Structural and Functional Dissection of Mif2p, a Conserved DNA-binding Kinetochoore Protein

Citation

Published Version
http://dx.doi.org/10.1091/mbc.E08-03-0297

Permanent link
http://nrs.harvard.edu/urn-3:HUL.InstRepos:3710321

Terms of Use
This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA

Share Your Story
The Harvard community has made this article openly available. Please share how this access benefits you. Submit a story.

Accessibility
Structural and Functional Dissection of Mif2p, a Conserved DNA-binding Kinetochore Protein


*Jack and Eileen Connors Structural Biology Laboratory, Harvard Medical School, and ‡Howard Hughes Medical Institute, Boston, MA 02115; and §Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139

Submitted March 19, 2008; Revised August 1, 2008; Accepted August 6, 2008

Mif2p is the budding-yeast orthologue of the mammalian centromere-binding protein CENP-C. We have mapped domains of Saccharomyces cerevisiae Mif2p and studied the phenotypic consequences of their deletion. Using chromatin immunoprecipitation (ChIP) and electrophoretic mobility shift assays, we have further shown that Mif2p binds in the CDEIII region of the budding-yeast centromere, probably in close spatial association with Ndc10p. Moreover, ChIP experiments show that Mif2p recruits to yeast kinetochores a substantial subset of inner and outer kinetochore proteins, but not the Ndc80 or Spc105 complexes. We have determined the crystal structure of the C-terminal, dimerization domain of Mif2p. It has a “cupin” fold, extremely similar both in polypeptide chain conformation and in dimer geometry to the domains of Mif2p are required for recruitment of many additional kinetochore components (Oegema et al., 2001; Cheeseman et al., 2004; Liu et al., 2006; Kwon et al., 2007). Mif2p is therefore both an integral part of the inner kinetochore and a critical subunit for assembling the entire kinetochore superstructure. We report here cellular, biochemical, and structural characterization of the domain organization of Mif2p. A region (residues 256–549) that includes the CENP-C signature motif and a likely DNA-binding domain is important for normal cell growth. Deletion N-terminal to this essential segment or deletion of a conserved C-terminal domain results in slow-growing, temperature-sensitive cells. High-resolution chromatin immunoprecipitation (ChIP) ex-
Determination of the Mif2p polypeptide chain create a dimer-containing nucleosomal assembly. Residues 438–549 at the ends show that this region is occupied instead by a Cse4p-2007; Mizuguchi et al.

Strains and Constructs

MATERIALS AND METHODS

Strains and Constructs

See Supplemental Table for details.

Analysis of Mif2p Truncation Phenotypes

An S288C haploid yeast strain was generated in which a chromosomal deletion of the endogenous MIF2 gene was rescued by MIF2 expressed from a URA3-CEN plasmid (to create pMIF2:URA3 mif2::KAN5). Deletion constructs under the control of the endogenous MIF2 promoter were then integrated at LEU2 in these cells, and the ability of deletions to substitute for wild-type (wt) MIF2 tested by counterselection for pMIF2:URA3 on 5-fluoroorotic acid. Plates were incubated overnight at 30 or 37°C and photographed.

Expression and Purification of Mif2p Fragments

DNA encoding residues 256–549, 365–549, or 365–530 of Mif2p was amplified using the polymerase chain reaction (PCR) and inserted into the pET28a(+) plasmid (Novagen, Madison, WI) downstream of a cleavable, N-terminal hexahistidine tag. The fragment was overexpressed in Rosetta pLysS cells (Novagen), induced with 1 mM isopropyl β-D-thiogalactoside at an optical density of 0.6, and purified from the bacterial lysate by nickel-nitrilotriacetic acid chromatography (Qiagen, Valencia, CA), by using 2 ml resin/l culture. The purified protein was concentrated in an Ultracel-10-kDa centrifugal filter (Millipore, Billerica, CA) to a volume of 1 ml and further purified by size exclusion chromatography on a 16/60 Superdex S75 column (GE Healthcare, Piscataway, NJ) equilibrated with 10 mM HEPES, pH 7.5, 800 mM NaCl (fragment 256–549), or 250 mM NaCl (fragment 365–549).

Crystallization of Mif2p(365–530) and x-Ray Structure Determination

Mif2p(365–530) crystallizes from 3 to 4% polyethylene glycol (PEG) 3000, 70–100 M Li2SO4, or Na2SO4, 100 mM imidazole, pH 8.0, in space group P213 (a = 106.0 Å), with a dimer in the asymmetric unit. For flash freezing, the mother liquor was changed stepwise to 25% glycerol, 6% PEG 3000, with other components as in the initial crystallization. We identified derivateis by using a native gel to test binding with a number of compounds (Boggon and Shaprio 2000). Of the compounds that altered protein mobility, two (K2OsCl6 and di-m-iodobis(ethylenediamine) diplatinum [PIP]) could be soaked into the Mif2p(365–530) crystals without causing damage or nonisomorphism. We collected multiwavelength data sets from these two derivatives atbeamline 8.2.2 at the Advanced Light Source and processed the data using HKL2000 (Otwinowski and Minor, 1997). The Os sites were located using SOLVE (Terviliger and Berendzen, 1997); phases from this derivative were used to locate the Pt sites in the PIP derivative. Multiple isomorphous replacement refinement yielded an interpretable map using structure factors to 3.2-Å resolution.

