Proc. Natl. Acad. Sci. USA
Vol. 93, pp. 502–507, January 1996
Biochemistry

Identification, purification, and molecular cloning of autonomously replicating sequence-binding protein 1 from fission yeast *Schizosaccharomyces pombe*

(DNA-binding protein/CENP-B)

YOTA MURAKAMI*, JOEL A. HUBERMAN†, AND JERARD HURWITZ‡

*Department of Viral Oncology, Institute for Virus Research, Kyoto University, Sakyo-ku, Kyoto 606, Japan; †Department of Molecular and Cellular Biology, Roswell Park Cancer Institute, Buffalo, NY 14263; and ‡Graduate Program in Molecular Biology, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, Box 97, New York, NY 10021

Contributed by Jerard Hurwitz, October 9, 1995

ABSTRACT **Autonomously replicating sequence (ARS) elements of the fission yeast *Schizosaccharomyces pombe* contain multiple imperfect copies of the consensus sequence reported by Maundrell *et al.* [Maundrell K., Hutchison, A. & Shall, S. (1988) *EMBO J.* 7, 2203–2209]. When cell free extracts of *S. pombe* were incubated with a dimer or tetramer of an oligonucleotide containing the ARS consensus sequence, several complexes were detected using a gel mobility-shift assay. The proteins forming these complexes also bind *ars3002*, which is the most active origin in the *ura4* region of chromosome III of *S. pombe*. One protein, partly responsible for the binding activity observed with crude extracts, was purified to near homogeneity. It is a 60-kDa protein and was named ARS-binding protein 1 (Abp1). Abp1 preferentially binds to multiple sites in ARS 3002 and to the DNA polymer poly[d(A·T)]. The cloning and sequence of the gene coding for Abp1 revealed that it encodes a protein of 59.8 kDa (522 amino acids). Abp1 has significant homology (25% identity, 50% similarity) to the N-terminal region (≈300 amino acids) of the human and mouse centromere DNA-binding protein CENP-B. Because centromeres of *S. pombe* contain a high density of ARS elements, Abp1 may play a role connecting DNA replication and chromosome segregation.**

In the budding yeast *Saccharomyces cerevisiae*, "replicators" have been identified as short stretches of DNA (≈100 bp), called autonomously replicating sequence (ARS) elements, that support the autonomous replication of plasmids in cells. A subset of ARS elements was shown to colocalize with origins on chromosomes (1). Fine structure analyses of the ARS1 and ARS307 origins revealed the presence of several short functional elements, including an essential A element that contains the 11-bp ARS consensus sequence and other important B elements (2–4). A multisubunit origin recognition complex (ORC) that binds to the A and B1 elements in the presence of ATP has been identified as a candidate for an "initiator protein," which facilitates formation of an initiation complex at replication origins (5).

In contrast, little is known about the sequence requirements for replication origins in animal cells due to the lack of reliable ARS assays and difficulty in the direct genetic manipulations of chromosomes. However, available data suggest that animal replicators may be less specific or function differently than those of budding yeast (6).

Because the fission yeast *Schizosaccharomyces pombe* is thought to be similar in some respects to higher eukaryotes than *S. cerevisiae*, a study of DNA replication in *S. pombe* may give us a better appreciation of this process in higher eukaryotic cells. ARS elements have been identified in *S. pombe* and were shown to colocalize with origins on chromosomes (7–9). The ARS elements of *S. pombe* are larger than those of *S. cerevisiae* (>500 bp vs. ≈100 bp) and are thought to contain multiple redundant sequence elements distributed over a longer region (9, 10), supporting the idea that *S. pombe* origins may be more related to mammalian origins than those of *S. cerevisiae.*

We have begun to identify origin-binding protein(s) from fission yeast using *ars3002*, which is the most active origin in the *ura4* origin region (containing three distinct origins) (7). Several binding activities have been detected that interact with *ars3002*, and one of these proteins has been purified to homogeneity. This 60-kDa protein, named Abp1, binds to multiple sites in *ars3002.* The gene for Abp1¶ has been cloned, and the deduced amino acid sequence reveals that the protein has homology to the human centromere DNA-binding protein CENP-B.

MATERIALS AND METHODS

Preparation of Substrates for Gel Mobility-Shift Assays. A set of 31-nt-long complementary oligonucleotides containing multiple overlapping matches to the *S. pombe* ARS consensus sequence (designated MMACS, for multiple Maundrell ARS consensus sequence, see Fig. 1) was annealed, phosphorylated by T4 kinase, and then ligated by T4 DNA ligase. The ligated DNA was digested with *Bam*HI and *Bgl* II. Because MMACS contains a *Bam*HI site at one end and a *Bgl* II site at the other end, head-to-tail ligation results in the loss of both restriction sites. As a result, digestion with both restriction enzymes yields only direct repeats of MMACS. After PAGE, DNA fragments, corresponding to the monomer, dimer, and tetramer forms of MMACS, were eluted and cloned into the *Bam*HI site of pBluescript (Stratagene). To label the multimer of MMACS, the plasmid DNA was digested with *Xho* I and *Xba* I and then end-labeled with T4 DNA polymerase and [α-^{32}P]dCTP. After PAGE, the labeled fragments were eluted from the gel and used as substrates in the gel mobility-shift assay.

