2930–2937 Nucleic Acids Research, 1994, Vol. 22, No. 15

Identification of the DNA-binding domains of the switch-activating-protein Sap1 from *S.pombe* by random point mutations screening in *E.coli*

Benoit Arcangioli*, Marjan Ghazvini and Véronique Ribes
Oncogenic Viruses Unit, URA 1644 of the CNRS, Department of Biotechnology, Pasteur Institute, 25 rue du Dr Roux, 75724 Paris Cedex 15, France

Received May 31, 1994; Revised and Accepted July 15, 1994

ABSTRACT

Mating type switching in fission yeast, *Schizosaccharomyces pombe*, is initiated by a site-specific double-strand break (DSB) at the *mat1* locus. The DSB is controlled from a distance by *cis*- and *trans*-acting elements. The switch-activating protein, Sap1 binds to the SAS1 *cis*-acting element which controls the frequency of the DSB at the *mat1* locus and, consequently the efficiency of mating type switching. We developed a general method for screening randomly mutagenized expression libraries of DNA-binding protein in *E.coli*. *Sap1* gene was mutagenized by PCR under conditions of reduced *Taq* polymerase fidelity. The mutated DNA was expressed in *E.coli* and screened for SAS1-recognition. This method was used to isolated 16 point mutations that abolished SAS1 interaction together with 18 mutations that did not affect binding. The position of these point mutations allowed the identification of three protein domains located in the N-terminal part of Sap1 that are essential for DNA-binding. Deletions and biochemical analysis showed that Sap1 is a dimer both in solution and when bound to SAS1 sequence. The dimerization domain was localized C-terminally to the three domains described above and when used in exess it inhibited DNA binding.

INTRODUCTION

Haploid *Schizosaccharomyces pombe* cells exist in one of two different mating-types, *P* (for plus) and *M* (for minus) (1). In dividing cells, mating-type switching occurs in such a way that only one cell among four grand-daughters has switched, a restriction that requires two consecutive cell divisions (2). The sister of a newly switched cell is competent for switching during its next division, so that a chain of recurrent switchings are produced (3, 4). Mating-type switching involves substitution of coding DNA at the *mat1* locus with sequences from the cryptic *mat2-P* or *mat3-M* loci (5–8). The site-specific double-stranded break (DSB) present at the *mat1* locus is thought to be the initiating event of mating-type switching (6, 9–11). Furthermore, genetic and molecular studies have shown that the DSB is asymmetrically distributed among the sister chromatids, explaining the different fates of *S.pombe* sister cells (12). A strand-specific imprinting event was proposed to be required to generate the DSB of the chromatid that inherits the parental imprinted strand (4, 12–14).

Functional analyses have revealed that at least two *cis*-acting sites, SAS1 and SAS2 (for switch-activating site), located within the 200 bp sequences located distal to *mat1*, are essential for generating the DSB (6, 15, 16). It was shown that a protein called Sap1 (for switch-activating protein) interacts specifically with SAS1 localised 160 bp away from the *mat1* locus. Sap1 protects two DNA sequences against DNAaseI digestion, domain α (4 bp) and domain β (12 bp) which are separated by 14 residues. Functional analysis has indicated that deletion of the SAS1 element reduces mating-type switching by adversely affecting the level of the double strand break (DSB) at *mat1* locus. Moreover, point mutations generated in SAS1 domains (α or β), that affect binding of Sap1 *in vitro*, were found to reduce the frequency of switching *in vivo* (16). The molecular mechanism by which Sap1 activates the DSB via SAS1 is unknown.

Since the discovery of DNA elements controlling the initiation of several biological processes, such as replication, transcription or site specific-recombination, it has become evident that the frequency and the position of these events involve regulatory proteins that interact with specific DNA sequences. The ability of these regulatory factors to discriminate between the DNA-binding site of a specific sequence and a random sequence is essential for correct function. Many of the DNA binding proteins cloned to date can be grouped into distinct classes based on the conservation of their primary sequence (for review see 17) and a number of protein motifs have been structurally and functionally characterised. For many of these proteins, structural domains or motifs have been characterised that mediate dimer (18) or trimer (19, 20) interactions. However, there are still many DNA-binding proteins that do not share obvious homology with any other known proteins.

*To whom correspondence should be addressed

Using a reverse-genetic approach, the *Sap1* gene was cloned and sequenced. An open reading frame coding for a 29 kDa protein, with no apparent homology with other canonical DNA binding protein motifs, was determined (21). Therefore, Sap1 provides an example of a possible novel type of DNA binding motif. Isolation of the Sap1 gene allowed a genetic and biochemical analysis of its structural and functional properties. Preliminary genetic experiments indicate that Sap1 gene is essential for growth, even in a strain in which mating-type switching is prohibited, suggesting that the Sap1 function is not limited to mating-type switching (21). This result makes the genetic analysis of Sap1 DNA-binding domain in *S.pombe* difficult.

Taking advantage that bacterially expressed Sap1 protein showed the same DNA binding specificities as the purified protein from *S.pombe*, we investigated the DNA-binding motifs of *Sap1* gene product *in vitro*. In this report we describe the characterization of Sap1 DNA-binding and dimerization domains.