Electron density for one of the two chains in the asymmetric unit was better defined than for the other chain. We initially built the clearest two strands of each molecule with the program O (Jones et al., 1991) and calculated a rotation matrix for noncrystallographic symmetry (NCS) averaging by using coordinates for residues in these strands and the NCS-related heavy-atom positions. We collected multiwavelength data sets from these two derivatives at beamline 8.2.2 at the Advanced Light Source and processed the data using HKL2000 (Otwinowski and Minor, 1997). The Os sites were located using SOLVE (Terviliger and Berendzen, 1997); phases from this derivative were used to locate the Pt sites in the PIP derivative. Multiple isomorphous replacement refinement yielded an interpretable map using structure factors to 3.2-Å resolution.

Figure 1. Domain organization of the Mif2p and CENP-C polypeptide chains. Top, the Mif2p polypeptide chain. Yellow, green, red, and blue blocks indicate regions of distinct function (assigned, in part, through results described herein). The two regions recognizably conserved in human CENP-C (middle) are a short “CENP-C signature sequence” (light green) and a C-terminal domain (blue) shown here to be a dimerizing element. The pink box in the human CENP-C diagram represents the presumptive DNA-binding domain. Various signature-motif sequences are shown in the bottom panel.
model for residues 437–530 was built into this map and refined, with cycles of rebuilding, using Refmac (CCP4). The final model has R cryst and Rfree of 21.0% and 24.2%, respectively; 90% of the residues are in the most favored region of the Ramachandran plot, and the remainder are in allowed regions. Coordinates and structure factors have been deposited in the PDB with accession number 2vpv. Data collection, phasing, and refinement statistics are shown in Table 1.

### Chromatin Immunoprecipitation
Chromatin Immunoprecipitation was performed based on the protocol of (Megge et al., 1999), with modifications. Specifically, cells were cross-linked with formaldehyde for 2 h at room temperature (RT) and then lysed using glass beads in a Bio101 FastPrep FP120. Genomic DNA was sonicated 6 times, then sheared, and the average length of the salmon sperm DNA was 500 base pairs. Immunoprecipitations (IPs) were performed using anti-myc (Ndc10p-myc; Santa Cruz Biotechnology, Mountain View, CA), anti-green fluorescent protein (GFP) (GFP-Cse4p and Cbf1p-GFP; Clontech, Palo Alto, CA), or no antibody (negative control; data not shown). An antibody fragment of Mif2p polypeptide, we designed a series of deletion constructs to substitute for wt MIF2 in the Mif2p polypeptide, we designed a series of deletion constructs to substitute for wt MIF2.

### Phenotypes of Mif2p Truncations
Phenotypes of Mif2p Truncations
To evaluate the roles of various protein sequence elements in the Mif2p polypeptide, we designed a series of deletion constructs based on multiple sequence alignments (Figure 1). A S288C haploid yeast strain was generated in which a deletion was created in the control of the endogenous MIF2. Deletion constructs under the control of the endogenous MIF2 promoter were then integrated at LEU2 in these cells, and the ability of the deletions to substitute for wt MIF2 was tested by counterselection for pMIF2:URA3 on 5-fluoroorotic acid. Two of the five regions of Mif2p were found to be essential for function: those between residues 267 and 347, a segment that contains a likely nuclear localization sequence (QRRKKQKK; Nair and Sorger), and those between residues 290 and 386, a region that contains the CENP-C signature box (green; Figure 1), and those between residues 267 and 347, a segment that contains a likely nuclear localization sequence (QRRKKQKK; Nair and Sorger, 2005).

### RESULTS

#### Phenotypes of Mif2p Truncations

To evaluate the roles of various protein sequence elements in the Mif2p polypeptide, we designed a series of deletion constructs based on multiple sequence alignments (Figure 1). A S288C haploid yeast strain was generated in which a chromosomal deletion of the endogenous MIF2 gene was rescued by MIF2 expressed from a URA3-CEN plasmid (to create pMIF2:URA3 mif2Δ−:KAN6). Deletion constructs under the control of the endogenous MIF2 promoter were then integrated at LEU2 in these cells, and the ability of the deletions to substitute for wt MIF2 was tested by counterselection for pMIF2:URA3 on 5-fluoroacetic acid. Two of the five regions of Mif2p were found to be essential for function: those between residues 267 and 347, a segment that contains the CENP-C signature box (green; Figure 1), and those between 331 and 438 (red; Figure 2). These regions contain the likely nuclear localization sequence (QRRKKQKK; Nair and Rost, 2005), but we did not explicitly test whether the sequence is essential for correct localization. The red region also contains an AT-hook, PRGPRKK, that is conserved in eight of nine sequenced fungal species with point centro-meres and in 19 of 31 fungal species overall (including

