Performs an Essential Function Distinct from That of the Histone Fold Domain

YINHUAI CHEN, RICHARD E. BAKER, KEVIN C. KEITH, KENDRA HARRIS, SAM STOLER, AND MOLLY FITZGERALD-HAYES

Department of Biochemistry and Molecular Biology, Program in Molecular and Cellular Biology, University of Massachusetts at Amherst, Amherst, Massachusetts 01003, and Department of Molecular Genetics and Microbiology, University of Massachusetts Medical School, Worcester, Massachusetts 01655

Received 29 February 2000/Returned for modification 4 April 2000/Accepted 30 May 2000

Cse4p is an evolutionarily conserved histone H3-like protein that is thought to replace H3 in a specialized nucleosome at the yeast (Saccharomyces cerevisiae) centromere. All known yeast, worm, fly, and human centromere H3-like proteins have highly conserved C-terminal histone fold domains (HFD) but very different N termini. We have carried out a comprehensive and systematic mutagenesis of the Cse4p N terminus to analyze its function. Surprisingly, only a 33-amino-acid domain within the 130-amino-acid-long N terminus is required for Cse4p N-terminal function. The spacing of the essential N-terminal domain (END) relative to the HFD can be changed significantly without an apparent effect on Cse4p function. The END appears to be important for interactions between Cse4p and known kinetochore components, including the Ctf19p/Mcm21p-Opk1p complex. Genetic and biochemical evidence shows that Cse4p proteins interact with each other in vivo and that nonfunctional cse4 END and HFD mutant proteins can form functional mixed complexes. These results support different roles for the Cse4p N terminus and the HFD in centromere function and are consistent with the proposed Cse4p nucleosome model. The structure-function characteristics of the Cse4p N terminus are relevant to understanding how other H3-like proteins, such as the human homolog CENP-A, function in kinetochore assembly and chromosome segregation.

The centromere in budding yeast consists of conserved centromere DNA elements CDEI, CDEII, and CDEIII, which span about 125 bp of the centromere DNA (CEN) in each of the 16 Saccharomyces cerevisiae chromosomes (12, 13). Mutations altering any of these DNA elements cause defective chromosome segregation. The effects of CDE mutations range from relatively mild increases in mitotic chromosome loss (2- to 20-fold in CDEI mutants) to complete loss of centromere function (single point mutation in the central CCG triplet of CDEIII) (11, 14, 16, 21, 31). The centromere DNA is the site of assembly of the kinetochore, a protein complex that connects the chromosome to the mitotic spindle. To date, at least 10 different S. cerevisiae kinetochore proteins have been identified, some of which bind directly to CEN DNA. A homodimer of Cbf1p binds to CDEI (2, 4, 8, 9, 32) and induces a bend in the DNA that enhances centromere function (24, 36). The CBF3 complex contains four essential protein subunits (p110, p64, p58, and p23), which assemble specifically on CDEIII DNA (26, 27). The CBF3-CDEIII complex is required but by itself is not sufficient to confer centromere function and accurate chromosome segregation (11, 14). The A+T-rich CDEII DNA, which has an intrinsic bend even in the absence of bound protein, is necessary in addition to CDEIII to form a functional kinetochore microtubule binding site (25). Candidate CDEII DNA binding proteins include Mif2p, an essential centromere protein containing an “AT-hook” motif (34), and Cse4p (44). In addition, Cse4p, Mif2p, and Cbf1p make connections to CBF3-CDEIII through interactions with the Ctf19p-Mcm21p-Opk1p complex at the centromere (37). Ctf19p may also be involved in regulating kinetochore-microtubule interactions (19). It is proposed that kinetochore assembly initiates with formation of the CBF3-CDEIII complex, which then recruits additional kinetochore proteins, including Cse4p, to the centromere (33, 37).

Yeast centromere DNA is organized into a distinctive chromatin structure (6). Current models propose that two copies of the histone H3-like protein Cse4p replace both copies of H3 in a centromere-specific nucleosome (23, 35). The idea that Cse4p is incorporated into specialized centromeric nucleosomes is consistent with the histone-like biochemical properties of the Cse4p protein (44) and with the results of Cse4p mutational analysis (23). In addition, a mutation in either Cse4p or H4 (hhf1-20) disrupts chromatin structure at the centromere (35, 42). Overexpression of Cse4p, but not that of H3, suppresses the conditional growth, chromosome mis segregation, and defective centromeric chromatin phenotypes of hhf1-20 mutant cells, providing strong evidence that Cse4p interacts with H4 at the centromere (35). Genetic interactions between cse4 and other kinetochore genes have been detected; Baker et al. (3) isolated a mutant cse4 allele that is synthetic lethal with cep1Δ (cbf1Δ), and Ortiz et al. (37) showed a two-hybrid interaction between Cse4p and Ctf19p. Recently, Cse4p was implicated in the association of cohesin with the centromere (47), a further indication that Cse4p is an integral part of a higher-order chromatin structure at the budding yeast centromere.