MATERIALS AND METHODS

Plasmid construction

PGEX::Sap1 was constructed by inserting the PCR DNA fragment encoding for *Sap1* into pGEX plasmid (22) in frame with the glutathione S-transferase (*GST*) gene. The N-terminus primer corresponded to nucleotides −16 to −4 and contained the *Bam*HI site, while the C-terminus primer was complementary to nucleotide 862 to 879 and is followed by *Hin*dIII site (21).

All of the 6xHis-tag Sap1 polypeptides were expressed from pET vector (23), containing 6 histine codons in 5′ of a *Nde*I and *Bam*HI cloning sites. Sap1 DNA fragments were generated by PCR, using synthetic oligonucleotides as primer. Each primer contained 18 nucleotides homologous to the *Sap1* gene plus 12 nucleotides containing either *Nde*I, or *Bam*HI restriction site, depending if they are used as a N-, or C-terminus primer, respectively. In addition, a stop codon was introduced for each C-terminus primer. Numbers in parenthesis following Sap1 denote the first and the last amino acid retained in the Sap1 protein.

Sap1 mutagenesis

The pGEX::Sap1 fusion protein expression vector was used as template in a PCR mutagenesis method. The N-terminus and the C-terminus primers overlap the *Bam*HI and the *Hin*dIII restriction sites, respectively. Two different conditions (0,1 and 0,5 mM $MnCl_2$) which reduced *Taq* polymerase fidelity were used to amplified Sap1 encoding DNA (24). The resulting mutagenized PCR products were digested with *Bam*HI and *Xho*I and cloned into pGEX::Sap1 to substitute the wild type sequence for the mutagenized one. These constructions fused *GST* gene with 5 untranslated *Sap1* codons and do not introduced mutations into the last 44 amino acids at the C-terminus of Sap1.

DNA sequencing

Plasmid DNA isolated from selected mutants was sequenced (25) with three synthetic primers that hybridize to the coding strand of *Sap1*.

Screening procedure

The two pools of amplified DNA fragments religated into the vector pGST::Sap1 were introduced by transformation into *E.coli* BL21 strain to generate two independent libraries. Transformed bacteria were spread on plates (about 200) and colonies were replica-plated onto nitro-cellulose filters. Expression of GST/Sap1 protein fusions were induced by IPTG (0,5 mM) for 2 hours at 30°C. Filters were transferred on 3 MM paper soaked into lysis buffer containing: 100 mM Tris pH 8, 150 mM NaCl, 0,1 % BSA, 2 mg/ml lysozyme and under chlorophorm vapours for 10 min. at room temperature. Filters were washed four times 5 min. at 4°C into washing buffer containing: 20 mM Tris pH8, 150 mM NaCl, 5 mM EDTA, 0,1% triton, 0,1% BSA and 2 μg/ml sonicated salmo sperm DNA. Filters were incubated for 30 min at 4°C in 5 ml/filter of washing buffer containing 100 ng of SAS1 ^{32}P labelled oligonucleotide (about 10^6 cpm/ng). Then they were washed at 4°C four times for 5 min. in washing buffer. Finally, the filters were autoradiographed to determined SAS1 DNA-binding activity. Sap1 expression was detected on the same filters with an antibody directed against Sap1, essentially as described (26).

Expression of Sap1 polypeptides

Wild type and mutant GST/Sap1 proteins were expressed in *E.coli* BL21 strain. Protein extract were prepared and fusion proteins were isolated as described by Smith and Johnson (22) on glutathione agarose-beads (Sigma).

6xHis-tag Sap1 wild type and derivatives were expressed in *E.coli* BL21 (DE3) pLysS strain, a gene expression system based on bacteriophage T7 RNA polymerase (23, 27). Sap1 purifications were done as described (28) using Nickel-NTA resin (Qiagen) To removed the imidazole, Sap1-containing fractions were dialysed against: 12 mM Hepes pH8, 60 mM KCl, 5 mM $MgCl_2$, 1 mM DTT, 10% glycerol.

Sap1 DNA binding and dimerization assays

Sap1 DNA-binding activity was monitored throughout this study by gel retardation assay essentially as previously described (29). Incubations were performed in a buffer containing 20 mM Tris pH8, 50 mM NaCl, 2,5% glycerol and 10 μg/ml sonicated salmon sperm DNA. The end-labelled probe (to about 5 fM) was added along with the purified or partially purified Sap1 protein. The reactions were mixed at room temperature for 5 min. and electrophoresed through 5% Tris−borate−EDTA polyacrylamide gels at 120 volt for 3 hr. Gels were then dried and autoradiographed.