### Electrophoretic Mobility Shift Assays (EMSAs) or Bandshift Assay

Bandshift/competition probes were based on Saccharomyces cerevisiae CENIII. For each probe, 1 pmol (each) of two complementary oligonucleotides were annealed in 40 μl of 20 mM Tris, pH 8.0, 2 mM MgCl₂, and 50 mM NaCl, by heating at 90°C for 10 min and then slowly cooling to room temperature. Samples from each reaction were separated on agarose gels to ensure the presence of a single band and to quantitate the annealed product (Image-Quant software; GE Healthcare). Random 88-bp DNA was generated by PCR from the pUC19, purified and quantitated. Probe DNA was labeled with 32P-dATP using T4 polynucleotide kinase (NEB). To measure DNA-binding affinity, 30-μl reactions containing 0.2–4 μg of recombinant Mif2p or CBF3 Complex, 50 fmol of DNA probe, 9 μg casein, 3 μg of sheared salmon sperm DNA, in 10 mM HEPES (pH 8.0), 6 mM MgCl₂, 10% glycerol, and adjusted with 1 M KCl to a final concentration of 150 mM were incubated at RT for 45 min and then loaded onto a 4% non-denaturing polyacrylamide bandshift gel (as described in Sorger et al., 1995). Competition experiments were carried out by mixing unlabeled wild-type or random DNA (pUC19) with labeled probe DNA and then adding Mif2p. To measure the ratio of sequence-specific to nonspecific binding, sheared salmon sperm DNA was used as a competitor. The average length of the salmon sperm DNA was 500 base pairs.

---

**Table 1. Data and refinement statistics**

<table>
<thead>
<tr>
<th>Data collection</th>
<th>Native</th>
<th>PIP</th>
<th>Os MAD A1</th>
<th>Os MAD A2</th>
<th>Os MAD A3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wavelength (Å)</td>
<td>9.7950</td>
<td>1.50030</td>
<td>1.14083</td>
<td>1.14048</td>
<td>1.14010</td>
</tr>
<tr>
<td>Space group</td>
<td>P2₁,₃</td>
<td>P2₁,₃</td>
<td>P2₁,₃</td>
<td>P2₁,₃</td>
<td>P2₁,₃</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>2.70 (2.80–2.70)</td>
<td>3.00 (3.00–2.90)</td>
<td>3.10 (3.21–3.10)</td>
<td>3.20 (3.34–3.20)</td>
<td>3.20 (3.34–3.20)</td>
</tr>
<tr>
<td>Unique observations</td>
<td>11,184</td>
<td>15,323</td>
<td>12,290</td>
<td>12,098</td>
<td>11,793</td>
</tr>
<tr>
<td>Redundancy</td>
<td>5.2 (4.0)</td>
<td>1.9 (1.6)</td>
<td>4.3 (2.1)</td>
<td>4.3 (2.0)</td>
<td>4.2 (2.4)</td>
</tr>
<tr>
<td>Completeness</td>
<td>99.3 (95.8)</td>
<td>89.7 (80.6)</td>
<td>88.3 (57.6)</td>
<td>86.9 (47.9)</td>
<td>84.0 (68.8)</td>
</tr>
<tr>
<td>R cryst (%)</td>
<td>26.7 (2.3)</td>
<td>7.9 (43.3)</td>
<td>5.7 (36.8)</td>
<td>6.2 (38.5)</td>
<td>6.0 (39.7)</td>
</tr>
<tr>
<td>R free (%)</td>
<td>30.0–2.70</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Refinement</td>
<td>186</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Refined residues</td>
<td>186</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Refined water molecules</td>
<td>29</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R cryst (%)</td>
<td>21.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R free (%)</td>
<td>24.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average B factors (Å²)</td>
<td>56.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mi2p</td>
<td>53.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Water molecules</td>
<td>53.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R.m.s. deviations from ideal geometry</td>
<td>0.012</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bond lengths (Å)</td>
<td>1.53</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bond angles (°)</td>
<td>18.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dihedral angles (°)</td>
<td>1.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Improper angles (°)</td>
<td>1.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**a** Numbers in parentheses refer to highest resolution shell.

**b** R cryst = [Σ(|Fo| – |Fc|)/Σ|Fo|] × 100, where |Fo| is the average intensity of I symmetry related observations of reflections with Bragg index h.

**c** R cryst = [Σ(|Fo| – |Fc|)/Σ|Fo|] × 100, where |Fo| – |Fc| are the observed and calculated structure factors.

**d** R free was calculated as for R cryst, but on 5% of data excluded before refinement.
Schizosaccharomyces pombe). A deletion in the N-terminal region (residues 3–39) had no effect on growth. Deletion of the PEST (Δ103-255) and C-terminal (Δ455-548) regions of the protein resulted in strains that were slow growing and temperature sensitive, with similar phenotypes (large, single-budded cells, indicating mitotic arrest) at the nonpermissive temperature (data not shown).