Centromere H3-like proteins are evolutionarily conserved. In addition to Cse4p, fly (Cid), worm (HCP-3), and mamma-
liant (CENP-A) homologs have been identified (7, 17, 46). Homology between these proteins is limited to the H3-like histone fold domain (HFD) and ranges from 34 to 57% identical amino acids (17). Although originating from different organisms, Cse4p-, CENP-A-, and HCP-3-green fluorescent protein (GFP) fusions all preferentially localize to pericentric heterochromatin when expressed in Drosophila cells while Cse4p and HCP-3 preferentially localize to heterochromatic sites (including centromeres) in human cells (17). All of the centromere H3-like proteins have distinctly different N termini with little homology to known proteins (17, 44). Studies of CENP-A show that the HFD, but not the N terminus, is required for centromere targeting (46). A similar conclusion has been drawn for the Cse4p N terminus (37). However, the Cse4p N terminus has at least one essential function, since deletion of the first 50 amino acids is lethal to the cell (23). By analogy to the known folded structure of H3 in the nucleosome, the HFD of Cse4p would assemble into the octamer with the N terminus of the protein extending away from the core between the wrapped DNA helices (28). In this conformation, the Cse4p N terminus would be available to interact with proteins involved in critical centromere functions, including kinetochore protein recruitment and assembly, kinetochore-microtubule connections, and sister chromatid cohesion. The Cse4p N terminus may also be a target for posttranslational modifications, as observed for the N termini of the standard core histones (50). Such modifications may be required for the integrity of kinetochore structure or for other centromere functions, such as checkpoint signaling.

Here we report a comprehensive mutagenesis study of the N-terminal 130 amino acids of Cse4p. Our results show that the Cse4p N terminus contains an essential region of 33 amino acids which performs a function distinct from that of the HFD. Surprisingly, the position of the essential N-terminal domain relative to the HFD is highly flexible although the N terminus must be physically linked to the HFD for the protein to function. We present genetic and biochemical evidence that Cse4p proteins interact directly in vivo. In addition, we describe results supporting interactions between the N terminus of Cse4p and the newly identified centromere protein complex Mcm21p-Ctf19p-Opk1p. Our results reinforce the proposed Cse4p-nucleosome model and provide new information with implications for understanding the function of centromere H3-like proteins and how they are targeted to the centromere DNA.

### MATERIALS AND METHODS

**Yeast strains and genetic methods.** The yeast strains used in this study are shown in Table 1. YC190 and YC121 contain the integrated cse4-39 and cse4-52 alleles, respectively. Integration was accomplished by recloning the mutant alleles into integrating vector pRS304 or pRS306 (41) and then deleting the HFD segments by cleaving with BsuBI and Sall and religating. The resulting truncated cse4 constructs were linearized by NdeI and then transformed into a wild-type yeast strain. Plasmid integration resulted in a genetic duplication at the CSE4 locus consisting of a full-length mutant cse4 allele and a 3′-truncated (nonfunctional) copy of the endogenous wild-type allele. The expected genome changes were confirmed by diagnostic PCR.

**Yeast genetic manipulations were carried out using standard procedures (40).** The media and growth conditions used were as previously described (39). Transformations were performed by the lithium acetate procedure (40). The restrictive temperatures for temperature-sensitive (Ts) and cold-sensitive (Cs) mutations were 38 and 15°C, respectively.

**CSE4 mutants and epitope tagging.** All mutant cse4 alleles were constructed by sequential PCR. The template for both reactions was plasmid pCSE4 DNA consisting of wild-type CSE4 in the polylinker of pRS314 (44). First, a PCR was carried out using a primer (Integrated DNA Technology) containing the desired mutation and a primer complementary to polylinker sequences of the vector (T7 or T3 promoter). This PCR product was then used as a “megaprimer,” along with the T7 or T3 primer from the opposite end of the polylinker, for the second PCR. The product of the second PCR was digested with BamHI and Sall and reintroduced into pRS314. All mutations were confirmed by sequencing. For those cse4 alleles which showed growth phenotypes, the entire mutant cse4 gene was sequenced.

Two forms of hemagglutinin (HA)-tagged CSE4 were used. CSE4-HA, which has three HA repeats inserted at CSE4 codon 83 (engineered NolI site), was described by Stoler et al. (44). CSE4-HA, constructed for this study, has the triple HA epitope inserted immediately after the initiating methionine in the N terminus. In CSE4-HA, an EcoRI site was incorporated into the junction between the HA4 and CSE4 sequences. The HAN versions of mutant cse4 alleles were constructed by replacing the wild-type CSE4 sequence by the mutant cse4 sequence, obtained by PCR using an upstream primer carrying an EcoRI site and a downstream polylinker primer. GFP-tagged CSE4 (CSE4GFP) was obtained by replacing the triple-HA segment of CSE4HA with DNA encoding GFP-Bex1 (1). The GFP-Bex1 insertion was carried out by PCR using primers designed with in-frame Egel sites, destroying the NolI site, and introducing an Spel sites at the 5′ side. CSE4Spel was constructed by replacing the GFP segment of CSE4GFP with DNA encoding the ZZ domain of protein A, obtained by PCR amplification of plasmid pYES2 (Invitrogen), retaining the unique Spel and NolI sites. The cse4-HpSpel allele was made by replacing the BamHI-Spel fragment of CSE4Spel with the BamHI-Xbal fragment of cse4-39. The cse4GSpel and cse4-107GFP alleles were made by replacing the BamHI-Spel or NolI-Sal fragment of CSE4GFP with the BamHI-Xbal fragment of cse4-107. The two mutant alleles were sequenced to confirm that the epitope tag was inserted properly in frame and that no undesired mutations were present.

