MOLECULAR AND CELLULAR BIOLOGY, Sept. 1995, p. 4939–4946 Vol. 15, No. 9
0270-7306/95/$04.00+0

The Essential DNA-Binding Protein sap1 of *Schizosaccharomyces pombe* Contains Two Independent Oligomerization Interfaces That Dictate the Relative Orientation of the DNA-Binding Domain

MARJAN GHAZVINI, VÉRONIQUE RIBES, AND BENOIT ARCANGIOLI*

Oncogenic Virus Unit, URA 1644, Centre National de la Recherche Scientifique, Department of Biotechnology, Pasteur Institute, Paris, France

Received 17 April 1995/Returned for modification 8 June 1995/Accepted 27 June 1995

The *sap1* gene from *Schizosaccharomyces pombe*, which is essential for mating-type switching and for growth, encodes a sequence-specific DNA-binding protein with no homology to other known proteins. We have used a reiterative selection procedure to isolate binding sites for sap1, using a bacterially expressed protein and randomized double-strand oligonucleotides. The sap1 homodimer preferentially selects a pentameric motif, TA(A/G)CG, organized as a direct repeat and spaced by 5 nucleotides. Removal of a C-terminal dimerization domain abolishes recognition of the direct repeat and creates a new specificity for a DNA sequence containing the same pentameric motif but organized as an inverted repeat. We present evidence that the orientation of the DNA-binding domain is controlled by two independent oligomerization interfaces. The C-terminal dimerization domain allows a head-to-tail organization of the DNA-binding domains in solution, while an N-terminal domain is involved in a cooperative interaction on the DNA target between pairs of dimers.

In the fission yeast *Schizosaccharomyces pombe*, two consecutive asymmetric cell divisions are required to restrict mating-type switching of one cell among four related cousin cells (17). Genetic and molecular analysis showed that the potential and effective mating-type switching segregates chromosomally rather than through other cellular components (6, 11, 12). A DNA strand-specific imprinting event located at the mating-type locus (*mat1*) was proposed to control the formation of a double-stranded break (DSB) essential for cell type changes (11, 12). Recent work has demonstrated that a mutation in the catalytic subunit (swi7) of DNA polymerase α lowers the frequency of switching by reducing the level of DSB at *mat1* (21), thus coupling DNA replication and DSB formation.

In a search for *cis*-acting sequences required for mating-type switching, we identified at least two switch-activating sites (SAS1 and SAS2), located next to the *mat1* locus. Both sites are required for generating a normal DSB frequency and consequently for efficient mating-type switching. Pedigree analysis of SAS1 mutants revealed that this element is equally required for the two asymmetric cell divisions (3). All the *cis*- and *trans*-acting mutants analyzed in this way to date have very similar detrimental effects on both the first-switching rate and subsequent-switching rates (13). The switch-activating protein (sap1) interacts in vitro with the SAS1 element located 140 bp away from the DSB site. DNase I protection experiments showed that sap1 interacts with two DNA regions, α (4 bp) and β (12 bp), which are separated by 13 bp. Point mutations generated in either region that affect sap1 interaction in vitro reduced the efficiency of switching in vivo. These results strongly suggested that sap1 is involved in the molecular event leading to mating-type switching (3).

The *sap1* gene, isolated by a reverse genetic approach, encodes a 30-kDa protein with no obvious homology to other known DNA-binding proteins. The DNA-binding activity of sap1 protein produced in *Escherichia coli* was indistinguishable from that in *S. pombe*. Genetic experiments showed that *sap1* is essential for growth even in a strain in which mating-type switching is blocked, indicating that *sap1* function is not limited to mating-type switching and has a more general role in cell growth (1). How sap1 promotes DNA cleavage through the SAS1 site is unknown. It remains possible that bound protein(s), such as sap1, dictates a differential chromatin reorganization at the *mat1* locus, concomitantly with or after DNA replication (5, 23), restricting DNA cleavage to only one of the two sister chromatids.

Analysis of eukaryotic site-specific DNA-binding proteins indicates that they are often composed of distinct functional domains. The DNA-binding function can be carried out by protein sequences forming at least six distinct motifs: helix-turn-helix, zinc finger, leucine zipper, β-ribbon, high-mobility group (HMG), and helix-loop-helix. sap1 does not appear to belong to any of the groups outlined above. Random point mutation analysis has been used to characterize sap1 domains responsible for DNA binding. Three clusters of point mutations (domains I, II, and III, from residue 30 to 126), which affected specific SAS1 recognition, define the DNA-binding motif. Furthermore, it was shown that sap1 exists as a dimer in solution as well as when bound to the SAS1 element. A dimerization domain (domain IV, from residue 136 to 203) was found in the carboxy-proximal region of the DNA-binding domain, and its presence is essential for SAS1 binding. This domain contains a heptad repeat of hydrophobic amino acids, which potentially forms a coiled-coil structure. The observation that this region remains dimeric with an NH_2-terminal disulfide linkage suggests that the helices are parallel. Finally, removal of the first 22 residues of sap1 (domain V, from residue 1 to 22) does not influence SAS1 recognition unless the dimerization domain of the protein is also partially removed (2).

* Corresponding author. Mailing address: Oncogenic Virus Unit, Institut Pasteur, 25 rue du Dr. Roux, 75724 Paris Cedex 15, France. Phone: (1) 45 68 84 54. Fax: (1) 45 68 87 90. Electronic mail address: barcan@pasteur.fr.

To further define the sap1 DNA-binding sequences, we have used a reiterative selection procedure to demonstrate that sap1 homodimers preferentially select direct repeat arrangement of a core binding motif, TA(A/G)CG. We propose that preferential recognition of direct repeat organization requires the sap1 dimerization domain located C terminally to the DNA-binding domain. A serendipitous observation of this study was that the removal of this dimerization domain revealed the presence of a second, independent dimerization interface consisting of the first 22 residues of the protein. The truncated C-terminal form of sap1 failed to recognize the direct repeat arrangement, concomitantly with the appearance of a novel DNA-binding specificity for a sequence containing the same core recognition motif organized as an inverted repeat. We demonstrate that, in the native protein, the carboxy-terminal domain promotes dimerization of sap1 in solution while the amino-terminal domain favors cooperative interaction between pairs of dimers.