Expression and Purification of Recombinant Mif2p Fragments

We were unable to express full-length Mif2p in either bacterial or insect cells. N-Terminal truncations of human CENP-C have been shown to yield increased expression levels (Lanini and McKeon, 1995); therefore, we chose to examine expression of truncated Mif2p. Residues 250–549 were selected for expression of a truncated protein based on sequence conservation among 14 fungal Mif2 proteins, which showed an N-proximal PEST region we presumed might destabilize the recombinant product (Figure 1). When fused to a C-terminal hexahistidine tag, the $M_r$ ~45-kDa fragment Mif2p(250-549) was indeed expressed at high levels in E. coli, and soluble protein could be recovered in good yield from extracts (Supplemental Figure 1A). After purification using a nickel-chelating resin, SDS-polyacrylamide gel electrophoresis (PAGE) also revealed the presence of a slightly more rapidly migrating species, shown by Edman degradation to correspond to Mif2p lacking residues 250–255 (data not shown), and to a species with $M_r$ ~37 kDa, which had lost the hexahistidine tag (Fig S1A). To identify additional unstructured segments in the Mif2p(250-549) polypeptide, purified protein was subjected to limited proteolysis using trypsin or chymotrypsin. A species with $M_r$ ~25 kDa was recovered as a stable product of digestion with either protease, and Edman sequencing showed it to corre-
spond to the C-terminal region of Mif2p in both cases (Supplemental Figure 1A). A polypeptide corresponding to the proteolytic endpoint, Mif2p(365-549)His6, was selected for further crystallographic studies, and other fragments were used for DNA binding assays.

Analytical equilibrium ultracentrifugation of bacterially expressed Mif2p(365-549)His6 yielded a molecular mass of 40.6 \pm 1.1 kDa (Supplemental Figure 1B); because the predicted mass of a single chain is 19.1 kDa, the protein was clearly dimeric. When SDS-PAGE was performed under reducing and nonreducing conditions, the apparent mass was 25 kDa in both cases, demonstrating that disulfide bond formation was not involved in dimerization. Gel filtration suggested a native mass closer to 70 kDa, indicating either an elongated structure or the presence of extended "arms"; crystallographic results, presented below, show the latter to be the case. From these biochemical studies, we conclude that the ~200 C-terminal residues in Mif2p correspond to a well folded domain that forms a stable, noncovalent dimer with extended (and potentially flexible) arms.

Crystallization and Structure Determination of Mif2p(365-530)

While performing a screen to find crystallization conditions for Mif2p(365-549)His6, we found that it was helpful to eliminate the C-terminal 19 residues of Mif2p from the expression construct and to remove the hexahistidine tag (by thrombin cleavage after purification). The resulting species, Mif2p(365-530), crystallized in 4% PEG 3000 at neutral pH in space group P213, with two molecules per asymmetric unit. The structure was determined as described in Materials and Methods, by using multiple isomorphous replacement with multiwavelength anomalous dispersion. The final model included density for residues 437–530 (Figure 3), but residues 365–436 were not visible, implying that they were disordered. Three pairs of chains (each pair defining an asymmetric unit of the P213 U cell) form a ring of six protein monomers (Supplemental Figure 2A), with a crystallographic threefold axis running through its center and a noncrystallographic twofold axis perpendicular to the threefold and intersecting it. One twofold relationship is determined by an extended protein interface; the other, by a tenuous contact that includes a disulfide. The disulfide seems to have formed during crystallization, because the crystals were grown in the absence of reducing agent, and the dimer in solution is noncovalent (see above). Nine water molecules were added in strong electron density, where the capacity to form two hydrogen bonds with groups on the protein was clear. The disordered N-terminal arm extends into the empty region of the crystal delimited by the hexamer packing (Supplemental Figure 2B). The resistance of this arm to both trypsin (Supplemental Figure 1A) and chymotrypsin (data not shown) suggests that it might be a flexibly tethered, folded unit rather than a fully disordered segment of polypeptide chain.

**The C-Terminal Domain of Mif2p Is a Dimeric, \( \beta \)-Jelly Roll**

Residues 441–530 of Mif2p fold as a nine-stranded \( \beta \)-jelly roll (Figure 3A), a domain first identified in virus capsids (Harrison et al., 1978) and subsequently in a variety of proteins, including transcription factors, lectins, and kinases (Soisson et al., 1997; Baker et al., 2001; Dunwell et al., 2001; Williams and Westhead, 2002). The jelly roll has come to be
called a “cupid” domain, because in many instances it has a barrel-like or cup-like aspect with a ligand-binding site at the open end of the cup (Dunwell et al., 2001). Indeed, multiple sequence alignments have predicted that the C-terminal region of CENP-C proteins might belong to this family (Dunwell et al., 2001). The Mif2p jelly-roll has a five-stranded, antiparallel β-sheet, with strand order ABIDG, and a four-stranded antiparallel sheet with strand order CHEF (Figure 3A). Many conserved residues are glycines or prolines, which break the hydrogen-bonding pattern of individual strands and allow the jelly roll to fold (Dunwell et al., 2001; Williams and Westhead, 2002). The positions of a number of aromatic and aliphatic residues are also conserved. Residues in the ABIDG sheet, which lies at the dimer contact, are better conserved than those in the CHEF sheet.

The ABIDG sheet (Figure 3A) creates an extended dimerization surface (Figure 3C). Conserved, generally nonpolar, residues line the interface. The surface area buried at the dimer interface is 1500 Å², comparable with the area buried in the complex of an antibody and a protein antigen (Janin and Chothia, 1990) and consistent with the observed stability of the dimer. The area buried at the other twofold contact is only 752 Å². The thermal parameters of the ABIDG interface residues are lower than those of other surface residues, consistent with their tightly complementary packing. The outer surface of the dimer has no obvious patches of conservation, however, nor does it have prominent insertions into the core cupin fold.