**Analyses of cse4 gene expression and function.** The ability of mutant cse4 alleles to complement a cse4 null allele was tested using a plasmid shuffle assay described previously (23). Plasmids carrying the cse4 mutant alleles were transformed into a cse4Δ:His3 strain (KC100) harboring wild-type CSE4 on a single-copy URA3 plasmid. Transformants were plated on 5-fluoroacetic acid (5-FOA) medium to select for loss of the URA3-CSE4 plasmid. 5-FOA-resistant cells, lacking the wild-type CSE4 plasmid, must rely on the mutant cse4 allele for viability. Three KC100 transformants for each cse4 allele were picked and sus-

### TABLE 1. Strains used in this study

<table>
<thead>
<tr>
<th>Strain</th>
<th>Genotype</th>
<th>Source or reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>YC190</td>
<td>MAT a de-2:101 his-3:11,15 leu-2:3 lys-2:801 trpA2::His3 ura-3:52 cse4::HIS3 (pC1I) (CSE4 in YCP50)</td>
<td>23</td>
</tr>
<tr>
<td>YC191</td>
<td>MAT a de-2:101 ade-2:24 his-3:11,15 leu-2:3 lys-2:801 trpA2::His3 ura-3:52 cse4::His3-TRP1</td>
<td>This study</td>
</tr>
<tr>
<td>YC121</td>
<td>MAT a de-2:101 his-3:11,15 leu-2:3 lysA::His3 ura-3:52 cse4::His3-TRP1</td>
<td>This study</td>
</tr>
<tr>
<td>YC320</td>
<td>MAT a de-2:101 ade-2:24 his-3:11,15 leu-2:3 cep1::His3</td>
<td>This study</td>
</tr>
<tr>
<td>YC222</td>
<td>MAT a de-2:101 his-3:11,15 leu-2:3 trpA2::His3 ura-3:52 mcm21::His3</td>
<td>This study</td>
</tr>
<tr>
<td>YC220</td>
<td>MAT a de-2:101 his-3:11,15 leu-2:3 trpA2::His3 ura-3:52 cse4::His3-TRP1 (p314)</td>
<td>This study</td>
</tr>
<tr>
<td>YMB120</td>
<td>MAT a de-2:101 his-3:200 leu-2::HAhis3::GFP ura-3:52 pspR42::His3 CFVII-URA3-SUP11</td>
<td>5</td>
</tr>
<tr>
<td>YPH1314</td>
<td>MAT a de-2:101 his-3:200 leu-2::HAhis3::GFP ura-3:52 cep1::His3</td>
<td>19</td>
</tr>
<tr>
<td>301-2A52</td>
<td>MAT a de-2:101 his-3:200 leu-2::HAhis3::GFP ura-3:52 okp1-5::TRP1 cyh2</td>
<td>38</td>
</tr>
<tr>
<td>YJL158</td>
<td>MAT a de-2:101 his-3:200 leu-2::HAhis3::GFP ura-3:52 okp1-5::TRP1</td>
<td>37</td>
</tr>
<tr>
<td>PKY031</td>
<td>MAT a de-2:101 his-3:200 leu-2::HAhis3::GFP ura-3:52 okp1-5::TRP1</td>
<td>22</td>
</tr>
<tr>
<td>PKY034</td>
<td>MAT a de-2:101 his-3:200 leu-2::HAhis3::GFP ura-3:52 okp1-5::TRP1</td>
<td>22</td>
</tr>
</tbody>
</table>

7038 CHEN ET AL. MOL. CELL. BIOL.
pended in 1.5 ml of water. The optical density at 660 nm (OD660) of the cell suspension was adjusted to 0.5. Six-microliter volumes of this suspension and of a 1:10-dilution were plated on medium lacking tryptophan and on 5-FOA medium, and the plates were photographed after 2 to 3 days at 30°C. These cse4 mutant alleles that could complement cse4-1/HIS3 were tested further for conditional growth phenotypes at 15 and 38°C. Tests for interallelic complementation were carried out using KC100 carrying the respective mutant alleles on separate 2.3-kb EcoRI fragments. These plasmids were transformed into strain PJ69-4A, which carries a Ctf19p-Gal4 DNA binding domain (BD) fusion protein (Ctf19p-BD). In both cases, the inserts were obtained by PCR, incorporating BamHI and Sall sites into the primers for cloning into the polylinker region of the vectors. Cse4p-AP and Ctf19p-BD were then transformed into strain JD16-4A, which carries a GALI-HIS3 reporter gene. Three transformants were picked and suspended in 1.5 ml of 5-FOA medium. The OD660 of the cell solution was determined and adjusted to 1.0, and then a series of fivefold dilutions was made. Six microliters of each dilution were plated on complete minimal medium and on minimal medium lacking histidine and supplemented with 0.5 mM 3-amino-triazole (to suppress growth due to basal HIS3 transcription). The plates were incubated at 30°C and photographed after 5 days.

RESULTS

Alanine scanning mutagenesis identifies an important region in the Cse4p N terminus. The N terminus of Cse4p contains several clusters of charged amino acids. To investigate the functional significance of these charged amino acids, groups of charged residues were changed to alanines and the resulting alleles were tested for Cse4p function in vivo. Haploid cells with a disrupted CSE4 gene (cse4Δ) null and carrying wild-type CSE4 on a URA3 plasmid to maintain viability were transformed with low-copy-number (CEN) plasmids carrying different mutated cse4 genes. The transformants were plated on agar medium containing 5-FOA to select for loss of the CSE4-URA3 plasmid and to obtain cells that were completely dependent on the function of the mutant cse4-encoded protein. In these 5-FOA cells, the effects of the cse4 mutations on cell growth and chromosome segregation were assessed without interference from the wild-type protein. Cells that cannot lose the URA3/CSE4 plasmid, and therefore cannot grow on 5-FOA medium, carry plasmids with lethal cse4 mutations. Nineteen different alanine scanning mutations (cse4-21 through cse4-39, Fig. 1A) were tested for the ability to rescue cse4 null cells in the plasmid shuffle assay. Most of the alanine substitution mutants exhibited wild-type phenotypes (data not shown). However, cse4-23 (R44A, R46A, and K49A) caused a leaky Ts growth phenotype and the cse4-22 mutant (D36A and R37A) showed a small increase in chromosome missegregation (data not shown). To determine whether the cse4-22 and cse4-23 mutants could be used to generate a new cse4-39 alanine substitution mutant (D36A, R37A, R44A, R46A, and K49A) which grew at 30°C but exhibited both Ts and Cs growth phenotypes (see Fig. 7A), these results suggested that a region of the N terminus encompassing cse4-39 is important for the function of Cse4p.