MATERIALS AND METHODS

sap1-binding site selection. We have used a reiterative selection procedure to isolate specific binding sites for sap1, by means of bacterially expressed protein His-sap1 (2) and randomized double-strand oligonucleotides and primers 5′ TTCTCACGTCGCCTCGAGN_{29}CTCGAGGACGCAGAGCCT (where *N* is a random nucleotide), 5′ AGGCTCTGCGTCCTCGAG, and 5′ TTCTCACGTC GCCTCGAG. Each oligonucleotide was gel purified prior to use. The random template was first radiolabelled and made double stranded by using *E. coli* Klenow fragment and one of two primers in the presence of the four deoxynucleoside triphosphates (dNTPs) (1 mM), [α-^{32}P]dATP, and [α-^{32}P]dGTP, and then purified on a nondenaturing 5% polyacrylamide gel. PCR amplifications were performed by using an amplification kit (Perkin-Elmer Cetus) with 30-s denaturation at 94°C, 1-min annealing at 55°C, and 1-min extension at 72°C for 15 cycles for the eluted DNA of the first round of selection, 10 cycles for those of the second round, and 7 cycles for those of the subsequent rounds. The amplified oligonucleotides were extracted by phenol-chloroform, precipitated by ethanol, and purified from a 5% polyacrylamide gel. The eluted oligonucleotides were precipitated and dissolved in 100 μl of buffer A (150 mM KCl) before being incubated with a bacterially synthesized chimeric protein consisting of six histidine residues fused to the N terminus of sap1 (6×His-sap1) (2) bound to nickel-nitrilotriacetic acid beads. The enriched DNAs after six cycles of selection-amplification were cloned within the *Sal*1 unique site of pUC19. Plasmid DNAs from 40 isolated colonies were digested by *Hin*dIII and *Bam*HI, labelled by *E. coli* DNA polymerase Klenow and [α-^{32}P]dATP, gel purified, and analyzed by gel mobility shift assay. Seventeen sites of high affinity and 23 sites of lower affinity were chosen and sequenced by using the primer −40 (17-mer) according to the double-stranded DNA sequencing protocol of the manufacturer (U.S. Biochemical, version 2.0). The alignment of these sequenced DNAs was performed by the computer program Multialign.

DNA-binding assays. Reactions for gel mobility shift assays (7, 8) were performed essentially as described elsewhere (4) with several modifications. Binding reactions were carried out in a 10-μl volume containing 10 mM HEPES (*N*-2-hydroxyethylpiperazine-N′-2-ethanesulfonic acid) (pH 8), 100 mM NaCl, 7 mM β-mercaptoethanol, 3% glycerol, 0.05% Triton, 0.5 ng of a radiolabelled DNA probe, and either 20 μg of protein crude extract from *S. pombe* or 20 ng of bacterial protein extract in the presence of 1 μg or 50 ng of sonicated salmon sperm DNA as nonspecific competitor DNA, respectively. The reaction mixes were incubated at room temperature for 5 min and electrophoresed through Tris-borate-EDTA–5% polyacrylamide gels at 120 V for 2 h 30 min. Gels were then dried and autoradiographed. For the cooperative binding activity, the radiolabelled synthetic oligonucleotides DR and (DR)×(IR) were incubated for 5 min at room temperature with an increasing amount of bacterium- or yeast-expressed sap1 prior to being loaded on a polyacrylamide gel. The protein-DNA complexes in the polyacrylamide gels were then measured by PhosphorImager analysis (Molecular Dynamics).

Synthetic oligonucleotides. The DNA sequences of synthesized wild-type SAS1 and (DR)×(IR) are as follows: 5′ tcgaCTCTAACGAGATATTTGCTTC GCTACGCTACGCAC and 5′ tcgaGCTCGCTACACCACGCTATACCGTTT ACGATGCCTAACGCG. All of the synthetic oligonucleotides, including DR1 to -3, IR1 and -2, and (DR)×(IR), are flanked at both ends by partial cohesive *Xho*I sites.

Methylation interference footprinting. Radiolabelled DNAs were treated with dimethyl sulfate as described elsewhere (9). Premodified DNAs were incubated with extracts and reaction mixes were electrophoresed as described above. Free and complexed DNAs were visualized by autoradiography of the wet gel. Labelled DNAs were excised and incubated at 37°C in 10 mM Tris-HCl (pH 8.0)–1 mM EDTA–0.3 mM NaCl–0.2% sodium dodecyl sulfate (SDS). The eluted DNAs were purified by phenol extraction and ethanol precipitation. The recovered DNAs were resuspended in 100 μl of 10% piperidine and incubated for 30 min at 90°C. The samples were lyophylized four times after being resuspended each time in 100 μl of H_2O. The pellet DNAs were resuspended in denaturing buffer, heated for 3 min at 94°C, and electrophoresed on an 8% polyacrylamide gel containing 8 M urea and 1× Tris-borate-EDTA in order to determine the positions of cleavage.

RESULTS

Isolation of DNA sequences that bind selectively to the sap1 protein. To identify specific DNA sequences that are recognized by the sap1 protein, we used a bacterially synthesized chimeric protein consisting of six histidine residues fused to the N terminus of sap1 (2). The 6×His tag allows the fusion protein to bind to nickel-nitrilotriacetic acid beads and facilitates its purification. To demonstrate the functional integrity of sap1 bound to the beads and the feasibility of selecting binding sites from random DNA, double-stranded oligonucleotides coding for SAS1 wild-type or mutated sequences were loaded on the beads in the presence of a DNA carrier and eluted with increasing KCl concentrations. The SAS1 mutated oligonucleotides were eluted from the beads at 200 mM, whereas the wild-type element was eluted from the beads at 600 mM (data not shown).

Next, random-sequence double-stranded oligonucleotides were incubated with the 6×His-sap1 protein adsorbed to the beads as described in Materials and Methods. Nonspecific oligonucleotides were eliminated by washing the beads with 200 mM KCl buffer, and specifically bound oligonucleotides were eluted with 600 mM KCl buffer. This fraction, representing about 0.5% of the input DNA, was amplified by PCR and subjected to a second selection cycle. After six rounds of selection-amplification, we enriched the oligonucleotides with a high affinity for sap1 from 0.5 to >50% (data not shown). These affinity-selected oligonucleotides were cloned within the *Sal*1 site of pUC19, and 40 isolated clones were further studied.

sap1 preferentially selects direct repeat sequences. We first analyzed the binding activities of the affinity-selected oligonucleotides by band shift electrophoresis using either the bacterially expressed 6×His-sap1 protein or a crude protein extract from *S. pombe* (data not shown). Among the 40 target DNAs analyzed, 17 oligonucleotides formed high-affinity complexes, similar to those observed with the SAS1 element, with both protein extracts containing sap1, while the other oligonucleotides formed weaker complexes. The sequences of the 17 high-affinity-selected oligonucleotides aligned for maximum homology are shown in Fig. 1A. It is clear from this analysis that nearly all the oligonucleotides share two related TAACG motifs, organized as direct repeats and spaced by 5 bp. The first motif (motif I) contains the sequence TA(A/G)CG, whereas the second motif (motif II) is composed of the sequence (C/T)AACG. In addition, a G was frequently found 2 nucleotides 3′ from each motif (positions 8 and 18 in Fig. 1A). Furthermore, the first motif was often preceded by A (position 1 in Fig. 1A), whereas no bias was observed at this position for the second repeat. Similarly, a T was favored at position 10 between the two repeats (Fig. 1A). By visual inspection, it was apparent that the selected oligonucleotide sequences are biased for purines and pyrimidines on the 5′ and 3′ sides, respectively. Figure 1B shows the sequences of the remaining 23 oligonucleotides with weaker affinities for sap1 protein. Nearly all of the clones shared the same core recognition sequence, TA(A/G)CG. However, this motif was found either alone or in combination with a second similar sequence in several different arrangements. Interestingly, the selected motif has higher se-