Dimerization In Vivo

Deletion of the C-terminal domain of Mif2p produces cells with a temperature-sensitive growth phenotype (Figure 2). To confirm that the effect is due to failure to dimerize, we introduced mutations Y451S, F452A, and F523A, designed to disrupt the dimer contact (see Supplemental Figure 3). As shown in Figure 2, these mutations indeed impaired cell growth at high temperature, whereas mutation of a conserved residue that lies outside the dimer interface, T488A, had no effect. The severity of our directed mutations is comparable with that of the classic temperature-sensitive mif2-3 mutation (a P505L change), which has previously been subjected to extensive genetic analysis (Brown et al., 1993). P505L should destabilize the conformation of a strand that connects one sheet to another and may destabilize the dimerization domain more generally. Thus, Mif2p dimerization is probably essential for function in vivo.

DNA Binding by Mif2p

Like its human homologue, CENP-C, Mif2p associates with centromeric DNA (Meluh et al., 1997). CEN binding by Mif2p in vivo requires both CBF3 (Meluh et al., 1997; Ortiz et al., 1999) and Cse4p (Westermann et al., 2003), and quantitative fluorescence microscopy estimates suggest the presence of one Mif2p dimer per CBF3 complex and hence one Mif2p dimer per kinetochore (Joglekar et al., 2006). We carried out high-resolution chromatin immunoprecipitation (ChIP) experiments to determine the approximate location of the Mif2p binding site. Cbf1p and Cep3p served as guides, because their centromeric binding sites are known (Bram and Kornberg 1987; Lechner and Carbon, 1991; Baker and Masison, 1990; Cai and Davis, 1990; Espelin et al., 1997). As shown in Figure 4A (top two panels), Cbf1p maps, as expected, to the CDEI-proximal end of the CENIV CDEI-II-III fragment (red line), whereas Cep3p binds toward the CDEIII-proximal end. The spread of a Gaussian fit to each histogram has a full-width at half maximum of ~350 base pairs, but the position of the centroid can be determined more precisely, as evident from the distributions in Figure 4 (e.g., Bevington and Robinson, 2002). Ndc10p and Mif2p generate very similar ChIP patterns, both biased toward the CDEIII-proximal side of the centromere (Figure 4A, bottom two panels). The centroid of the Mif2p distribution lies just outside the core CDEI-II-III interval of CENIV, but still within the “extended” (~80-base pair) region of CDEIII (Espelin et al., 1997). Ndc10p has a closely related distribution, but it also exhibits some tendency to immunoprecipitate the fragment at the opposite end of the CDEI-II-III interval. The relatively large size and extended shape of Ndc10p (Espelin et al., 1997) might account for this pattern. Alternatively, Ndc10p might have a somewhat bimodal binding distribution. Ndc10p is an essential component of the CDEIII-binding, CBF3 complex, but it also can bind CDEII (Espelin et al., 2003). Binding to CDEII in some of the cells (e.g., those cells in G2 that have not yet assembled mature kinetochores) could account for an apparent bimodality. The principal occupant of CDEII is, however, a Cse4p-containing nucleosome (Figure 4B; see below).

The presence of an AT-hook in Mif2p suggests binding to A:T-rich sequences. Preliminary EMSA demonstrated association of Mif2p(256-549) both with a DNA fragment that included CDEI, -II, and -III (data not shown) and also with an 88-base pair DNA fragment that included only CDEIII (Figure 5A). All binding experiments were performed in the presence of excess random sequence DNA to ensure specificity of the observed interactions. The amount of complex (as indicated by the intensity of the mobility shifted band) on the 88-bp CDEIII probe increased more or less linearly with the concentration of Mif2p, over the full probe concentration range, consistent with our conclusion from the ChIP data (above) that CDEIII contains the principal Mif2p binding site. The addition of excess “cold,” wild-type competitor DNA eliminated the Mif2p-CDEIII binding, whereas addition of DNA with a random sequence did not, further demonstrating specificity of the interaction (Figure 5A). Moreover, we could detect no binding to a radiolabeled 88-base pair negative-control fragment derived from plasmid sequences (Figure 5A). We estimate that Mif2p(256-549) binds wild-type, 88 base-pair CDEIII DNA probes, with a Kd of ~0.5 nM (Figure 5D).

Mif2p(256-549) forms a stable DNA–protein complex with 88-base pair CDEIII DNA; Mif2p(345-549), which contains the AT-hook but lacks the signature region, does not bind as strongly as Mif2p(256-549) (data not shown). The requirement of an AT-hook for CEN binding by dimeric Mif2p led us to examine whether the binding site is itself twofold symmetric, with a defined spacing between A:T-rich regions. The data in Figure 5, B and C, suggest that Mif2p binding requires at least 30 base pairs of DNA containing almost exclusively A:T