Deletion mutagenesis further delineates a 33-amino-acid region essential and sufficient for wild-type Cse4p function. As a complement to the alanine scanning mutagenesis, we also performed a systematic deletion analysis of the Cse4p N terminus. The first set of nested deletions retained the Cse4p initiating methionine but removed successively larger segments of the N terminus (cse4-Δ13 to cse4-Δ129, Fig. 1B). Results of plasmid shuffle assays showed that, in agreement with previous results (23), removing residues 2 to 27 from the N terminus (cse4-Δ27) was without apparent effect, but removing residues 2 to 55 (cse4-Δ55) was lethal (Fig. 1B and 2A).

To define precisely the essential domain between residues 27 and 55, we made the following series of small deletions spanning the region: cse4-548 (Δ28-35), cse4-541 (Δ36-40), cse4-542 (Δ41-45), cse4-543 (Δ46-50), and cse4-544 (Δ51-55) (Fig. 1B). Two of these cse4 N-terminal mutants, those with the cse4-541 and cse4-542 deletions, were viable in the plasmid shuffle assay but formed very small colonies on 5-FOA medium that were distinctly different from the normal-size colonies produced by the deletion cse4-548 and cse4-543 mutants (Fig. 2A, Upon...
proximal \textit{cse4} deletion mutants were designed to have common endpoints at residue 130 (Fig. 1B). Throughout the rest of this report, we will consider residues 130 to 135 to be part of the HFD. All of these \textit{cse4} N-terminal deletion mutants grew like the wild type, revealing that the 70 amino acids proximal to the Cse4p HFD are not required for Cse4p function (data not shown).

Having found that residues 2 to 27 and 51 to 129 in the Cse4p N terminus are dispensable, we next identified the minimal sequences sufficient to provide the essential N-terminal function. Cse4-559p contained amino acids 28 to 60 fused directly to the HFD (Fig. 1B) and was functional at all of the temperatures tested (Fig. 2B). Additional deletion of amino acids 51 to 60 (cse4-560) or 28 to 35 (cse4-561) impaired but did not abolish function. The cse4-560 and cse4-561 mutants, while viable, exhibited conditional growth phenotypes and consistently formed small colonies on 5-FOA medium at the permissive temperature (Fig. 2A and B). The region of overlap between cse4-560 and cse4-561 did not confer the essential N terminal function (cse4-562, Fig. 2A). We concluded that amino acids 28 to 60 are sufficient to provide the essential function of the Cse4p N terminus, and we have named this region the essential N-terminal domain (END).

Expression of \textit{cse4} N terminus mutant proteins. Expression of the mutant \textit{cse4} proteins was assayed by immunoblotting. With one possible exception, \textit{cse4} mutant phenotypes could not be attributed simply to altered protein expression levels. As shown in Fig. 2C, alleles cse4-541, -542, -560, and -561, all of which cause detectable growth and/or segregation phenotypes, produce protein at levels comparable to that of the wild-type allele. The exceptional case is the lethal allele cse4-562, which did not express a detectable level of protein. The presence of wild-type Cse4p, necessary where the mutant Cse4p was non-functional, resulted in significantly reduced steady-state levels of the mutant proteins (asterisks, Fig. 2C). Interestingly, the location of the epitope tag affected the migration of the protein in the SDS-gel. When the triple-HA tag was inserted at residue 83, the resulting Cse4pHA protein migrated with an apparent molecular weight of 43,000, in agreement with previous results (23, 44). When the same tag was fused to the extreme N terminus, epitope-tagged Cse4p migrated with an apparent molecular weight of 37,000.

Mutations in the N terminus of Cse4p cause increased chromosome missegregation. Cse4p plays a critical role in centromere function during mitosis, and mutations in the HFD cause increased chromosome loss rates (23, 35, 44). To determine if chromosome segregation is impaired in strains carrying Cse4p N-terminal mutations, we screened the collection of Cse4p N-terminal alanine scanning and deletion mutants for increased chromosome missegregation levels using a qualitative colony color sectoring assay (18). The frequencies of sectored colonies made by diploid \textit{cse4} null cells expressing different \textit{cse4} N-terminal mutant proteins were visually compared to those of cells expressing wild-type \textit{CSE4}. We assigned tentative chromosome loss phenotypes to 3 of the 19 alanine scanning mutants (cse4-39, cse4-22, and cse4-23), two partial \textit{END} deletions (cse4-541 and cse4-542), and two \textit{END} boundary mutants (cse4-560 and cse4-561). Chromosome loss rates were quantified for three of these mutants (cse4-39, cse4-541, and cse4-542), all of which exhibited increases in chromosome loss rates of 12- to 14-fold (Table 2). In contrast, an only 1.7-fold increase in chromosome loss was found for the Cse4p-559p protein, which contains the minimal \textit{END} fused to the HFD.