Gel-Shift Assay. Reaction mixtures (15 μl) containing 40 mM Hepes·NaOH (pH 7.5), 5 mM magnesium acetate, 2 mM dithiothreitol, 5% glycerol, 2% polyethylene glycol (M_r 20,000), 0.5 μg of poly[d(A-C)] as competitor, 2 μg of bovine serum albumin, 5 fmol of end-labeled monomer, dimer, or tetramers of MMACS (5–10 × 10^3 cpm/fmol), and indicated amounts of crude extracts or purified fractions were incubated for 30 min at 32°C. Loading buffer [2 μl, 0.5% bromophenol

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: Abp1, ARS-binding protein 1; ARS, autonomously replicating sequence; ORC, origin recognition complex; MMACS, multiple Maundrell ARS consensus sequence.

¶The sequence reported in this paper has been deposited in the GenBank data base (accession no. U39079).

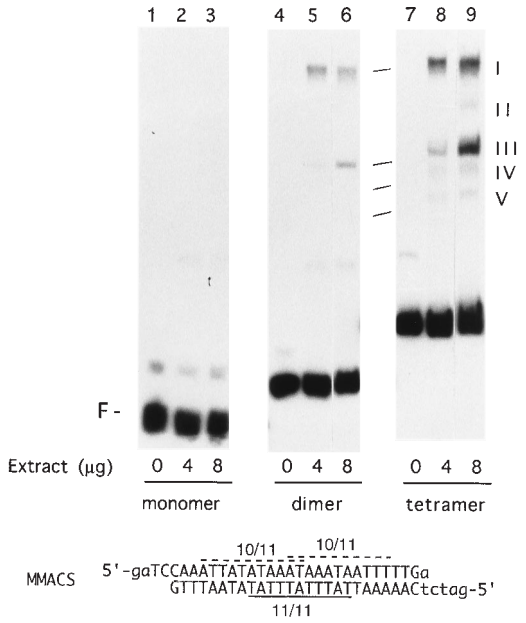

FIG. 1. (*Upper*) Detection of MMACS binding activities in crude extracts of *S. pombe* by gel-mobility shift. Indicated amounts of whole-cell extract from *S. pombe* were incubated with a monomer, dimer, or tetramer of labeled MMACS oligonucleotides and subjected to PAGE as described. Positions of distinct DNA–protein complexes I–V are indicated. (*Lower*) Sequence of the MMACS oligonucleotide used. Underlined and dashed lines indicate the perfect match and one-base mismatch sequences, respectively, related to the ARS consensus sequence reported by Maundrell *et al.* (11).

blue/0.5% xylene cyanol/40% (vol/vol) glycerol] was added, and the reaction mixtures were electrophoresed through a 5% polyacrylamide gel containing 5% glycerol in 0.5× TBE (1× TBE is 91 mM Tris base/89 mM boric acid/2 mM EDTA) for 2 hr at 10 V/cm. After electrophoresis, gels were dried on Whatman 3MM paper and autoradiographed.

Purification of Abp1. The DNA-binding activity of Abp1 was measured by the gel mobility-shift assay described above using the MMACS dimer as the substrate. All buffers used contained 0.1 mM phenylmethanesulfonyl fluoride, 1 mM benzamidine, 10 mM $NaHSO_3$, leupeptin at 0.5 μg/ml, and 0.5 μg of antipain. Logarithmically growing 972 h^- cells (800 g, wet cells) were frozen in liquid nitrogen and then pulverized in a Waring laboratory blender in liquid nitrogen. After being ground, the fine powder was suspended in the same volume of extraction buffer [0.1 M Hepes·KOH, pH 7.5/10 mM magnesium acetate/4 mM dithiothreitol/1 M KCl/2 mM EDTA/2 mM EGTA/0.04% Nonidet P-40/20% (vol/vol) glycerol] and stirred at 4°C for 1 hr. After centrifugation at 34,000 rpm at 4°C in a 50.2 Ti rotor, the supernatant was dialyzed for 16 hr against 12 liters of buffer H [50 mM Hepes·KOH, pH 7.5/5 mM magnesium acetate/1 mM EDTA/1 mM EGTA/0.02% Nonidet P-40/10% (vol/vol) glycerol]. The dialyzed extract was loaded onto a S-Sepharose column (Pharmacia LKB, 5 × 16 cm) equilibrated with buffer H/0.1 M KCl. The column was washed with 3 liters of buffer H/0.1 M KCl, and protein was eluted stepwise with 1.5 liters each of buffer H/0.25 M KCl, buffer H/0.4 M, and buffer H/0.8 M KCl. The binding activities that formed complex I, complex III, and complexes IV and V (see Fig. 1) were eluted with buffer H plus 0.25, 0.4, and 0.8 M KCl, respectively. The fraction that eluted with buffer H/0.4 M KCl was dialyzed against 8 liters of buffer H/0.1 M KCl for 16 hr and loaded onto a Q-Sepharose column (Pharmacia-LKB, 2.5 × 10 cm) equilibrated with buffer H/0.1 M KCl. After being washed with 0.5 liter of buffer H/0.1 M KCl, the column was eluted with a 500-ml linear gradient of KCl (0.1–0.6 M) in buffer H. Active fractions (0.25 M KCl) were diluted to 0.1 M KCl with buffer H and loaded onto a sequence-specific DNA-affinity column (1 × 1.5 cm) equilibrated with buffer H/0.1 M KCl. The sequence-specific DNA column was prepared as described by Lusky *et al.* (12) with a self-complimentary biotinylated oligonucleotide 111 nt in length (5′-biotin-GGAAATTATATAAATAAATAATTTT-TGGCCAAATTATATAAATAAATAATTTT-TGGCCAAAAATTATTTATTTATATAATT-TGGCCAAAAATTATTTATTTATATAATTTC-3′). The biotinylated oligonucleotide contained four tandem repeats of nt 3374–3398 of ARS 3002 (10), similar to the nucleotides in the MMACS monomer (3371–3398). Protein was eluted with a 30-ml linear gradient of KCl (0.1–0.6 M), and the active fractions (0.3 M KCl) were diluted to 0.1 M KCl and again passed through the sequence-specific DNA column and eluted as described above. Active fractions, measured by gel mobility shift, were pooled. Half of the pooled fraction was diluted to 0.1 M KCl, concentrated by passage through a 0.2-ml Q-Sepharose column equilibrated in buffer H/0.1 M KCl, followed by step elution with 1 ml of buffer H/0.4 M KCl. The peak fractions, measured by the gel-shift assay, were applied to a 5-ml, 15–35% glycerol gradient in buffer H/0.2 M KCl. After centrifugation at 37,000 rpm in a SW40 rotor for 18 hr, fractions (0.16 μl) were collected from the bottom of the gradient. A summary of the purification procedure is presented below in Table 1. Active fractions stored at −80°C retained 50% of their activity over a period of 3 mo with repeated freezing and thawing.