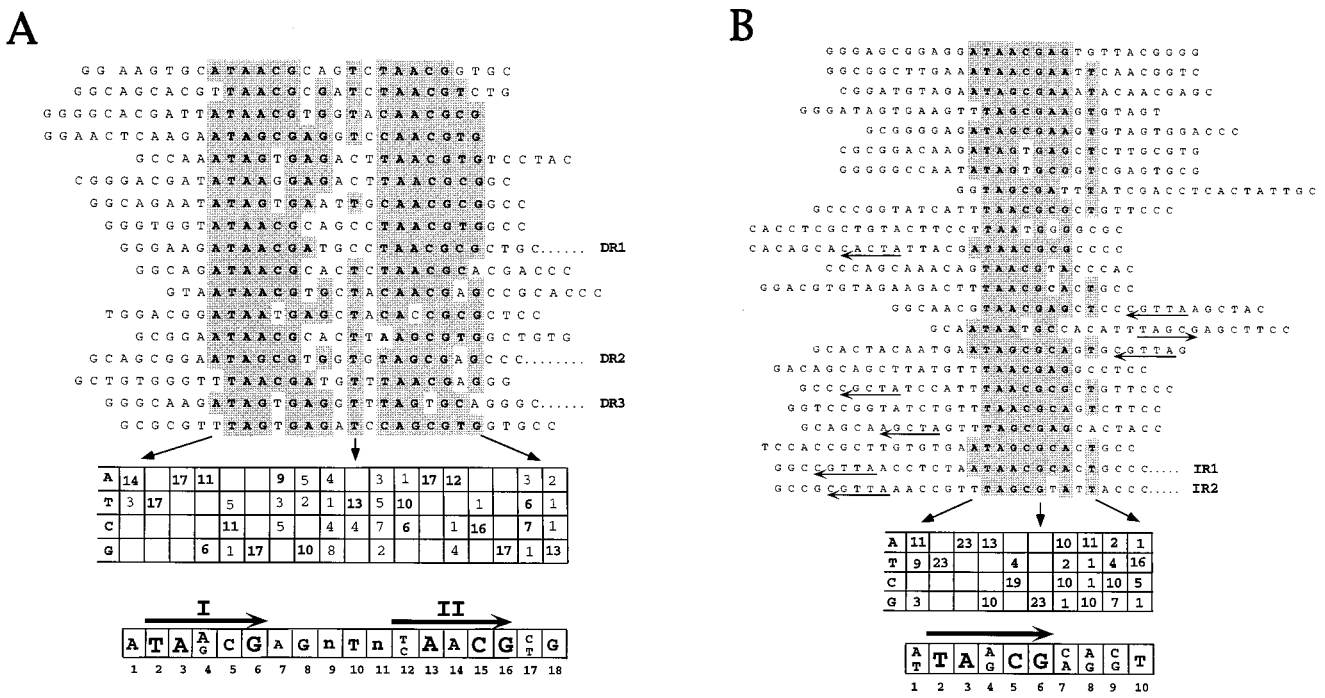

FIG. 1. sap1-selected DNA sequences. (A) DNA sequences of 17 high-affinity clones. The initial target oligonucleotide contains flanking PCR primer sites which bracket a central core of 29 degenerated bases. This oligonucleotide was converted to double-strand DNA and subjected to six rounds of selection and amplification. The randomized and selected sequences of 17 clones are aligned, and conserved nucleotides are shaded. Oligonucleotides DR1 to -3 are indicated. The number of times that each base occurs at a position and the resulting consensus are shown below. The consensus sequence can be defined as containing two half-sites (I and II), organized as direct repeats and spaced by 5 bp. (B) DNA sequences of 23 medium-affinity clones. Thin horizontal arrows, potential core recognition motifs. Oligonucleotides IR1 and IR2 are indicated.

quence homology with motif I of the direct repeat. In addition, it seems that a T was favored in the 3′ position (compare positions 10 in Fig. 1A and B) of the core sequence.

Importantly, the initial SAS1 sequence contains three core recognition motifs, one on the upper strand and two on the lower strand (called a, b, and c, respectively), organized as direct repeats. Motif a overlapped with the SAS1 α subregion, whereas motifs b and c overlapped with the SAS1 β subregion (see the introduction and reference 3).

We have previously shown that sap1 exists as a dimer both when free in solution and when bound to the SAS1 sequence (2). Similarly, sap1 formed a dimer with the DR2 sequence (data not shown).

To further analyze protein-DNA interactions between the sap1 dimer and direct repeat oligonucleotides, methylation interference experiments were performed (see Materials and Methods). The patterns observed for sap1 with three randomly chosen direct repeat oligonucleotides (DR1 to -3 in Fig. 1A) are shown in Fig. 2. Significant interference was observed only on the sense strand with guanine residues within the core recognition sequence TA(A/G)CG (no methylation interference on the antisense strand was observed [data not shown]). In addition, methylation within repeat I abolished totally sap1 binding, while methylation within repeat II allowed residual binding. These results indicate that sap1 interacts mainly with the major groove of the core recognition sequence and that the two repeats are not equally important for sap1 binding; also, we cannot exclude minor-groove DNA interaction. Furthermore, since the two direct repeats are spaced by 5 bp, sap1 dimers contact their DNA targets predominantly from one side of the double helix.

In order to confirm the preferential role of repeat I, we

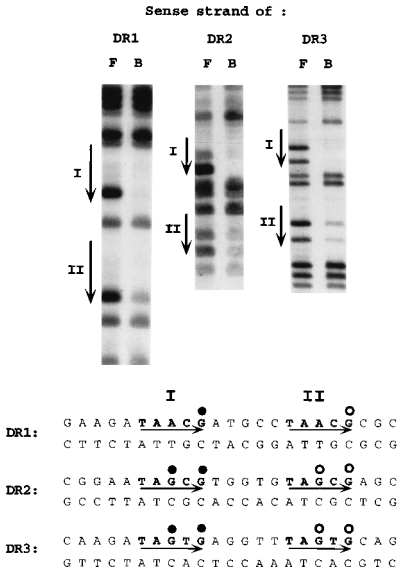

FIG. 2. Methylation of guanine residues in the direct repeat motif interferes with full-length sap1(1-254) binding. Oligonucleotides DR1 to -3 containing direct repeats (Fig. 1) were labelled on the sense strand and methylated with dimethyl sulfate. The methylated oligonucleotides were incubated with sap1(1-254), and free (F) and bound (B) labelled DNAs were separated on native gels, eluted, cleaved with piperidine, and analyzed on denaturing polyacrylamide gels. The positions of the two direct repeats (I and II) (arrows) and positions of complete interference (solid circles) and partial interference (open circles) caused by N-7 methylation of guanine are indicated.