Is the Cse4p N terminus posttranslationally modified? The N-terminal tails of standard core histones are posttranslationally modified, and some modifications are critical for histone

restreaking, the \textit{cse4-541} and \textit{cse4-542} mutant cells grew slower than control strains at 30°C and were Ts and Cs (Fig. 2B). One deletion mutant, the \textit{cse4-548} mutant, grew well at 30°C but slowly at 38 and 15°C. These results were consistent with those of the alanine scanning mutagenesis and conclusively identified the region located between residues 28 and 46 in the Cse4p N terminus, epitope-tagged Cse4p migrated with an apparent molecular weight of 43,000, in agreement with previous results (23, 44). When the same tag was fused to the extreme N terminus, epitope-tagged Cse4p migrated with an apparent molecular weight of 37,000.
function in transcriptional regulation, cell division, and chromatin structure (48, 50). The Cse4HA protein migrates on SDS-PAGE with an apparent molecular mass (43 kDa) much larger than that predicted for the HA-tagged protein (31 kDa) (23, 44), suggesting that Cse4p is modified in vivo. The Cse4p N-terminal amino acid sequence contains several potential posttranslational modification sites. As most of the Cse4p N terminus is dispensable, it is unlikely that modification of any of these nonessential residues is functionally critical; however, the END contains two serines, two threonines, and one lysine, potential sites of phosphorylation (serine or threonine) or acetylation (lysine). To test this possibility, we changed these residues to alanines in three different \(cse4\) alleles (\(cse4-559a\) to \(-559c\), Fig. 3). In no case did the mutations affect the growth of the cells or the expression levels or migration of the proteins on SDS-gels (Fig. 3). We concluded that neither phosphorylation nor acetylation of residues within the END is essential for Cse4p function, although other possible modifications are not excluded. The modification mutants were also screened for chromosome segregation defects using the visual colony sectoring assay. The \(cse4-559a\) and \(cse4-559b\) mutants exhibited a detectable increase in chromosome loss (approximately two- to threefold) compared to the parent (\(cse4-559\)), while the sectoring phenotype of \(cse4-559c\) was identical to that of \(cse4-559\). The relatively mild chromosome segregation defects observed

<table>
<thead>
<tr>
<th>CSE4 allele</th>
<th>No. of chromosome loss events/cell division (10^3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CSE4</td>
<td>0.17 (1)</td>
</tr>
<tr>
<td>(cse4-39)</td>
<td>2.42 (14)</td>
</tr>
<tr>
<td>(cse4-541)</td>
<td>2.28 (13)</td>
</tr>
<tr>
<td>(cse4-542)</td>
<td>2.00 (12)</td>
</tr>
<tr>
<td>(cse4-559)</td>
<td>0.29 (1.7)</td>
</tr>
</tbody>
</table>

* The values in parentheses are fold increases relative to wild-type \(CSE4\).
could be due either to the lack of modification or to alteration of the ENd structure caused by the mutations.

Evidence for a functional multimeric complex containing at least two molecules of Cse4p. Current genetic evidence is consistent with the idea that Cse4p binds to H4 and forms H4-Cse4p dimers, which assemble into H4-Cse4p-Cse4p-H4 tetramers and ultimately into (H4-Cse4p)2(H2A-H2B)2 octamers (35). In the case of a cell expressing two mutant cse4 alleles, this model predicts that two types of Cse4p octamers would form: homotypic complexes containing two molecules of the same mutant cse4 protein and heterotypic complexes containing one each of the two different mutant cse4 proteins. If the different cse4 mutations affect different biochemical activities of the protein, wild-type function might be partially or fully recovered through the formation of the mixed (i.e., heterotypic) complex (interallelic complementation).

To test for interallelic complementation between cse4 N terminus and HFD mutations, we coexpressed the Ts alleles cse4-39 and cse4-107, which carry mutations in the ENd (Fig. 1A) and HFD (glutamine 219 changed to aspartate), respectively. Control cse4 null cells carrying either two cse4-39 plasmids or two cse4-107 plasmids failed to grow at 38°C. In contrast, the same yeast strain carrying both the cse4-39 plasmid and the cse4-107 plasmid could grow at 38°C, although not as well as cells expressing wild-type Cse4p (Fig. 4A). Apparently, the essential N-terminal function was provided by the wild-type N terminus of the cse4-107-encoded protein, while the wild-type HFD of the cse4-39-encoded protein conferred the essential HFD function, possibly through heterodimer formation with the defective HFD of Cse4-107p. Interallelic complementation was not observed when a frameshift mutation terminating translation between the N terminus and the HFD was introduced into the cse4-107 gene (Fig. 4A). Thus, the wild-type N terminus of the cse4-107 protein could not rescue the cse4-39 defect when separated from its HFD, consistent with the prediction that the HFD mediates the physical interactions between Cse4p proteins.

Expression of the Cse4p HFD alone was sufficient to rescue...
cse4-107 temperature sensitivity. Cells coexpressing one of the lethal cse4 N-terminal deletion mutants (cse4∆55, cse4∆80, or cse4∆129) and the Ts cse4-107 allele were viable and grew at both 30 and 38°C (Fig. 4A). The cse4-107-encoded protein apparently provided the END function, while the mutant proteins with N-terminal deletions complemented the HFD function. This implies that only one functional Cse4p N terminus is required in each Cse4p nucleosome. To confirm that the N-terminally truncated cse4-encoded proteins were actually present in Cse4p complexes at functional centromeres, the cellular localization of Cse4D55p tagged with GFP (Cse4D55GFP) was determined in cells dependent on both Cse4-107p and Cse4D55GFP for viability. In agreement with previous indirect immunofluorescence analysis of Cse4HA (35), we found that wild-type Cse4GFP was localized as a single subnuclear focus in unbudded and small-budded cells and as a short bar or two distinct foci in preanaphase cells (Fig. 5, left panels). These localization patterns correspond to the clustered centromeres visualized in yeast nuclei when the centromeres are labeled by fluorescence in situ hybridization (15). In cells expressing Cse4-107GFP and grown at the permissive temperature, the mutant protein localizes like the wild-type protein (Fig. 5, middle panels) although the strain exhibits a mild growth defect and a high proportion of large-budded cells. When cse4-107 cells were rescued by the cse4D55GFP allele at 38°C, localization of Cse4D55GFP was very similar to that of Cse4-107GFP at 30°C (Fig. 5, right panels), supporting the idea that Cse4D55p is incorporated into complexes (presumably nucleosomes) at centromeres. Localization of Cse4D55GFP at the DNA level was further analyzed by CHIP. In cse4-107 mutant cells coexpressing cse4D55GFP, the GFP-tagged protein showed specific cross-linking to CEN DNA (Fig. 6) at both growth temperatures. Thus, by both biochemical and morphological criteria, Cse4D55GFP is properly localized to centromeres under these conditions.