Vol. 19, October 2008 4485
We examined the consequences in vivo of deleting or mutating the region between residues 256 and 365 (Figure 2E). The AT-hook did not seem to be essential for viability, but combined deletion of the AT-hook and the sequence link-
Figure 5. Binding of Mif2p to *S. cerevisiae* CENIV. Recombinant proteins were incubated with radiolabeled DNA fragments, and complexes were resolved on nondenaturing bandshift gels (see Materials and Methods). (A) Binding of Mif2p to CENIV. Recombinant Mif2p(256-549) (lanes A–E), Mif2p(365-549) (lane G), or CBF3 complex (lane F), was mixed with 32P-radiolabeled 88-base pair wild-type CDEIII alone or in addition to unlabeled wild-type 88bp CDEIII (lane D) or 88-bp random plasmid DNA (lane E). Recombinant Mif2p(256-549) (lane H) or CBF3 complex (lane I) was mixed with 32P-radiolabeled 88-base pairs random plasmid DNA. (B and C) DNA requirements for Mif2p binding. Recombinant Mif2p(256-549) was incubated with 32P-radiolabeled DNA probes of indicated sequence and binding evaluated as described in A. The boxed CCG is the conserved binding triplet for one of the Cep3p Zn clusters (cf. Figure 6; McGrew et al., 1986, Ng and Carbon, 1987;
ing it to the dimerization domain had a pronounced effect on cell growth. In general, AT-hooks are adjunct DNA-binding elements (Aravind and Landsman, 1998), implying that the region of Mif2p between the AT-hook and the dimerization domain might also contribute to DNA binding. Indeed, this region might be the principal DNA-binding element, because the AT-hook is not absolutely conserved among budding yeast.

Co-occupancy of centromeres with CBF3 may enhance the sequence selectivity of Mif2p. We carried out the EMSA “supershift” experiment reported in Figure 5E to show that CBF3 and Mif2p can together occupy the same DNA fragment. The two bands present in all CBF3-containing lanes represent complexes with one Skp1p:Ctf13p heterodimer, one Cep3p homodimer, and one or two Ndc10p homodimers, respectively (Espelin et al., 1997). Only the band with a single Ndc10p shows a detectable supershift with Mif2p (Figure 5E). Mif2p and (Ndc10p)2 may be mutually exclusive on the second (Ndc10p)2 site, which includes the segment that is altered in sequence E of Figure 5C. Thus, one interpretation of the combined data in Figures 5C, lane E, and 5E is that the complete complex of DNA-binding proteins on a yeast centromere includes one Ndc10p dimer and one Mif2p dimer. It is possible, however, that experimental factors involving in vitro conditions and/or kinetics of binding might have obscured a supershift of the upper band.

We included the variant, H3-like histone Cse4p in our ChIP experiments. The immunoprecipitation histogram in Figure 4B (bottom) has a maximum just between those for Cbf1p and Cep3p, and it is thus centered precisely on CDEII. Moreover, a similar immunoprecipitation experiment with antibodies to H3 shows depletion exactly complementary to Cse4p occupancy (Figure 4B, top). These results, which are consistent with the recently published evidence for a single centromeric nucleosome in budding yeast (Furuyama and Biggins, 2007) and with earlier quantitative fluorescence measurements, suggesting that no more than one nucleosome is present per CBF3 complex (Joglekar et al., 2006), provide additional markers for validating the relative position of Mif2p. Our data also support the notion that the Cse4p-containing nucleosomal assembly binds CDEII.

Role of Mif2p in Kinetochore Assembly
To examine the role of Mif2p in kinetochore assembly, we used ChIP and fluorescence microscopy of GFP-tagged proteins to analyze CEN binding by inner and outer kinetochore proteins in mif2−3 and mif2−4d degron mutants under permissive (room temperature) and nonpermissive (37°C) conditions (Table 2 and Supplemental Figure 4). As specificity controls, we analyzed kinetochore recruitment of the CBF3 subunit Cep3p, which binds to CEN DNA independently of R. L. Cohen et al.

<table>
<thead>
<tr>
<th>Mutant background</th>
<th>Mutated protein (complex)</th>
<th>Analyzed protein (complex)</th>
<th>Temp</th>
<th>Positive control (Cep3p)</th>
</tr>
</thead>
<tbody>
<tr>
<td>cse4−1</td>
<td>Cse4-Scm3 nucleosome</td>
<td>Mif2p (Mif2p dimer)</td>
<td>24°C</td>
<td>+ a</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Okp1p-GFP (COMA)</td>
<td>37°C</td>
<td>− a</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mtw1p-GFP (MIND)</td>
<td></td>
<td>+ a</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ndc80p-GFP (Ndc80 complex)</td>
<td></td>
<td>+ a</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Spc105p-GFP (Spc105 complex)</td>
<td></td>
<td>+ a</td>
</tr>
<tr>
<td>mif2−3</td>
<td>Mif2p dimer</td>
<td>Cse4p-GFP (Cse4p-Scm3p nucleosome)</td>
<td></td>
<td>+ a</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Spc105p-GFP</td>
<td></td>
<td>+ a</td>
</tr>
<tr>
<td>mif2−4d</td>
<td>Mif2p dimer</td>
<td>Mtw1p-GFP</td>
<td></td>
<td>+ n.d. b</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ndc80p-GFP</td>
<td></td>
<td>+ n.d. d</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ctf19p-GFP (COMA)</td>
<td></td>
<td>+ n.d. d</td>
</tr>
<tr>
<td>ndc80−1</td>
<td>Ndc80 complex</td>
<td>Mif2p-TAF</td>
<td></td>
<td>+ n.d. d</td>
</tr>
<tr>
<td>spc105−15</td>
<td>Spc105 complex</td>
<td>Mif2p</td>
<td></td>
<td>+ n.d. d</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Okp1p-GFP</td>
<td></td>
<td>+ n.d. d</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mtw1p-GFP</td>
<td></td>
<td>+ n.d. d</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ndc80p-GFP</td>
<td></td>
<td>+ n.d. d</td>
</tr>
</tbody>
</table>