Cloning of Abp1. Five distinct tryptic peptide sequences were obtained from the purified preparation of Abp1. Using the information derived from two of these (see Fig. 5 for amino acid sequences), degenerate oligonucleotides (26-mer) were synthesized and used for PCR reactions with a λgt11 genomic library (Clontech). A PCR product 720 bp in length was cloned into a plasmid vector and sequenced from both ends to identify the amino acid sequences contained within peptides used for the synthesis of the degenerate oligonucleotide. Three positive clones were identified after screening of the λgt11 genomic

Table 1. Summary of the purification of Abp1

Fraction	Volume, ml	Protein, mg	Total units*	Specific activity, units × 10^3/mg
Whole-cell extract	1300	7900	130,000	0.017
I S-Sepharose	92	147	36,800	0.25
II Q-Sepharose	43	47	34,400	0.73
III 1st-MMACS-agarose	12	0.54	18,000	33
IV 2nd-MMACS-agarose	55	0.06	6,600	110
V Glycerol gradients†	0.64	0.0012‡	190	160

*One unit was defined as the amount of protein required to convert 50% of the substrate (MMACS dimer) to complexes in the standard gel mobility-shift assay described in text.
†In this step, only half of the second MMACS-agarose fraction was used, and the results presented were calculated assuming that the entire fraction IV was used.
‡Protein concentration was estimated based on silver staining.

library with the PCR products. All three clones were derived from the same genomic sequence. An insert from the longest clone (≈2 kb) was subcloned into pBluescript (Stratagene), and the entire sequence was determined. The cloned fragment contained a long open reading frame that included all of the tryptic sequences obtained from Abp1 but lacked the C-terminal region. A DNA fragment containing the missing C-terminal region of Abp1 was obtained from a cosmid clone that was identified by screening a set of cosmid DNAs that covered the entire *S. pombe* genome (13).

RESULTS

Detection of ARS-Binding Proteins in *S. pombe*. Because *ars3002* is the most active of the three different origins located in the *ura4* origin region on chromosome III in *S. pombe* (7), we wished to use it to identify ARS-binding proteins. However, the *ars3002* origin is >600 bp long and therefore too large for conventional gel-mobility shift or DNase I footprinting assays. Although the functionally important sequence motifs within *S. pombe* ARS elements have not been fully defined, the 11-bp consensus sequence identified by Maundrell *et al.* (11) is suspected to have functional significance. Perfect matches to this consensus are found in *S. pombe* ARS elements far more frequently than expected by chance (10, 11), and multiple 9/11 or better matches are found in all known *S. pombe* ARS elements (10, 11). In *ars3002*, there are 15 matches of 9/11 or better (10). For this reason we used a 31-nt oligonucleotide that included a 27-bp duplex stretch of *ars3002* [nt 332–3398 (10)] with one 11/11 match and two 10/11 matches (Fig. 1 *Lower*) to identify putative ARS-binding proteins.

Monomer, dimer, and tetrameric forms of MMACS were used as substrates in the gel mobility-shift assay. When the MMACS monomer was incubated with crude extracts from *S. pombe*, little binding activity was detected. However, four complexes were detected when crude extracts were incubated with the MMACS dimer. The same complexes were detected with the MMACS tetramer, though an additional complex, complex II, was observed (Fig. 1, lanes 4–9).

Specificity of the Binding Activities. To check the binding specificity of the complexes formed with the crude extracts, competition experiments with different unlabeled polynucleotides were used (Fig. 2*A*). Each unlabeled oligonucleotide was self-ligated to produce multimers that were screened as competitors in the gel mobility-shift assay. Multimers of unlabeled MMACS blocked or reduced the formation of all complexes produced with the labeled MMACS tetramer as substrate (lanes 3–5). The introduction of three point mutations into the perfect *S. pombe* ARS consensus and flanking sequences in MMACS blocked its ability to compete in the mobility-shift assay (Fig. 2*A*, lanes 6–8), suggesting an important role of the consensus sequence in complex formation. In support of this, multimerized oligonucleotides containing only the consensus sequence (MMACS-S in Fig. 2*A*) competed in complex formation almost as efficiently as the MMACS multimer (lanes 9–11). In contrast, multimerized oligonucleotides containing the A element of *S. cerevisiae* ARS1, which includes an ARS consensus sequence (SACS in Fig. 2*A*), did not prevent complex formation although the A element, like MMACS, is A+T rich (lanes 12–14). The unlabeled MMACS monomer blocked the formation of all complexes but at an efficiency 100-fold lower than that observed with ligated oligonucleotides (data not presented). These results further support the conclusion that complexes were formed preferentially on multimeric MMACS.