Biochemical evidence supporting our interpretation that the interallelic complementation results from the formation of mixed complexes of Cse4p mutant proteins was obtained from coprecipitation assays using Cse4p proteins tagged with the divalent staphylococcal protein A analogue ZZ (SpA) (43). The Cse4SpA protein is fully functional in cse4 null cells and is efficiently precipitated from crude cell extracts by IgG-Sepharose beads. The cse4-39SpA allele was coexpressed with the HA-tagged cse4-107 allele in cse4 null cells. As previously observed, neither single mutant grew at 38°C but when expressed together, Cse4-39SpA and Cse4-107HA permitted growth at

FIG. 5. Localization of GFP-tagged Cse4p proteins. Cells were fixed and prepared for microscopy as described in Materials and Methods. The top panels show GFP fluorescence, the middle panels show DAPI staining, and the bottom panels show the overlaid GFP and DAPI images. (Left) Yeast cells (KC100) carrying only wild-type Cse4GFP were grown at 30°C and shifted to 37°C for 5 h (three to four doublings) before fixation. (Middle) KC100 cells expressing only Cse4-107GFP grown at the permissive temperature (30°C). (Right) KC100 cells carrying cse4-107HA and cse4D55GFP on separate URA3 and TRP1 vectors, respectively, were streaked and grown at 38°C, the restrictive temperature for both mutations. A single colony was picked, inoculated into prewarmed medium, and grown for an additional 15 h (three to four doublings) at 38°C before fixation and microscopy.
38°C (data not shown). As shown in Fig. 4B, Cse4-107HA and Cse4-39SpA coprecipitated from extracts prepared from cells grown at both 30 and 38°C, demonstrating that the two mutant cse4 proteins formed a mixed complex in vivo. The fact that the heteromeric complexes were also observed at 30°C, a temperature at which both proteins are functional, suggests that complex formation is not an artifactual peculiarity of one or the other inactive mutant protein at the restrictive temperature.

Genetic interactions between cse4 N-terminal alleles and genes encoding subunits of the Mcm21p-Ctf19p-Okp1p complex. To identify proteins that might potentially interact with the N terminus of Cse4p, we looked for dosage suppressors of the Ts phenotype of cse4-23 cells (see Materials and Methods). Among the candidate genes isolated in the primary screen was MCM21, a nonessential gene originally identified in a screen for minichromosome maintenance mutants (29). MCM21 in high copy number also suppressed the Ts and Cs phenotypes of cse4-39. MCM21 suppression was selective for the END alleles; high-copy-number MCM21 did not suppress the Ts phenotype of the cse4-107 (glutamine 219 to aspartate) or cse4-1 (alanine 221 to threonine) HFD mutant (data not shown). A second genetic interaction, synthetic lethality, was also observed between mcm21 and cse4-39 (Fig. 7B; Table 3).

Since MCM21 encodes a subunit of the Mcm21p-Ctf19p-Okp1p centromere complex (37), we tested for genetic interactions between cse4 N-terminal mutations and other components of the complex, as well as other known centromere proteins. We observed synthetic lethal interactions of cse4-39 with ctf19 Δ and mcm21 Δ and of cse4-542 with okp1-5, cep1 Δ, and mcm22 Δ (Table 3). MCM22 is a nonessential gene whose function is not known that was isolated in the same genetic suppressor screen as MCM21 (29, 38). Synthetic lethality was
not observed between cse4-39 and cac1 or cac2, genes encoding chromatin assembly factors (22), or *spq4*, which is involved in the structure of centromere chromatin (5). In addition to the observed synthetic lethality between *cse4* END alleles and *mcm21Δ*, *ctf19Δ*, *okp1-1*, and *cep1Δ*, subsequent plasmid shuffle assays (similar to those of Fig. 7B) revealed that *cse4-39* was synthetically lethal with *nsc1Δ*<sup>sd</sup>, *ctf13-30*, and *cep3<sup>sd</sup>*<sup>d</sup>, all genes encoding subunits of the CBF3 complex (data not shown). The *ctf13-30* and *cep3<sup>sd</sup>*<sup>d</sup> interactions were END specific, as these alleles were not synthetic lethal with *cse4-107* and other HFD alleles (data not shown). Surprisingly, in contrast to high-copy-number suppression, synthetic lethal interactions between *mcm21Δ* and *cse4* were not END specific; some HFD mutant alleles were also synthetic lethal with *mcm21Δ* (as tested by plasmid shuffle assay; data not shown).