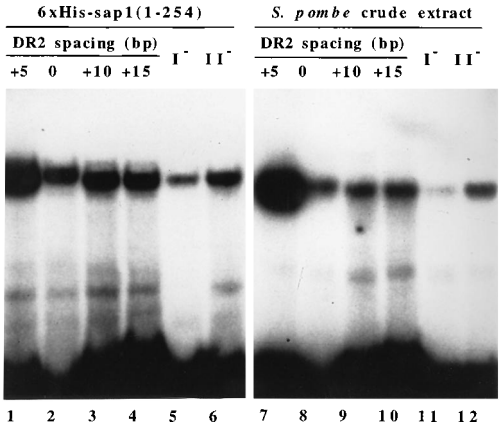

FIG. 3. Electrophoretic mobility shift analysis of sap1 proteins from *E. coli* and *S. pombe* showing that DNA-binding activity depends on the space between two direct repeats and their sequence integrity. The oligonucleotide used as reference was DR2, which naturally contains two TAGCG sequences spaced by 5 bp (lanes 1 and 7). The numbers of nucleotides separating the repeat motifs are indicated, as are motif I and II mutations (−). We used the partially purified 6×His–sap1(1-254) produced in *E. coli* (lanes 1 to 6) and the S100 extract from *S. pombe* containing full-length sap1 (lanes 7 to 12).

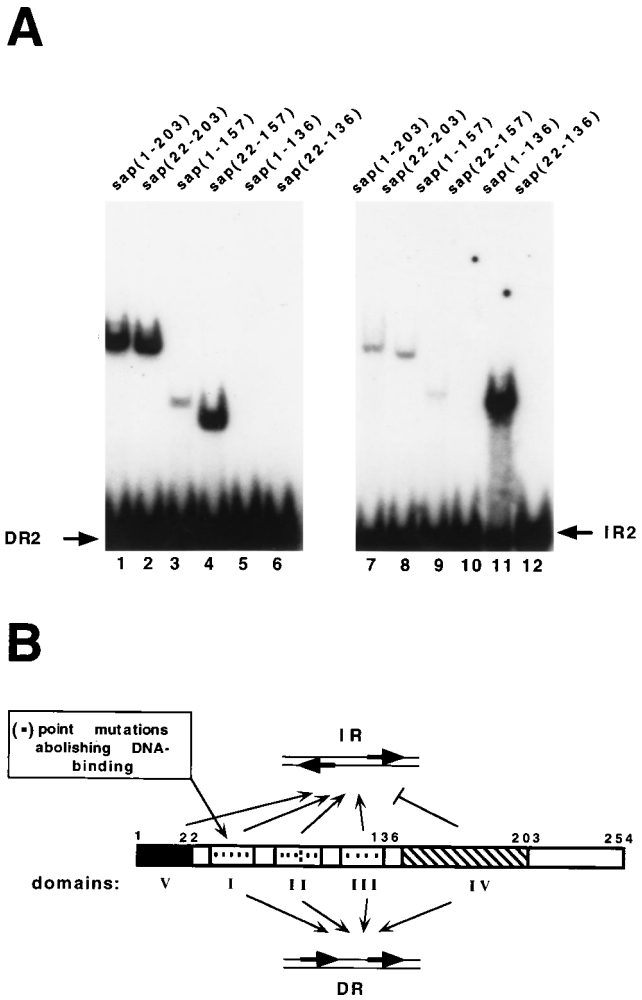

FIG. 4. sap1 protein domains required for direct repeat (DR) and inverted repeat (IR) recognition. (A) Partially purified *E. coli* protein extract containing truncated sap1 fused to a 6×His tail at the N terminus was prepared, and the sap1 concentration was estimated by SDS-polyacrylamide gel electrophoresis followed by Coomassie blue staining (data not shown). Electrophoretic mobility shift assays were carried out with equal amounts of sap1 polypeptides and with the DR2 (lanes 1 to 6) or IR2 (lanes 7 to 12) double-strand, end-labelled oligonucleotide. The amino acids retained are indicated in parentheses. (B) sap1 domain properties. sap1 domains I to V (2) are indicated. Thin arrows, sap1 domains required for specific binding.

replaced the first adenosine residue of the core recognition motif TAACG (conserved among the 40 selected oligonucleotides) in each repeat with a cytosine. Figure 3 shows that such an A-to-C mutation reduced sap1-binding activity for both mutated oligonucleotides. The point mutation in repeat I affects sap1 interaction more severely than does the point mutation in repeat II (compare lane 5 with lane 6 and lane 11 with lane 12), in agreement with the methylation interference results. We further investigated the contribution of the spacing between the repeats to sap1 binding. Figure 3 shows that *S. pombe* sap1 and bacterial sap1 formed complexes with weaker efficiencies when the direct repeats were separated by 0, 10, and 15 bp than with DR2, which contained an initial spacing of 5 bp (compare lane 1 with lanes 2 to 4 and lane 7 with lanes 8 to 10). Note that changing the space between direct repeats (even by 15 bp) decreased sap1-binding activity to a level close to that observed with the oligonucleotide containing a mutation in motif II (lanes 6 and 12). Taken together, these results suggest that the sap1 dimer is still capable of interacting with DNA when it contains only one core recognition motif (perhaps motif I), in agreement with the selection by sap1 of the low-affinity-selected oligonucleotides harboring a single motif.

Removal of the sap1 C-terminal dimerization domain switches its DNA-binding specificity. The relative orientation and spacing of the core recognition motifs seem to play essential roles in determining the specificity of sap1 DNA recognition. We have previously shown that a sap1 protein [sap(1-136)] in which the C-terminal dimerization domain had been deleted failed to bind the SAS1 sequence (2). To further investigate the contributions of the different domains of sap1 to the selectivity of DNA recognition, we analyzed the DNA-binding activities of various sap1 truncated forms with the DR2 and IR2 oligonucleotides (Fig. 4). When DR2 was used as a probe, an intermediate deletion in the dimerization domain [sap1(1-157)] strongly decreased DNA binding (compare lanes 1 and 3) while complete deletion of domain IV [sap1(1-136)] abolished DNA binding (lane 5). Removal of the 22 N-terminal residues (domain V) did not change the DNA-binding activity (lanes 1 and 2) unless domain IV was partially deleted, whereas DR binding activity was greatly enhanced, to the level observed with full-length sap1 (compare lanes 1 and 2 with 3 and 4). Similar results were obtained with the other high-affinity-selected oligonucleotides (data not shown). However, the opposite DNA-binding activity was found for the sap1 truncated forms when the IR2 oligonucleotide was used as a probe. Surprisingly, while the sap1 C-terminal dimerization domain was essential for direct repeat binding, its presence inhibited IR2 recognition and its removal [sap1(1-136)] resulted in formation of a high-affinity complex with the IR2 oligonucleotide. This specific interaction was weaker with IR1 (Fig. 1B) and absent with all the other selected oligonucleotides (data not shown). In addition, deletion of the 22 N-terminal residues, which did not change or increase DNA binding to direct repeats, abolished binding to the inverted repeat sequence (Fig. 4A; compare lanes 11 and 12).