To examine the interaction between proteins in the crude extract and the native ARS element, competition experiments were done using a plasmid containing a 582-bp-long *ars3002* [nt 3294–3875 (10)] (Fig. 2*B*). The plasmid containing *ars3002* was an effective competitor, preventing complex formation with the labeled MMACS dimer (lanes 3–5), whereas the vector DNA did not compete (lanes 6–8). We conclude that proteins that form complexes with the MMACS multimers bind to the *ars3002* sequences as well.

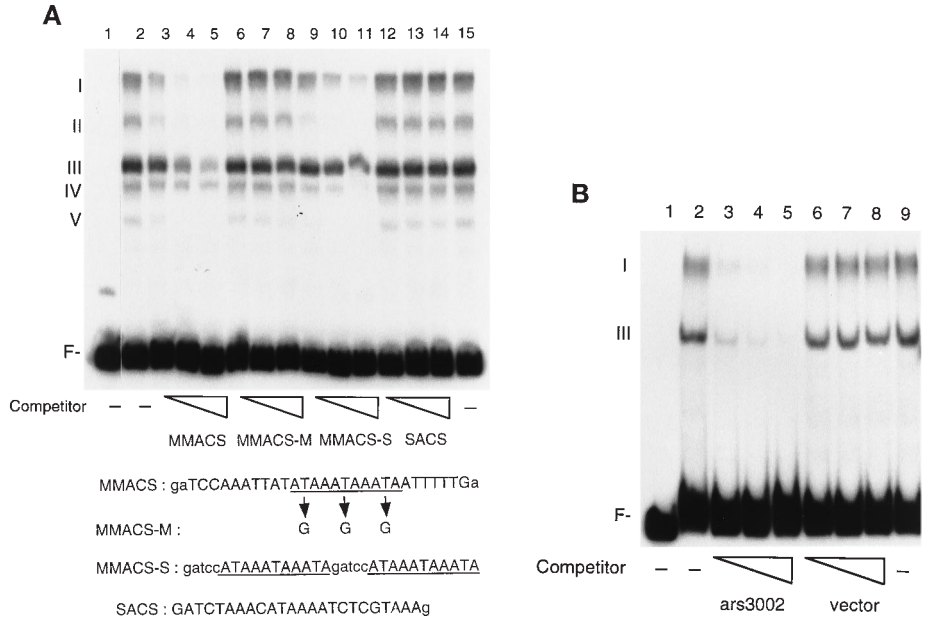

FIG. 2. Binding specificity of DNA–protein complexes formed on multimers of MMACS. (*A*) Increased levels of unlabeled oligonucleotide multimers whose sequences are shown below were added to reactions containing 8 μg of whole-cell extract and 5 fmol of labeled MMACS tetramer. Reaction mixtures were subjected to PAGE as described. The amounts of oligonucleotide added were in 5-, 25-, and 100-fold molar excess to the labeled substrate (in terms of nucleotides). The underlined nucleotides indicate a perfect match to the ARS consensus sequence. For the oligonucleotide MMACS-M only the mutated sequences are shown. (*B*) Increased levels of plasmid DNAs, containing either a 582-bp-long *ars3002* sequence or only vector sequence, were added to reactions containing 4 μg of whole-cell extract and 5 fmol of labeled MMACS dimer. The amounts of plasmid DNAs added were 50, 100, and 200 ng.

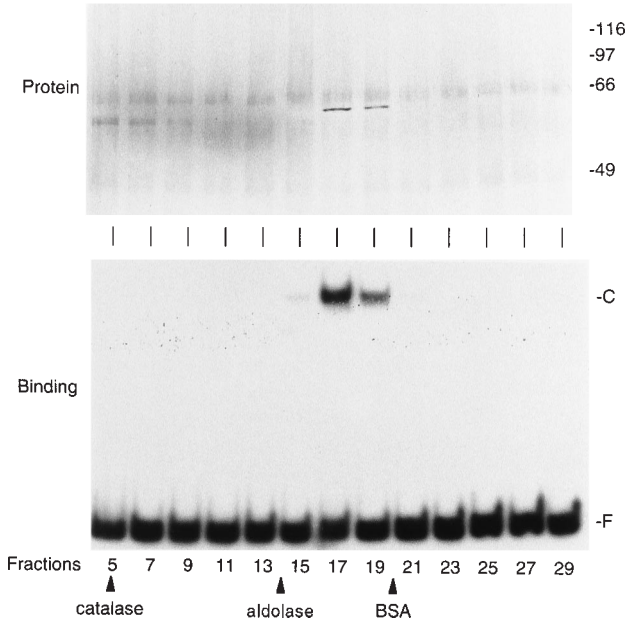

FIG. 3. Glycerol-gradient sedimentation of fractions obtained after the second MMACS-agarose column. In this experiment, 50 μg of protein in 0.1 ml was loaded onto a 15–35% glycerol gradient (5 ml, containing buffer H/0.1 M KCl). After centrifugation for 18 hr at 37,000 rpm in a SW40 rotor, 30 fractions (170 μl each) were collected from the bottom of the gradient. Samples (50 μl) of each fraction were precipitated with acetone and subjected to SDS/PAGE followed by silver staining of proteins (*Upper*). Positions of molecular-weight-marker proteins are indicated at right. C indicates the position of the Abp1 complex. Another aliquot (2 μl) of each fraction was used for gel mobility-shift assays with the labeled MMACS dimer as substrate (*Lower*). Positions of marker proteins in parallel glycerol gradients are indicated. BSA, bovine serum albumin.