+ A protein localizes to CENIV (ChIP) and/or kinetochores (fluorescence microscopy). Recruitment of all proteins was scored by ChIP, except for Cse4p-GFP in mif2−3 and Mtw1p-GFP, Ndc80p-GFP, and Ctf19p-GFP in the Mif2p degron strain mif2−4d, for which recruitment was analyzed using wide-field Deltavision fluorescence microscopy (De Wulf et al., 2003). − , protein does not localize to CENIV (ChIP).

a Confirms a previous result reported by Westermann et al. (2003).
b n.d., not determined.
c See Westermann et al. (2003).
Figure 6. Diagram summarizing binding data in this article, in previous work on CBF3–centromere interactions (Espelin et al., 1997), and in the literature. CBF3 comprises a Skp1p:Ctf13p heterodimer associated closely with a Cep3p dimer; addition of one or two dimers of Ndc10p generates “core” and “extended” complexes, respectively. The Cse4p nucleosome contains Scm3p, Cse4p, and H4 in 1:1:1 proportion, but it seems to lack H2a and H2b (Mizuguchi et al., 2007). An interaction of Scm3p and Ndc10p has been proposed to be the mechanism by which CBF3 recruits the Cse4p nucleosome (Cameron et al., 2007). We discuss in the text whether adding a second Ndc10p dimer, to form an extended complex, is compatible with recruitment of Mif2p; the possibility, that the two are mutually exclusive is suggested by a white hatching across the second Ndc10p dimer. The guide at the bottom of the figure relates the spatial organization diagrammed in the upper part to the sequence and organization of centromeric DNA in yeast.

Mi6p, as well as Mi6p recruitment in ndc10-1 and cse4-1, two mutants in which Mi6p is lost from kinetochores at 37°C. We also analyzed Bim1p-GFP recruitment in mif2-3, because Bim1p requires Mi6p to associate with kinetochores (De Wulf, unpublished). These experiments confirm that CBF3 and Cse4p are required for CEN binding by Mi6p (Table 2). We found that Mi6p function is required to recruit the conserved, inner kinetochore, COMA complex (Ctf19p, Okp1p, Mcm21p, and Ame1p) as well as the conserved outer kinetochore MIND complex (Mt1p, Nfn1p, Sfn1p, and Dsn1p). The conserved outer kinetochore complexes, Ndc80 (Ndc80p, Nuf2p, Spc24p, and Spc25p) and Spc105 (Spc105p and Krc28p), associate with CEN DNA even under nonpermissive conditions (Table 2). We conclude that although the CBF3 complex is required for the recruitment of all kinetochore proteins (including Cse4p and Mi6p) to budding yeast centromeres, the Cse4p-Mi6p axis seems to be involved in recruiting only a subset. This selective recruitment corresponds to the phenotypes of mutations in proteins recruited by Mi6p and Cse4p. Although CBF3 is required to activate the spindle checkpoint, Cse4p, Mi6p, COMA, and MIND components are not (Gardner et al., 2001, Scharfenberger et al., 2003). Moreover, like mif2-3 and cse4-1, mutants in the COMA and MIND complexes arrest at G2/M with kinetochore attachment (either bipolar or monopolar) established, whereas mutants in the Ndc80 and Spc105 complexes pass through the checkpoint, for which both these complexes are essential (Kiyomitsu et al., 2007; De Wulf, unpublished data). Thus, Mi6p (together with upstream component Cse4p) recruits kinetochore proteins/complexes that contribute to sister chromatid biorientation and force generation in mitosis (e.g., COMA, MIND), proteins that depend on COMA and MIND (e.g., Ctf3), and proteins that protect centromeric cohesion during meiosis I (Iml3p, Ch4p). Kinetochore association by proteins involved in microtubule attachment and spindle checkpoint activity (e.g., the Ndc80p and Spc105p complexes) does not seem to require Mi6p, but only CBF3 (Supplemental Figure 4).

Proteins Related to Mi6p
Psi-BLAST searches, using the C-terminal domain sequence of Mi6p as a query, identified a large number of potential relatives in various species. These presumably fall into one or more of the 18 cupin-domain subfamilies tabulated previously (Dunwell et al., 2001). One of the 18 subfamilies contains helix-turn-helix (HTH) transcription factors, with the cupin domain at the C terminus (Aravind and Koonin, 1999). A search with the DALI server (Holm and Sander, 1995) by using the Mi6p dimerization domain as probe uncovers one of the members of this subfamily, Hth-3 from Vibrio cholerae, as a structural homologue. Indeed, Hth-3 contains a dimeric β-jelly roll, with essentially the same dimer interface as Mi6p (Figure 3C). Hth-3 has not been characterized functionally: it is the unpublished product of a structural genomics effort. We can assume that it is indeed a DNA-binding protein, because its HTH domain is convincingly similar to those of known bacterial repressors. Another class of cupin-domain containing transcription factors, in which the DNA-binding domain is C-terminal to the cupin domain, includes AraC (Dunwell et al., 2001). It thus seems that Mi6p is a “borrowed” transcription factor, with a conserved dimerization element present in a number of bacterial DNA-binding proteins.