Figure 4B summarizes the data obtained. The DNA-binding motif (domains I to III) was defined earlier. Domain IV is responsible for switching the DNA-binding specificity of sap1. This protein domain, able to dimerize in solution, is essential

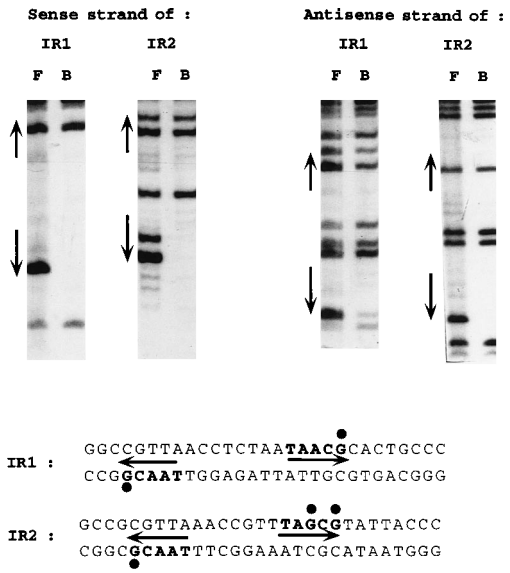

FIG. 5. Methylation of guanine residues in the inverted repeat motif interferes with sap1(1-136) binding. Oligonucleotides IR1 and IR2 containing inverted repeats (Fig. 1) were labelled on the sense or antisense strand and methylated with dimethyl sulfate. The methylated oligonucleotides were incubated with sap1(1-136), and free (F) and bound (B) labelled DNAs were separated on native gels, eluted, cleaved with piperidine, and analyzed on denaturing polyacrylamide gels. Arrows, inverted repeats; circles, positions of complete interference caused by N-7 methylation of guanine.

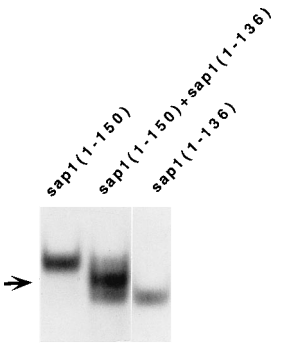

FIG. 6. Electrophoretic mobility shift assay for dimer formation. Partially purified sap1(1-136) and sap1(1-150) were preincubated alone or mixed for 10 min before incubation with the ^{32}P-labelled DR2 oligonucleotide prior to analysis by mobility shift assay. Arrow, sap1(1-136)/sap1(1-150) heterodimer–IR2 complex.

for SAS1 and direct repeat recognition and prevents binding to the IR2 sequence. Removal of domain IV abolished binding to SAS1 and the direct repeat, concomitantly with the emergence of a new DNA-binding specificity for a sequence containing the core recognition motif organized as an inverted repeat. Finally, domain V is essential for inverted repeat recognition.

The N-terminal part of sap1 binds to the inverted repeat motif of the core recognition motif. In order to further define sap1(1-136)–DNA interactions, the methylation interference patterns were determined for the sap1(1-136) complex with IR1 and IR2 (Fig. 5). Significant interference with the guanine residues within the two core recognition motifs was observed on both DNA strands. In contrast to the asymmetric interference patterns on the direct repeat motifs, the degree of interference on the inverted repeat motif is strong and the levels of interference are similar for the two repeated elements, indicating that sap1(1-136) interacts similarly with both motifs, possibly as a dimer. These data indicate that the sap1(1-136) polypeptide, which is monomeric in solution (2), may retain a potential symmetric dimerization interface mediating cooperative interactions when binding to core recognition motifs arrayed in an inverted repeat and spaced by 7 or 8 bp. We also noticed that at least three other oligonucleotides in addition to IR1 and -2 contain two motifs organized as an inverted repeat but with a different spacing (Fig. 1B). However, these oligonucleotides showed very weak or no binding to the sap1(1-136) polypeptide (data not shown), suggesting that spacing is an important feature for inverted repeat recognition.

Domain V is involved in a second dimerization interface. In an attempt to determine the multimerization state of sap1(1-136) bound to inverted repeat sequences and to localize the potential oligomerization interface, we used the strategy described by Hope and Struhl (10). In this experiment, the short polypeptide consisted of sap1(1-136), whereas the long polypeptide consisted of sap1(1-150). Figure 6 shows that both polypeptides interact with the IR2 oligonucleotide, forming complexes with different mobilities. To obtain similar binding activities with both polypeptides, we used 10 times more sap1(1-150) than sap1(1-136). When the two polypeptides were incubated together prior to binding, one additional complex, with an intermediate mobility, was observed. This result supports the conclusion that sap1 forms a dimer when bound to inverted repeat motifs.

To determine whether this potential dimerization interface is functional, not only in the context of the truncated form sap1(1-136) but also in full-length sap1 protein, we reexamined the DNA-binding properties of sap1(1-254) and sap1(22-254), in which the 22 first residues had been removed. First, we measured the dissociation constant by titrating a fixed amount of the DR2 oligonucleotide with increasing amounts of either sap1(1-254) or sap1(22-254). For three different titrations, the concentrations at half saturation were similar for the two proteins and estimated to be 4×10^{-10} M on average (data not shown).

Next, we synthesized an oligonucleotide [(DR)×(IR)] whose sequence includes the features of both the DR2 oligonucleotide and the IR2 oligonucleotide, as it contains two direct repeats in an inverted arrangement (see Fig. 7C for the sequence and Fig. 8). Increasing concentrations of sap1 protein, from either *S. pombe* or *E. coli*, were incubated with this oligonucleotide and analyzed by gel retardation assay.