Purification of Abp1. A protein that formed complex II with the MMACS dimer was purified to near homogeneity as described. The procedure described above resulted in purification of the protein $\approx 10^4$-fold with a recovery of 0.1% (Table 1). Fig. 3 shows a mobility-shift assay of the fractions obtained after glycerol gradient centrifugation, the final step in purification of this protein. The peak of the binding activity (using the MMACS dimer as substrate) comigrated with a 60-kDa protein. From these results, we concluded that the 60-kDa protein was responsible for the binding activity and named the protein Abp1 (ARS-binding protein 1). The activity sedimented between aldolase and bovine serum albumin with a sedimentation coefficient of 5.8.

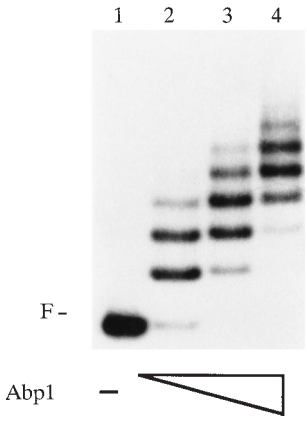

FIG. 4. Binding properties of purified Abp1. Purified Abp1 (8, 16, and 32 ng) was incubated with 2 fmol of labeled 582-bp-long *ars3002* DNA. The reaction mixtures were subjected to 1.5% agarose gel electrophoresis as described.

Characterization of Binding Activity of Abp1. The DNA-binding activity of Abp1 was stimulated by Mg^{2+}. The optimal concentration of Mg^{2+} was 5 mM; binding activity decreased in the presence of lower or higher levels of Mg^{2+}, and 25% of the maximal activity was detected in the presence of 10 mM Mg^{2+} or 10 mM EDTA. The addition of ATP did not influence the binding activity. However, in the presence of 10 mM Mg^{2+}, 2- to 4-fold stimulation was observed at high levels of ATP (2–4 mM), which likely reduced the inhibitory effects of high Mg^{2+} concentrations. No ATPase activity was detected in glycerol gradient fractions.

The interaction between purified Abp1 and the 582-bp-long ars3002 was examined (Fig. 4). Multiple bands were observed with increased amounts of Abp1 in the gel-shift assay. Seven discrete bands were detected at the highest concentration of Abp1 added. Presumably, the slower migrating bands contain multiple protein molecules per molecule of the 582-bp-long *ars3002* substrate, suggesting that Abp1 binds to multiple sites in *ars3002.*

Because *S. pombe* ARS and MMACS sequences are highly (A+T) rich, it is likely that Abp1 recognizes (A+T)-rich sequences in DNA. For this reason, we examined the binding of Abp1 to synthetic homopolymers in competition experiments. Of the synthetic polymers tested, only poly(dA-dT) efficiently competed and markedly reduced formation of the Abp1–MMACS complex, whereas poly(dA)-poly(dT) did not (data not presented). This result suggests that Abp1 preferentially binds to the alternating poly (dA-dT) sequence.

Cloning of Abp1. The sequences of five different tryptic peptides derived from Abp1 were obtained. With this information, the genomic DNA containing the complete coding region of Abp1 was cloned as described. During the process of cloning, a set of cosmid DNAs covering the entire *S. pombe* genome was screened. This localized the *abp1* gene to contig 8 of chromosome II (13). No known CDC mutant has been mapped to this position.

The nucleotide sequence of the cloned DNA is shown in Fig. 5*A*, along with the predicted amino acid sequence. The expected protein contains 522 amino acids, corresponding to 59.8 kDa, and includes amino acid sequences identical to the five peptides obtained by direct sequencing of the isolated protein (which are underlined). No introns were detected in this open reading frame. A putative TATA box was identified 105 bp upstream of the expected methionine initiation codon.

Abp1 Has Homology to Human and Mouse CENP-B. Homology search of the data base indicated that Abp1 possesses significant homology to the human and mouse CENP-B, which binds to specific sequences (CENP-B box, C/TTTCGTTG-GAARCGGGA) present in the highly repeated α-satellite DNA of the centromere region (14). The CENP-B and Abp1 proteins were $\approx$25% identical and 48% similar in their N-terminal (130 amino acids) and central regions (110 amino acids) (Fig. 5*B*). Several short amino acid stretches, as indicated, are highly conserved.

DISCUSSION

In this report we have identified and purified a 60-kDa protein, Abp1, which binds to multimers of the *S. pombe* ARS consensus sequence (11) and to multiple sites in an *S. pombe* origin, *ars3002.* We cloned the genomic DNA encoding Abp1. We are confident that the isolated genomic clone encodes Abp1 for the following reasons. The DNA contained a single open reading frame that encodes a protein of 59.8 kDa, consistent with the mass of purified Abp1. The amino acid sequence of the open reading frame contained five distinct tryptic peptides derived from purified Abp1. Polyclonal antibodies raised against the bacterially expressed protein reacted