Ortiz et al. (37) demonstrated an interaction between Cse4p and Ctf19p by yeast two-hybrid assay. We used a similar two-hybrid assay to test the dependence of the Cse4p-Ctf19p interaction on the Cse4p N terminus. Gal4 AD fusions of wild-type and mutant Cse4p proteins were coexpressed with Ctf19p fused to the Gal4 BD in yeast cells containing a HIS3 reporter. Cells expressing the full-length AD-Cse4p and BD-Ctf19p fusion proteins were His<sup>+</sup> and grew well on indicator medium containing 3-amino-triazole, confirming Cse4p-Ctf19p interaction. Cells expressing the Ad-Cse4p-559p fusion protein containing the minimal END attached to the HFD and BD-Ctf19p grew as well as cells expressing the wild-type AD-Cse4p and BD-Ctf19p proteins. However, when the END was deleted from the AD-Cse4p fusion (AD-Cse4Δ436-559p), cell growth was barely detectable (Fig. 7C). We observed no growth of cells coexpressing either the wild-type CSE4 or cse4-559 AD fusion with an empty BD vector. We concluded that Cse4p and Ctf19p interact in vivo and that the END of the N terminus is an important determinant of this protein-protein interaction.

**DISCUSSION**

Cse4p is novel among known histone H3-like proteins in having a unique 135-amino-acid-long N terminus extending from the conserved HFD homology region. Unlike the N terminus of yeast H3, which can be deleted without loss of cell viability (30), the Cse4p N terminus is essential (23). Here we have characterized the Cse4p N terminus by systematic mutagenesis, revealing important new information about its function. First, the Cse4p N terminus contains a spatially flexible 33-amino-acid domain, the END, that is essential for Cse4p function. Second, mutations affecting the END and HFD define distinct functions of the protein. Third, the END appears to be involved in the interaction between Cse4p and the Mcm21p-Ctf19p-Okp1p centromere complex. These results are consistent with the current hypothesis that Cse4p replaces H3 in a specialized nucleosome and mediates an essential interaction(s) with other components of centromere chromatin (35).

**An essential domain in the Cse4p N terminus.** The Cse4p N terminus contains a large proportion of charged amino acids, especially the region between residues 54 and 132, where 48% of the amino acids are charged. The N terminus also contains a putative bipartite nuclear localization signal between residues 115 and 132 (44), and in addition to many possible posttranslational modification sites and a high concentration of serines within the first 22 amino acids. The combined mutagenesis results show that none of these features is critical; in fact, most of the Cse4p N terminus is dispensable. Only the region between amino acids 28 and 60, the domain we refer to as the END, is essential. Another region of possible functional importance includes residues 130 to 135, which, by analogy to the structure of the conventional nucleosome (28), would exit the core and pass between the two DNA helices wrapped around the histone octamer. Because of this, we avoided altering residues 130 to 135 in the mutagenesis study, although we now know that deletion of residues 130 to 135 does not cause growth or chromosome loss phenotypes in cells with wild-type centromeres (data not shown). This suggests that the positioning of the DNA gyres on the surface of the putative Cse4p-containing octamer is sufficiently flexible to accommodate totally different polypeptide chains passing between them. Interestingly, although the wild-type Cse4p N terminus is much longer than those of H3 and CENP-A, the Cse4p mutant protein Cse4p-559p, which consists of the 33-amino-acid END fused directly to the HFD (including residues 130 to 135), confers wild-type function and is very similar in length to CENP-A and H3.

In mammals, H3 phosphorylation is required for the initiation of chromosome condensation (48) but phosphorylated H3 is excluded from the chromatin subjacent to the inner kinetochore plate (49). The centromere chromatin is also under-acetylated in mammals (49). The nonphosphorylated, hypoacetylated chromatin zone corresponds to that region of the centromere associated with CENP-A. CENP-A lacks the phosphorylation epitope (serine 10) found in the N terminus of H3, as well as the acetylation sites associated with “transcriptional”
methionine located between the END and the HFD (residue 93). A protein initiated at this methionine would have almost the same length as H3. It is therefore possible that Cse4p is the evolutionary result of a gene fusion event through which H3, or a variant of H3, acquired the END function. That the END and the HFD behave to some degree independently, as suggested from the seeming lack of spacing constraints, is consistent with this idea.

**Cse4p END function is distinct from that of the HFD.** Current genetic data agree with the model in which Cse4p replaces H3 in a centromere-specific nucleosome (35). This model also predicts that the Cse4p HFD mediates Cse4p-Cse4p and Cse4p-H4 interactions in the histone octamer, which is supported by results of systematic HFD mutagenesis (23). Additional evidence that Cse4p proteins interact in vivo and probably form nucleosome-like structures is supplied by the interallelic complementation that we observed between END and HFD cse4 alleles. These results argue that functional heterotypic octamers can be assembled in which one of two Cse4p molecules is defective in the HFD and the other is defective in the END. Biochemical support for this interpretation is provided by the fact that coexpressed mutant cse4 proteins were copurified from cell extracts. The nucleosome model further predicts that the END is incorporated into centric chromatin by virtue of the attached HFD. Consistent with this model, introducing a frameshift mutation into Cse4p between the END and HFD abolishes interallelic complementation. The incorporation of Cse4p molecules into CEN chromatin in complemented cells is probably not random. If it were, half of the nucleosomes assembled would be homotypic (nonfunctional) and the probability of all 16 centromeres incorporating a functional, heterotypic octamer would be low, with severe effects on growth. Since this is not observed, functional heterotypic Cse4p nucleosomes are probably assembled preferentially. Alternatively, the centromere region may have multiple Cse4p nucleosomes, increasing the probability of at least one functional END at each centromere.