At a low concentration of full-length sap1 protein isolated from *E. coli* or *S. pombe*, two complexes were apparent (Fig. 7A). The lowest complex (dimeric [2:4]) migrated at the same position as the complex formed with the DR2 oligonucleotide (data not shown) and probably corresponds to a sap1 dimer bound to one of the two direct repeat motifs. The highest complex (tetrameric [4:4]), which is also apparent at a low sap1 concentration, might be a tetramer of sap1 bound to the two direct repeat motifs, as suggested by dimethyl sulfate interference analysis (data not shown). A third complex, which has an intermediate mobility, appears only with the bacterially expressed full-length protein and not with the protein extract prepared from *S. pombe*. This suggests that some of the sap1 protein present in the *E. coli* extract can exist as a monomer in solution and may allow the formation of a trimeric form of sap1 on the probe. To determine the cooperativity parameters, with the assumption that association constants for both direct repeats are equal, we quantitated with a PhosphorImager the labelled species detected in band shift electrophoresis. As shown in Fig. 7B, the fraction of the oligonucleotide bound by

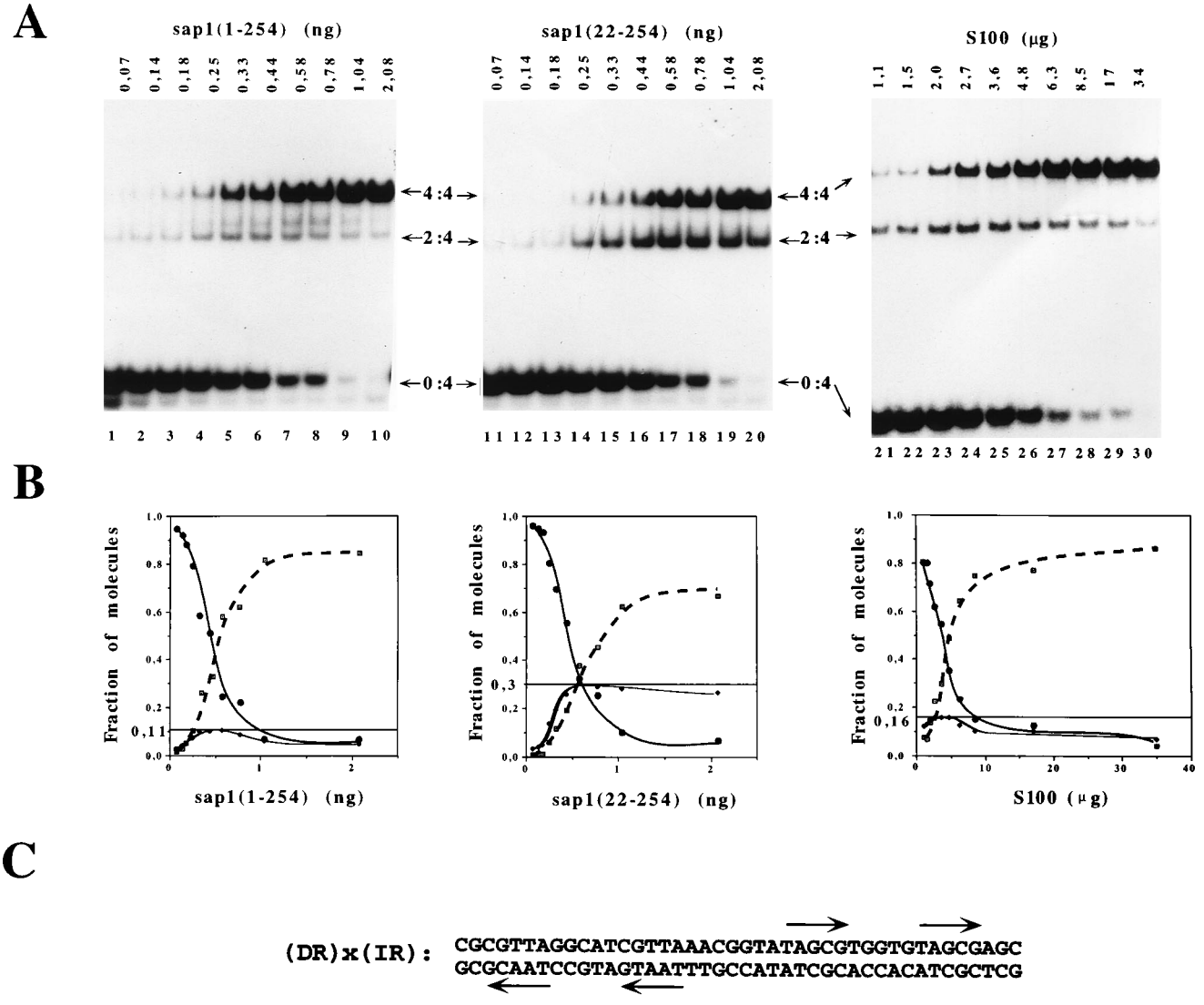

FIG. 7. Titration of the labelled oligonucleotide containing two direct repeats in a palindromic arrangement with the full-length sap1 protein and the N-terminally truncated sap1(22-254) polypeptides. (A) A fixed amount of a labelled 46-bp oligonucleotide (0.5 ng) was titrated in a 10-μl volume with increasing protein concentrations. Bacterium- and *S. pombe* (S100)-expressed sap1 proteins were quantitated by Coomassie blue staining and Bradford assay, respectively. The labelled species are indicated as follows: 0:4, free probe; 2:4, complex with one dimer bound; 4:4, complex with two dimers bound. The intermediate-mobility complex with sap1(1-254) protein never exceeded 5% of the total product. (B) Quantitative PhosphorImager analysis of the gel shifts from panel A. ●, free probe; ◆, 2:4 complex; ⊡, 4:4 complex. Maximum fractions of complexes containing one bound dimer (2:4): 0.11, 0.3, and 0.16 for sap1(1-254), sap1(22-254), and sap1 from *S. pombe*, respectively. The cooperativity parameter (k_{coop}) was calculated as $([1/(2{:}4_{max})]^{-1})^2$, with the assumption that the dissociation constants for the two direct repeats are equal. k_{coop}s were 65.5, 5.3, and 27.5 for sap1(1-254), sap1(22-254), and sap1 from *S. pombe*, respectively. Similar analyses have been performed for Oct-2 and steroid hormone receptors (15, 22). (C) DNA sequence of the (DR)×(IR) oligonucleotide containing two direct repeats in an inverted arrangement. Arrows, core recognition motifs.

sap1 dimers (2:4) increases first and reaches maximum values of 0.11 and 0.16 for the full-length sap1 proteins from *E. coli* and *S. pombe* extracts, respectively. The slower-mobility tetrameric complex then accumulates, approaching a value of 1.0, with a decrease in the concentration of the dimeric complexes. By using these data, we determined that the relative association constants of the tetrameric forms of sap1 are 65- and 27-fold higher than that of the dimeric form of sap1 for the bacterial and yeast proteins, respectively (see reference 21 for the method of calculation). The higher cooperativity parameter value obtained with sap1 isolated from *E. coli* may be explained by the presence of the intermediate-mobility complex; also, we cannot exclude the possibility that the six histidines fused at the N terminus of the bacterially expressed protein stabilized the two dimers on the DNA or that the sap1 protein present in the yeast extract is not fully active.