DISCUSSION
We summarize in Figure 6 our conclusions from the present analysis of Mi6p–centromere interactions, together with the earlier detailed mapping of CBF3 contacts (Espelin et al., 1997) and other published data, as referenced in the figure caption. The DNA-binding properties of Mi6p depend on contacts from an AT-hook and probably from additional structural elements in the region of the molecule between the AT-hook and the dimerization domain. Optimal binding requires an AT-rich DNA segment of ~30 base pairs. In particular, the sequence between 25 and 35 base pairs to the right of the conserved CCG seems to be essential, at least for in vitro binding to CENIV. This palindromic sequence is appropriate for binding a Mi6p homodimer, but because it is not conserved among all centromeres, the symmetry could be coincidental. One possible binding model would place Mi6p across the major-groove dyad of the palindrome, with the AT-hooks inserted symmetrically into minor grooves one turn apart. Alternatively, other parts of the protein might interact with major-groove sequences one turn apart, with the AT-hooks inserted into the flanking minor grooves. The protein would then cover at least 20–25 base pairs. The AT-rich region of CDEIII to which Mi6p binds tightly is within the extended footprint of Ndc10p, which also binds AT-rich DNA. Some of our data suggest that Mi6p might exclude the second Ndc10p dimer in CBF3, which generates the extended coverage (Espelin et al., 1997). In any case, fully specific association of Mi6p with a centromere probably requires Ndc10p or other components of CBF3, on which
Mi2p depends for binding in vivo. The arrangement and overlap of centromeric binding sites for transcription-factor-like proteins recalls the properties of a transcriptional enhancer (Panne et al., 2007), as first suggested over a decade ago (Meluh and Koshland, 1995). In some enhanceosomes, the identity of proteins occupying particular sites can vary with the physiological state of the cell; likewise, in yeast kinetochores it is possible that Ndc10p and Mi2p occupy the same or overlapping sites at different cell cycle stages.

What about CDEII? The recent demonstration that the Cse4p-containing nucleosome, recruited jointly by CBF3 and Scm3p, lacks H2A and H2B, suggests that it might cover only ~80 base pairs of DNA, which corresponds closely to the required length of CDEII in *S. cerevisiae* (Gaudet and Fitzgerald-Hayes, 1987; Mizuguchi et al., 2007). Moreover, in two budding yeasts with point centromeres, *Eremothecium gossypii* and *Kluveromyces lactis*, CDEII is twice that length and might therefore contain two such tetramer-core nucleosomes (Meraldi et al., 2006). Wrapping around a histone tetramer core would probably exclude Ndc10p and Mi2p from CDEII, even though Ndc10p can associate with empty CDEII in vitro (Espelin et al., 2003).

The parts of Mi2p that are clearly conserved in CENP-C proteins of higher eukaryotes are the CENP-C signature element and the dimerization domain. The presumptive DNA-binding domain lies between these regions in yeast but N-terminal to the CENP-C signature in metazoans and plants. Because DNA-binding regions of many yeast transcription factors (e.g., the Zn2Cys6 clusters of Gal4p and its relatives, including the Cep3p component of CBF3) seem to have evolved divergent structures, and as CBF3 itself is yeast specific, the transposition of the two essential regions is still consistent with conserved overall function.

Mi2p is not required for recruitment of Cse4p, but it is required for centromeric association of the MIND and COMA complexes and of proteins that depend on MIND and COMA for kinetochore localization (e.g., Ctf13 complex, Imp3p, Ch14p; Meadlay et al., 2002) (Supplemental Figure S4). Although the presence of Mi2p at a centromere leads to recruitment of only a subset of kinetochore components in budding yeast, its orthologues, Cnp1 in fission yeast and CENP-C in animal cells, participate in recruitment of all kinetochore subunits (except for the upstream Cse4p orthologues Cnp1 and CENP-A) (Oegema et al., 2001; Cheeseman et al., 2004; Liu et al., 2006; Kwon et al., 2007). Thus, CEN-P-C in higher eukaryotes seems to have functions that are shared between Mi2p and CBF3 in budding yeast.

The dimerization domain is necessary for proper Mi2p function, but it is not essential for viability. It is therefore unlikely to be the recruitment structure. Indeed, its surface exhibits no particularly extended patches of conserved residues, as one might expect for a protein association domain. We suggest instead that the conserved CENP-C signature element, present in homologues of every species examined, creates at least part of the contact for the next members of the assembly hierarchy.

ACKNOWLEDGMENTS

We thank Nick Larsen for help with data collection; John Bellizzi, Piotr Sliż, and Ben Spiller for advice on crystallographic methods; Jawdat Al-Bassam, for help with analytical ultracentrifugation; Kevin Corbett, for some final adjustments and refinement of the model; Caroline George for help with the kinetochore recruitment experiments; members of Fred Winston’s laboratory (Harvard Medical School) for advice on yeast-strain constructs; and JJ Mirzam for discussion. P.K.S. acknowledges National Institutes of Health grants GM-51464 and CA-84179. S.C.H. is an Investigator in the Howard Hughes Medical Institute. P.D.W. and K.T.S. acknowledge fellowships from the Charles King Trust.

REFERENCES


R. L. Cohen et al.