Besides providing genetic support for the nucleosome model, interallelic complementation between cse4 END and HFD alleles suggests that the Cse4p END and HFD have distinct functions: the HFD mediates nucleosome assembly, while the END is required for a different Cse4p function, such as nuclear localization, centromere targeting, or interactions with other kinetochore components. The END function could be required at a different time from that of the HFD, either before or after Cse4p dimer formation. Interestingly, cse4 alleles entirely lacking the N-terminal sequences can complement the cse4-107 HFD mutation, indicating that only one functional END per nucleosome is sufficient for Cse4p function at the centromere. The conclusion that the Cse4p END and HFD have different functions is also supported by genetic suppression and synthetic lethal analyses. The screen for high-copy-number suppressors of the cse4-23 END mutation yielded MCM21. In contrast, similar screens for high-copy-number suppressors of HFD mutants (e.g., cse4-1) consistently yielded SCM3 but not MCM21 (C. D. DeFalco et al., unpublished data). SCM3 does not suppress END mutants such as cse4-23 and cse4-39, while MCM21 does not suppress HFD mutants, including cse4-1 and cse4-107. Allele specificity is also observed with synthetic lethal gene interactions. The END allele cse4-39 is synthetically lethal with all of the centromere protein mutants tested, while cse4-107 and other HFD alleles are not synthetic lethal with any of the centromere protein mutants tested, while cse4-39 and other HFD alleles are not synthetic lethal with any.

**A network of kinetochore protein interactions involving the Cse4p END.** Recently, Ortiz et al. (37) reported that the Ctf19p-Mcm21p-Okp1p protein complex mediates protein-protein in-
teractions at the yeast centromere. They showed by both two-
hybrid analysis and communoprecipitation that Ctf19p, Mcm21p, and Okp1p interact with each other, with one or more subunits of CBF3, with Mit2p, and with Cse4p. This network of protein-
protein interactions potentially accounts for the localization of all known kinetochore components, including Cse4p, to the centromere. In addition, they demonstrated by CHIP that CDEIII is necessary and sufficient to localize the Ctf19p com-
plex and Cse4p to the centromere.

Our results suggest that the Cse4p END is involved in the interaction(s) between Cse4p and the Ctf19p-Mcm21p-Okp1p complex. MCM21 is a dosis suppressor of END mutations, and a mutation in any of the three components of the Ctf19p-Mcm21p-Okp1p complex is synthetic lethal with a end mutation. Our two-hybrid results confirm the Ctf19p-Cse4p interaction (37) and further show that it is abolished when the END is deleted from the Cse4p-AD fusion protein, implying that the Cse4p END is required for this interaction. Protein-
protein interactions between the END of Cse4p and other centromere proteins may be important for recruiting additional kinetochore proteins to the centromere or for local-
ization of Cse4p itself. Although the Cse4p HFD alone (Cse4Δ55GFP) can be properly targeted to the centromere, as demonstrated by CHIP analysis, this should not be interpreted to mean that the END is not required for centromere localization. Indeed, results from the interallelic complementation and coprecipitation experiments suggest that Cse4Δ55GFP may be targeted by heterodimerization with another Cse4p molecule bearing an intact END (e.g., Cse4-107p or wild-type Cse4p).

An END in other H3-like centromere proteins? The HFDs of centric histone variants are conserved from yeast to humans despite great divergence in the structures of the respective centromeres. While the HFD of CENP-A might be sufficient to target CENP-A to mammalian centromeres (46), a possible function for the CENP-A N terminus cannot be ruled out. N-terminally truncated CENP-A might lack an essential func-
tion and still be able to localize properly (e.g., Cse4Δ55GFP). A second possibility is that the CENP-A “END” became sepa-
parated from the HFD during evolution and the END function is now supplied to human centromeres in trans by another, as yet unidentified, protein. The Cse4p END does not function in trans when expressed as a separate polypeptide (Fig. 8), pos-
sibly because the N terminus lacks targeting information or because the N-terminal peptide by itself is unstable. It would be interesting to determine if the Cse4p END function could be supplied in trans by fusing the Cse4p N terminus to another kinetochore protein, such as Cbf1p or the CBF3 components. Finally, it is also possible that the function of the Cse4p N terminus is unique to the “point” centromeres of budding yeast and that no comparable activity is necessary for the function of vertebrate kinetochores. Regardless of a possible cis or trans CENP-A END function, the similarities and differences be-
 tween the centromere-specific CENP-A and Cse4p proteins continue to reveal features in common between the two very diverse kinetochore structures.

ACKNOWLEDGMENTS

We thank Jinyun Chen, Tom King, Ben Liu, and Wayne Decatur for technical advice and other members of the M.F.-H. laboratory for comments and suggestions. We are grateful to P. Hieter, J. Lechner, P. Sinha, M. A. Basrai, and F. D. Kaufman for yeast strains and plasmids and to Kellie Kosewski for constructing the GFP-tagged CSE4 allele.

This work was supported by grants to M.F.-H. from the National Institutes of Health (GM54776) and to R.E.B. from the National Science Foundation (MCB-9406050).

REFERENCES

6. Bloom, K. S., and J. Carbon. 1982. Yeast centromere DNA is in a unique and highly ordered structure in chromosomes and small circular minichro-
10. Carlson, M., and D. Botstein. 1982. Two differentially regulated mRNAs with different 5′ ends encode secreted and intracellular forms of yeast in-
235–244.
17. Henikoff, S., K. Ahmad, J. S. Platero, and B. van Steensel. 2000. Hetero-
23. Keith, K. C., R. E. Baker, Y. Chen, K. Harris, S. Stoler, and M. Fitzgerald-
15–28.