In order to determine whether the first 22 residues of sap1 influence the cooperativity parameter, a similar binding study was performed with the sap1(22-254) polypeptide and the (DR)×(IR) oligonucleotide as probe. When increasing concentrations of sap1(22-254) were incubated with this oligonucleotide, only two sap1-DNA complexes, corresponding to dimeric (2:4) and tetrameric (4:4) bound sap1 complexes, were formed (Fig. 7A). The fraction of the oligonucleotide bound by the sap1(22-254) dimer (2:4) increases first and reaches a maximum value of 0.3 (Fig. 7B). The calculated value of the cooperativity parameter is 5, which is 10-fold lower than the value obtained for the full-length bacterial protein. Taken together,

these data show that removal of the first 22 residues of sap1 strongly reduces the cooperativity parameter and demonstrate that the N-terminal part of sap1 is involved in a second oligomerization interface responsible for the cooperativity of binding of two dimers. However, the residual cooperativity found for sap1(22-254) suggests that additional protein domains may still contribute to the formation of the tetrameric complex.

DISCUSSION

sap1 consensus recognition motif. In this study we have demonstrated that the *S. pombe* sap1 protein, which participates in mating-type switching control, interacts preferentially, as a homodimer, with a direct repeat of the core sequence, TAACG, spaced by 5 bp. In addition, binding studies showed that motif I interacts more avidly with sap1 than does motif II (at least for three oligonucleotides randomly chosen). These data demonstrate that the two sap1 subunits interact differently with the DNA, reflecting an asymmetric protein-protein interaction underlying the molecular basis for specific binding of the sap1 homodimer to the tandem repeat of the core recognition motif. The orientation and the restricted spacing of the selected DNA targets appear to arise from the organization of the DNA-binding domains.

Other binding sites for sap1 in the *S. pombe* genome were anticipated, since it was shown that sap1 is essential for growth, in addition to its role in mating-type switching. However, we were unable to unambiguously define multimeric sap1-binding sites, by computer analysis, among the genomic sequences from *S. pombe*. It will be of interest to identify these sites and to analyze their functions relative to cell growth.

Determinants of sap1 DNA-binding domain orientation. For most DNA-binding proteins, oligomerization is an essential part of the binding interaction and its regulation. We have previously demonstrated that domain IV (residues 136 to 203) contains a region, from residue 143 to 173, which displays a strong potential to form an amphipathic α-helical structure. Furthermore, we observed that domain IV alone can dimerize and form an NH_2-terminal disulfide linkage (involving Cys-140 or Cys-142), indicating that the helices are parallel. This structure should naturally organize the DNA-binding domains (residues 23 to 136) of the sap1 dimer in a head-to-head and not a head-to-tail configuration, as anticipated from the preferential interaction with direct repeat sequences. It is possible that another protein motif, probably located in the DNA-binding domain, contributes to asymmetric protein interaction, leading to the preferential binding of sap1 to direct repeats.

In this report, we provide evidence for the existence of a second dimerization interface which involves the first 22 residues of sap1 (domain V). Domain V is located at the N terminus of sap1 protein and is crucial in establishing the orientation and spacing preferences for cognate inverted repeat sequences. Our data suggested that the two domains compete for opposite orientations of the DNA-binding domains, but, since domain IV allows the dimerization of sap1 in solution, the direct repeat arrangement of the core recognition motif is preferred for DNA binding.

We propose a model (Fig. 8) in which sap1 protein domains I to III fold independently of domains IV and V and can interact with the core recognition motif, TAACG. Domain IV, which promotes dimerization of sap1 in solution, may position the DNA-binding motif in such a way that it can recognize the two half-sites only when they are arrayed as direct repeats but not when they are inverted. Removal of domain IV, which renders the protein monomeric in solution, allowed domain V

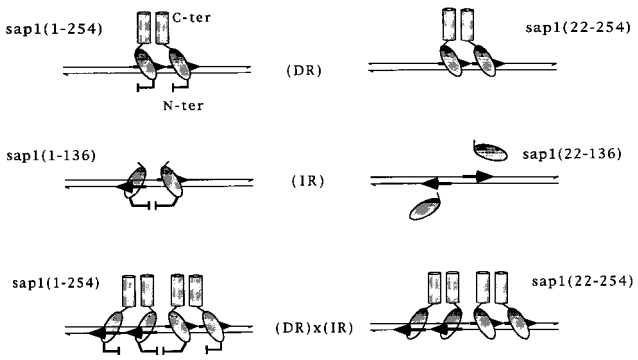

FIG. 8. sap1 dimerization interfaces required for DNA-binding recognition. Three oligonucleotides, containing a direct repeat (DR), an inverted repeat (IR), or two direct repeats in a palindromic arrangement (DR)×(IR), are shown. The C terminus (domain IV) mediates dimerization in solution and leads to head-to-tail arrangement of the DNA-binding domain, allowing recognition of the direct repeat motif, independently of the presence of the N-terminal part of the protein (domain V). Removal of the dimerization domain leading to a monomeric polypeptide in solution allows recognition of the inverted repeat motif. In contrast to direct repeat recognition, stable binding to the inverted repeat required the presence of domain V, which probably provides a dimerization interface between monomers when they are bound to the inverted repeat. This model is consistent with our data for the (DR)×(IR) oligonucleotide, with which the binding of the first sap1 dimer to one direct repeat enhances the binding of a second sap1 dimer to the other direct repeat in a cooperative manner, and the observation that removal of domain V leads to an important loss of the DNA-binding cooperativity between sap1 dimers.

to reorganize the two DNA-binding domains head-to-head on the inverted repeat sequence. However, when the protein is intact, domain V provides a dimerization interface promoting cooperative binding of two dimers with an oligonucleotide containing a palindromic sequence of two direct repeat elements. The cooperativity might result from direct protein-protein interactions or be the product of an altered DNA conformation induced by sap1 binding. Both mechanisms depend on the presence of the 22 N-terminal residues and are likely to be influenced by the spacing between the two direct repeat sequences.

Although the sap1 protein and its target DNA diverge from other DNA-binding proteins, it is worth noting that sap1 DNA-binding properties are, to some extent, reminiscent of those of other nuclear factors. Recent studies of the yeast activator HAP1 indicate that this factor binds DNA in a direct repeat orientation (24). Similarly, the DNA-binding properties of the nuclear-receptor family (retinoid X receptor, thyroid hormone receptor, retinoic acid receptor, and vitamin D_3 receptor) have revealed that the relative orientation and spacing of the core recognition sequence play essential roles in the specificity of DNA binding. The retinoid X receptor promotes binding at direct repeats spaced by 1 to 5 nucleotides by forming homodimers as well as heterodimers with different members of the nuclear-receptor family. Homodimerization and heterodimerization, in solution, are dependent on a carboxy-terminal motif present within the ligand-binding domain of each receptor, whereas the orientation and the spacing of DNA-binding elements are dictated primarily by a protein motif present within or next to the DNA-binding domain (14, 16, 18, 19, and references therein). The functional relevance of the intrinsic asymmetry created by either factor for binding to direct repeat DNA is unknown. The mechanism of sap1 recognition of direct repeats is likely to be different from the other factors, since sap1 does not contain zinc finger DNA-binding motifs but harbors a new type of DNA recognition domain.