A

```
ATATTACGTTACGCTGTTCAGTATAATATTACGAACTCGTGACTGTTAGTGTCGTTTTAT 60
TAAGGCAAGAAACCTTTGCTCTGCTATCCTTATATTTTCACTTCAAGACATTAGAAATAA 120
TAAAGCTATCGAGTAATGGGAAAAATCAAAAGAAGAGCCATTACAGAACATGAAAAACGC 180
                M  G  K  I  K  R  R  A  I  T  E  H  E  K  R

GCTCTTAGGCACTACTTTTTTCAACTTCAAAACCGTTCCGGTCAACAAGATCTTATCGAA 240
A  L  R  H  Y  F  F  Q  L  Q  N  R  S  G  Q  Q  D  L  I  E

TGGTTCCGTGAAAAGTTTGGCAAAGATATTTCACAGCCTTCCGTCTCTCAGATCTTATCC 300
W  F  R  E  K  F  G  K  D  I  S  Q  P  S  V  S  Q  I  L  S

TCCAAATACTCGTATCTGGACAATACTGTAGAAAAACCATGGGACGTGAAACGTAACAGA 360
S  K  Y  S  Y  L  D  N  T  V  E  K  P  W  D  V  K  R  N  R

CCACCGAAATACCCTCTTTTAGAGGCCGCTTTATTTGAATGGCAAGTTCAACAGGGTGAC 420
P  P  K  Y  P  L  L  E  A  A  L  F  E  W  Q  V  Q  Q  G  D

GATGCTACACTTTCCGGTGAAACAATTAAGAGAGCCGCCGCTATTCTTTGGCACAAAATT 480
D  A  T  L  S  G  E  T  I  K  R  A  A  A  I  L  W  H  K  I

CCCGAGTATCAAGATCAACCGGTTCCGAACTTTAGCAATGGATGGTTGGAAGGTTTCAGG 540
P  E  Y  Q  D  Q  P  V  P  N  F  S  N  G  W  L  E  G  F  R

AAGCGCCATATACTTCATGCAATTAATGAACAACCCACCGAGTCTGTGGTTTTGAATAAT 600
K  R  H  I  L  H  A  I  N  E  Q  P  T  E  S  V  V  L  N  N

ACGGAACCACCTAATGATCCTCTCTCTCGTGTTTACGATGTCACGCGATTAACGAATATT 660
T  E  P  P  N  D  P  L  S  R  V  Y  D  V  T  R  L  T  N  I

AATGATATCTTCACAATGCAAGAGACCGGTCTTTTTTGGAAATTGGTTCCAAATGGCACT 720
N  D  I  F  T  M  Q  E  T  G  L  F  W  K  L  V  P  N  G  T

CCCGAAGTCGAAGATATAAAAGGTATTACTCGATTCAAAGCTCGTATTACACTAACCGTG 780
P  E  V  E  D  I  K  G  I  T  R  F  K  A  R  I  T  L  T  V

TGTTGTAATGCTTCGGGAACCGAACGTTTGCCTTTATGGGTTATTGGATACTCACAATCT 840
C  C  N  A  S  G  T  E  R  L  P  L  W  V  I  G  Y  S  Q  S

CCCAGAGTGTTTCGGGCTGCTAATGTAAAACCCGAAGTAATGAATTTCAAATGGAGAAGT 900
P  R  V  F  R  A  A  N  V  K  P  E  V  M  N  F  K  W  R  S

AATGGTAAGGCATCCATGACCACAGCTATAATGGAGGAATGGCTACGTTGGTTTGATGCG 960
N  G  K  A  S  M  T  T  A  I  M  E  E  W  L  R  W  F  D  A

TGTATGGAAGGTCGCAAAGTCATACTTCTTATTGATAGCTATACTCCACATTTGCGTGCT 1020
C  M  E  G  R  K  V  I  L  L  I  D  S  Y  T  P  H  L  R  A

GTTGAAAACATTCGAAACTCTGGTAATGACCTCAGAAATACAACTGTAATTACTTTACCA 1080
V  E  N  I  R  N  S  G  N  D  L  R  N  T  T  V  I  T  L  P

TCTACTTCAGCCTCTATTTCCCAGCCTTGTAGTGAGGGTGTGATCTATGCTCTAAAAGCT 1140
S  T  S  A  S  I  S  Q  P  C  S  E  G  V  I  Y  A  L  K  A

AGCTATCGCAAACACTGGGTTCAATATATATTGGAACAAAATGAGTTGGGACGCAATCCC 1200
S  Y  R  K  H  W  V  Q  Y  I  L  E  Q  N  E  L  G  R  N  P

TATAATACTACTAATGTTTTAAGAGCTATTTTATGGTTGGTGAAAGCATGGACGACAGAC 1260
Y  N  T  T  N  V  L  R  A  I  L  W  L  V  K  A  W  T  T  D

ATATCACCAGAAATTATTGAAAATGCCTTCAATCTTTCCGGTGTTCTTGGTCTATTCAAT 1320
I  S  P  E  I  I  E  N  A  F  N  L  S  G  V  L  G  L  F  N

GAATCTGCCGTTACCTCGCGTGCCTTGGATGAAATGATTCATCCACTTCGTGAGCTCGTC 1380
E  S  A  V  T  S  R  A  L  D  E  M  I  H  P  L  R  E  L  V

TCAGAGTTTTCTGTTCAAGCTGCTATGCGTATAGAGGATTTTATAAGTCCTTCAGAAGAG 1440
S  E  F  S  V  Q  A  A  M  R  I  E  D  F  I  S  P  S  E  E

AATATTGTTGATTCCTCTGAGGATATAATCAATCAAATTGCTTCGCAGTACATGGATGAT 1500
N  I  V  D  S  S  E  D  I  I  N  Q  I  A  S  Q  Y  M  D  D

CGCGCTTTCGAGACGGACGAAGAAGAATCTACTGAATTCCAGATTACCACTAAAGACGCT 1560
R  A  F  E  T  D  E  E  E  S  T  E  F  Q  I  T  T  K  D  A

ATGAAGGCGATCGAGCTTTTGCTAAACTATGAAGCTCAACAACCCGATGGAAATCCTGCA 1620
M  K  A  I  E  L  L  L  N  Y  E  A  Q  Q  P  D  G  N  P  A

ATCACGATTAGTCTATTAAATTATCAGAAGCTCCTTGAGGCAAGGGGTGGTAACGTGAAT 1680
I  T  I  S  L  L  N  Y  Q  K  L  L  E  A  R  G  G  N  V  N

CTTTCTCGTTTGAGAAGCACCTAATGTCTGTATTCGTTGTGCATATTTGACTTTTTTTGT 1740
L  S  R  L  R  S  T  Stop

TTTTTGGTTTGATTTTAATTTTCTTTTTTGTTGAGTTT
```