The importance of sap1 DNA-binding cooperativity. Since sap1 may be directly involved in the molecular mechanism which controls the asymmetric division of the fission yeast, it is tempting to speculate on the functional consequences of sap1 DNA-binding cooperativity. We wish to propose a model for the asymmetric distribution of the DSB at the *mat1* locus. The sap1 protein-DNA complexes are likely to be disrupted by the replication fork. Considering that sap1 nuclear concentration is low and tightly regulated, this opens a window in which the sap1 protein molecules can reassemble cooperatively on only one of the two new sister DNAs. This asymmetric distribution possibly reorganizes the two chromatids differently, producing two nonequivalent sister chromosomes soon after replication. The sap1-bound chromatin becomes accessible for DNA cleavage, while the unbound chromatin is not. The short delay between leading-strand replication and lagging-strand replication may also be an important parameter. Conceivably, the DNA strand first replicated will be recurrently the one cooperatively bound by sap1, for every round of DNA replication. According to this scheme, the DNA replication of the *mat1* locus must be specifically organized in order to keep the same strand always the one first replicated. Although this simple model does not explain how the DSB at *mat1* is formed, the asymmetric distribution of the sap1 protein-DNA complex can explain the strand segregation model, involving a form of chromosomal imprinting, proposed by Klar (11, 12). A simple prediction of this model is that overexpression of *sap1* may compromise the asymmetric distribution of the protein. However, preliminary experiments indicate that overexpression of the sap1 full-length or truncated form is toxic for growth (20). Further development of our working model awaits construction and analysis of inducible or conditional *sap1* mutants.

ACKNOWLEDGMENTS

We are grateful to G. Langsley, S. Whiteside, and M. Yaniv for critical reading of the manuscript.

M.G. was supported by a fellowship from the Ministère de la Recherche et des Techniques. This work was supported by grants from the Association pour la Recherche sur le Cancer and the Centre National de la Recherche Scientifique.

REFERENCES

1. **Arcangioli, B., T. D. Copeland, and A. J. S. Klar.** 1994. Sap1, a protein that binds to sequences required for mating-type switching, is essential for viability in *Schizosaccharomyces pombe*. Mol. Cell. Biol. **14:**2058–2065.
2. **Arcangioli, B., M. Ghazvini, and V. Ribes.** 1994. Identification of the DNA-binding domains of the switch-activating-protein sap1 from *S. pombe* by random point mutations screening in *E. coli*. Nucleic Acids Res. **22:**2930–2937.
3. **Arcangioli, B., and A. J. S. Klar.** 1991. A novel switch-activating site (SAS1) and its cognate binding factor (sap1) required for efficient *mat*1 switching in *Schizosaccharomyces pombe*. EMBO J. **10:**3025–3032.
4. **Arcangioli, B., and B. Lescure.** 1985. Identification of proteins involved in the regulation of the yeast iso-1-cytochrome C expression by oxygen. EMBO J. **4:**2627–2633.
5. **Brown, D. D.** 1984. The role of stable complexes that repress and activate eucaryotic genes. Cell **37:**359–365.
6. **Egel, R.** 1984. The pedigree pattern of mating-type in *Schizosaccharomyces pombe*. Curr. Genet. **8:**205–210.
7. **Fried, M., and D. M. Crothers.** 1981. Equilibria and kinetics of *lac* repressor-operator interactions by polyacrylamide gel electrophoresis. Nucleic Acids Res. **9:**6505–6525.
8. **Garner, M. M., and A. Revsin.** 1981. A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions. Application to components of the *Escherichia coli* lactose operon regulatory system. Nucleic Acids Res. **9:**3047–3060.
9. **Hendrickson, W., and R. Schleif.** 1985. A dimer of AraC protein contacts three adjacent major groove regions of the *ara1* DNA site. Proc. Natl. Acad. Sci. USA **82:**3129–3133.
10. **Hope, I. A., and K. Struhl.** 1987. GCN4, a eukaryotic transcriptional activator protein, binds as a dimer to target DNA. EMBO J. **6:**2781–2784.
11. **Klar, A. J. S.** 1987. Differentiated parental DNA strands confer developmental asymmetry on daughter cells in fission yeast. Nature (London) **362:**466–470.
12. **Klar, A. J. S.** 1990. The developmental fate of fission yeast cells is determined by the pattern of inheritance of parental and grandparental DNA strands. EMBO J. **9:**1407–1415.
13. **Klar, A. J. S., M. Bonaduce, and R. Cafferkey.** 1991. The mechanism of fission yeast mating type interconversion: seal/replicate/cleave model of replication across the double-stranded break site at *mat1*. Genetics **127:**489–496.
14. **Kurokawa, R., V. C. Yu, A. Näär, S. Kyakumoto, Z. Han, S. Silverman, M. G. Rosenfeld, and C. K. Glass.** 1993. Differential orientations of the DNA-binding domain and carboxy-terminal dimerization interface regulate binding site selection by nuclear receptor heterodimers. Genes Dev. **7:**1423–1435.
15. **LeBowitz, J. H., R. G. Clerc, M. Brenowitz, and P. A. Sharp.** 1989. The Oct-2 protein binds cooperatively to adjacent octamer sites. Genes Dev. **3:**1625–1638.
16. **Mader, S., J. Y. Chen, Z. Chen, J. White, P. Chambon, and H. Gronemeyer.** 1993. The patterns of binding of RAR, RXR and T4R homo- and heterodimers to direct repeats are dictated by the binding specifities of the DNA binding domains. EMBO J. **12:**5029–5041.
17. **Miyata, H., and M. Miyata.** 1981. Modes of conjugation in homothallic cells of *Schizosaccharomyces pombe*. J. Gen. Appl. Microbiol. **27:**365–371.
18. **Perlmann, T., P. N. Ragarajan, K. Umesono, and R. M. Evans.** 1993. Determinants for selective RAR and TR recognition of direct repeat HREs. Genes Dev. **7:**1411–1422.
19. **Rastinejad, F., T. Perlmann, R. M. Evans, and P. B. Sigler.** 1995. Structural determinants of nuclear receptor assembly on DNA direct repeats. Nature (London) **375:**203–211.
20. **Ribes, V., and B. Arcangioli.** Unpublished data.
21. **Singh, J., and A. J. S. Klar.** 1993. DNA polymerase-a is essential for mating-type switching in fission yeast. Nature (London) **361:**271–273.
22. **Tsai, S. Y., M. J. Tsai, and B. W. O'Malley.** 1989. Cooperative binding of steroid hormone receptors contributes to transcriptional synergism at target enhancer elements. Cell **57:**443–448.
23. **Weintraub, H.** 1985. Assembly and propagation of repressed and derepressed chromosomal states. Cell **42:**705–711.
24. **Zhang, L., and L. Guarente.** 1994. The yeast activator HAP1-a GAL4 family member binds DNA in a directly repeated orientation. Genes Dev. **8:**2110–2119.