B

```
Abp1        MGKIKRRAITEHEK-RALRHYFFQLQNRSGQQDLIEWFREKFGKDISQPS 49
CENP B      MG-PKRRQLTFREKSRIIQEVEENPDLRKGEIARRFNIPPSTLSTILK-N 48
            **  *** .* .** * ..   .  * *.          * .

Abp1        VSQILSSKYSYLDNTVEKPWDVKRNRPPKYPLLEAALFEWQVQQGDDATL 99
CENP B      KRAILASERKYGVASTCR----KTNKLSPYDKLEGLLIAWFQQIRAAGLP 94
              **.*   *   .  .    * *.   *  **  *  *  *

abp1        SGETIKRAAAILWHKIPEYQDQPVPNFSNGWLEGFRKRH           138
CENP B      VKGIILKEKALRIAEELGMDD--FT-ASNGWLDRFRRRH           130
              * .  *.          *      *****. **.**

Abp1   (177)DIFTMQETGLFWKLVPNGTPEVEDIKGITR-FKARITLTVCCNASGTERL 226
CENP B (185)DVFSATETSLWYDFLPDQAAGLCGGDGRPRQATQRLSVLLCANADGSEKL 235
            *.*.  ** *   .*  .  .    *  *    *... .*.** *.*.*

Abp1        PLWVIGYSQSPRVFRAANVKPEVMNFKWRSNGKASMTTAIMEEWLRWFDA 276
CENP B      PPLVAGKSAKPRAGQAG------LPCDYTANSKGGVTTQALAKYLKALDT 279
            *  * * *  **  .*       .     .* *  .**  .   *.  *.

Abp1        CMEG--RKVILLIDSYTPHLRAVENIRNSGNDLRNTTVITLPSTSASISQ 324
CENP B      RMAAESRRVLLLAG-----RLAAQSLDTSG--LRHVQLAFFPPG---TVH 319
             *    *.*.**         * . . .**  **.  .   *      .

Abp1        PCSEGVIYALKASYRK                                  339
CENP B      PLERGVVQQVKGHYRQ                                  334
            *   **    *  **
```

FIG. 5. Genomic DNA sequence and deduced amino acid sequence of Abp1 and its similarity to CENP-B. (*A*) The open reading frame starts at nt 146 and ends at the termination codon at nt 1702. The amino acid sequence is shown immediately below the nucleotide sequence. Underlined amino acid sequences were derived from tryptic peptides of the isolated 60-kDa protein in the purified Abp1 fraction. Thick-underlined sequences indicate peptide sequences used for PCR. (*B*) Homology between Abp1 and the human CENP-B. Stars under amino acids indicate an exact match, whereas dots denote conserved replacements. Dashes within the coding sequence indicate spaces inserted into the sequence by the ALIGN program to achieve an optimal match.

with the 60-kDa protein present in purified fractions of Abp1. In the presence of these antibodies the Abp1–DNA complex was super-shifted in the gel mobility-shift assay (data not shown).

S. pombe ARS elements differ from budding yeast ARS elements. They are larger (<500 bp) and usually lack an obvious boundary that defines the ARS element; longer fragments tend to have greater ARS activity than shorter fragments (8–11). At present, consensus sequences have not been shown to be functionally important for ARS activity (10, 11). In view of these properties of *S. pombe* ARS elements, it is possible that ARS-specific binding protein of *S. pombe* has less stringent binding specificity than in *S. cerevisiae*, and this reduced affinity may be compensated for by redundant binding sites in ARS elements (8–11). The binding specificity of Abp1 is not stringent because it can bind MMACS multimers, poly(dA-dT) and to multiple sites in *ars3002.* Thus, the binding properties of Abp1 are consistent with those postulated for a *S. pombe* ARS-binding protein. It should be noted that the prokaryotic initiator proteins—for example, the *Escherichia coli* DnaA and λ O proteins—bind to multiple sites in their respective replication origins.

S. cerevisiae ORC is a multisubunit complex that requires ATP to bind to *S. cerevisiae* ARS elements (5). Abp1 is a single-peptide protein, and ATP had virtually no effect on its binding activity. We think Abp1 is different from ORC, although it is possible that Abp1 is a DNA-binding component of a complex formed on ARS elements. In *S. cerevisiae* it has been suggested that a protein(s) other than ORC, which recognizes the B3 or B3 and/or part of the B1 element in the origin, may also be involved in the initiation of DNA replication (3, 15). Thus, Abp1 might be a similar protein in *S. pombe.*

Abp1 has significant homology to the human and mouse centromere DNA-binding protein CENP-B. CENP-B, as a dimer, binds to a 17-bp motif (the CENP-B box) present in centromeric α-satellite DNA (14, 16, 17). The DNA-binding and dimerization domains of CENP-B are located at the N- and C-terminal regions, respectively (17, 18). The role of CENP-B in the function of centromeres is unclear, but it is thought to maintain the higher-order structure of large cen-

tromere DNA (19). To date no homologue of CENP-B has been reported in either budding or fission yeast. Preliminary results indicate that the N-terminal domain of Abp1 contains the DNA-binding domain, whereas the C-terminal region functions as a multimerization site (K. Tanka and Y.M., unpublished data). These properties suggest that there is a similarity of domain structure and function between Abp1 and CENP-B, although there appears to be little similarity in the DNA sequences required for their binding to DNA. Although *S. pombe* centromeres are much smaller than those of humans (20–30 kb vs. 200 kb), the centromeres of *S. pombe* also contain repeated sequences (20, 21). Interestingly, *S. pombe* centromeres contain multiple ARS elements (22, 23), some of which are active as chromosomal replication origins (23), although their significance in the function of centromeres is unknown. It is possible that Abp1 binds to centromere regions and plays a role similar to that of CENP-B. Perhaps, in addition, it plays a role in replication at centromeres and possibly also elsewhere in the genome.

In addition to Abp1, other proteins capable of complexing MMACS multimers were detected (Fig. 1). Formation of these complexes was blocked in competition experiments in a manner similar to that observed with Abp1. We have purified one of the binding activities, Abp2, which forms complex I with MMACS dimers and tetramers. The peptide sequence of Abp2 is distinct from that of Abp1 (J. P. Sanchez and J.H., unpublished data). Further genetic and biochemical analyses of Abp1 and Abp2 should define the functions of both these proteins in the cell.

We thank Drs. Y. Nakaseko and M. Yanagida for helpful discussions and for preparations of cosmid clones covering the entire *S. pombe* genome. We also thank Dr. Yeon-Soo Seo for suggestions in the preparation of the sequence-specific DNA-affinity column. We are indebted to David Valentin for the preparation of yeast cells in these studies and to Christine Brun and Dharoni Dubey for their gift of the plasmid containing a 582-bp stretch of *ars3002.* This work was supported by Grants 5 R37 Gm34559 (J.H.) and GM 49294 (J.A.H.). J.H. is an American Cancer Society Professor.

1. Rowley, A., Dowell, S. J. & Diffley, J. F. (1994) *Biochim. Biophys. Acta* **1217,** 239–256.
2. Theis, J. F. & Newlon, C. S. (1994) *Mol. Cell Biol.* **14,** 7652–7659.
3. Rao, H., Marahrens, Y. & Stillman, B. (1994) *Mol. Cell Biol.* **14,** 7643–7651.
4. Marahrens, Y. & Stillman, B. (1992) *Science* **255,** 817–823.
5. Bell, S. P. & Stillman, B. (1992) *Nature (London)* **357,** 128–134.
6. Burhans, W. C. & Huberman, J. A. (1994) *Science* **263,** 639–640.
7. Dubey, D. D., Zhu, J., Carlson, D. L., Sharma, K. & Huberman, J. A. (1994) *EMBO J.* **13,** 3638–3647.
8. Caddle, M. S. & Calos, M. P. (1994) *Mol. Cell Biol.* **14,** 1796–1805.
9. Wohlgenmuth, M. S., Bulboaca, G. H., Moghadam, M., Caddle, M. S. & Calos, M. P. (1994) *Mol. Biol. Cell* **5,** 839–849.
10. Zhu, J., Carlson, D. L., Dubey, D. D., Sharma, K. & Huberman, J. A. (1994) *Chromosoma* **103,** 414–422.
11. Maundrell, K., Hutchison, A. & Shall, S. (1988) *EMBO J.* **7,** 2203–2209.
12. Lusky, M., Hurwitz, J. & Seo, Y. S. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 8895–8899.
13. Mizukami, T., Chang, W. I., Grakavtsev, I., Kaplan, N., Lombardi, D., Matsumoto, S., Niwa, W., Kounosu, A., Yanagida, M., Marr, T. G. & Beach, D. (1993) *Cell* **73,** 121–132.
14. Matsumoto, H., Masukata, H., Muro, Y., Nozaki, N. & Okazaki, T. (1989) *J. Cell. Biol.* **109,** 1963–1973.
15. Rowley, A., Cocker, J. H., Harwood, J. & Diffley, J. (1995) *EMBO J.* **14,** 2631–2641.
16. Muro, Y., Masumoto, H., Yoda, K., Nozaki, N., Ohashi, M. & Okazaki, T. (1992) *J. Cell Biol.* **116,** 585–596.
17. Yoda, K., Kitagawa, K., Masumoto, Y., Muro, Y. & Okazaki, T. (1992) *J. Cell Biol.* **119,** 1413–1427.
18. Kitagawa, K., Masumoto, H., Ikeda, M. & Okazaki, T. (1995) *Mol. Cell. Biol.* **15,** 1602–1612.
19. Ikeno, M., Masumoto, H. & Okazaki, T. (1994) *Human Mol. Genet.* **3,** 1245–1257.
20. Nakaseko, Y., Adachi, Y., Funahashi, S., Niwa, O. & Yanagida, M. (1986) *EMBO J.* **5,** 1011–1021.
21. Clarke, L., Amstutz, B., Fishel, B. & Carbon, J. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 8253–8257.
22. Takahasi, K., Murakami, S., Chikashige, Y., Funabiki, H., Niwa, O. & Yanagida, M. (1992) *Mol. Biol. Cell* **3,** 819–835.
23. Smith, J. G., Caddle, M. S., Bulboaca, G. H., Wohlgemuth, J. G., Baum, M., Clarke, L. & Calos, M. P. (1995) *Mol. Cell. Biol.* **15,** 5165–5172.