The SAS1 oligonucleotide sequences were:

TCGAGCCTCTAACGAGATATTTGCTTCGCTACGCTACGC and
TCGAGCGTAGCGTAGCGAAGCAAATATCTCGTTAGAGG

E.coli crude protein extracts containing GST/Sap1 wild type or mutants were mixed with ^{35}S *in vitro* labelled Sap1 protein (TNT T7 Coupled Reticulocyte Lysate System, from Promega) in MTPBS: 150 mM NaCl, 16 mM Na_2HPO_4, 4 mM Na_2HPO_4 (pH7.3), 1 mM PMSF, 3 μM leupeptin, 3 μM pepstatin, 0.1% triton X-100; for 10 min. at 37°C. The GST/Sap1 protein fusions were purified through glutathione-agarose beads (Sigma), eluted with buffer containing 5 mM reduced glutathione (Sigma) as described (22) and analysed by SDS-PAGE. The gel was stained with Coomassie blue to analyse the GST/Sap1 protein fusions, dried and autoradiographied to detect the ^{35}S-labelled Sap1 protein.

DNA-binding interference and dimerization experiments were achieved by mixing two different purified or partially purified

6xHis/Sap1 polypeptides (about 1 μg) as described in the text for 10 min. at 37°C in 10 μl of dialyse buffer (see above). The reaction were diluted 10 times in the same buffer and 1 μl samples were used for gel retardation assay as described above.

Cross-linking experiments were carried out essentially as described (30) by incubating purified 6xHis/Sap1 polypeptide in dialysis buffer (30 μl) as described above at 37°C for 10 min. Cross-linking reaction was initiated by addition of DSP (purchased from Pierce) from 30× stock solution in dimethylsulfoxide to a final concentration of 0.1, 0.3, 1 and 3 mM. The reactions were incubated 20 min at room temperature and then quenched with 20 mM lysine. Laemmli buffer without β-mercaptoethanol was added and complexes were separated by SDS-PAGE and stained with Coomassie blue.

RESULTS

Sap1 mutagenesis and screening strategy

In order to introduce random point mutations within the Sap1 gene, we used a modified PCR procedure. The general scheme of the method used to mutagenize Sap1 is shown in Figure 1. Two amplification conditions were used (see Materials and Methods). The first used 0.1 mM $MnCl_2$ in order to decrease slightly the fidelity of DNA synthesis by Taq DNA polymerase, while the second condition used 0.5 mM $MnCl_2$, increasing the overall frequency of mutations (24). The two amplified and mutagenized pools of DNA fragments encoding the Sap1 gene were subsequently digested with *BamH*1 and *Xho*I restriction enzymes in order to substitute the wild type sequence present in the expression vector pGex::Sap1 expressing a GST−Sap1 fusion protein (See Materials and Methods), thereby generating two independent mutant expression libraries. Library I and library II contain slightly and highly mutated Sap1 sequences, respectively. This strategy did not introduce mutations into the last C-terminal 44 amino-acids; however, this should not interfere with Sap1 DNA-binding analysis, since we have previously shown that Sap1 DNA-binding activity does not require this part of the protein (21).

In order to analyze Sap1 DNA-binding activity, we screened bacterial colonies from libraries I and II on nitrocellulose filters with a radiolabelled double-stranded oligonucleotide encoding the SAS1 sequence (see Materials and Methods). A short exposure of the nitrocellulose filters allowed detection of positive and negative colonies for specific SAS1-binding activity (Fig. 2B). This protocol revealed that libraries I and II contained 10% and 55% mutant colonies respectively. In principle, two classes of mutations may reduce Sap1 DNA-binding activity, mutations affecting SAS1 recognition and mutations affecting Sap1 expression, such as non-sense mutations, deletions and insertions. In order to distinguish between these, the nitrocellulose filters used to detect Sap1 binding activities were incubated with antibodies directed against Sap1 (Fig. 2A). In doing so, we can disregard mutants which reduce or abolish Sap1 expression and further analyse only mutants affected for SAS1 recognition.

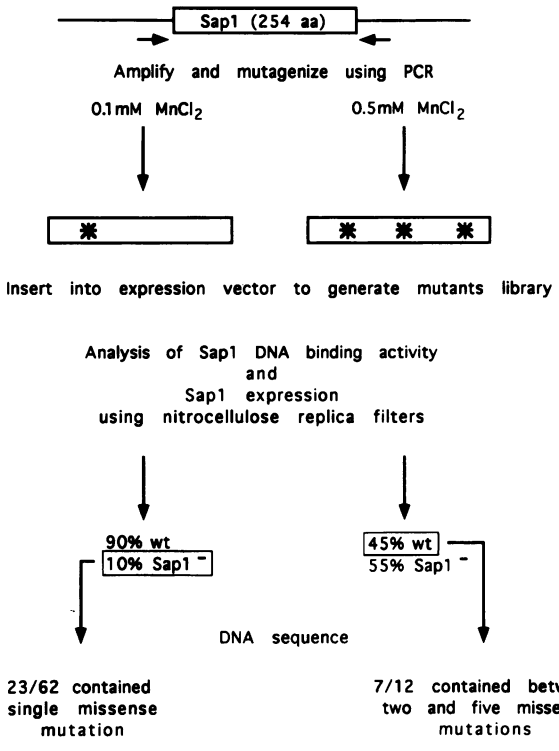

Figure 1. Outline of the strategy for generating random point mutations into *Sap1* open reading frame and screening for DNA-binding mutants, using the pGex expression vector. The relevant steps are indicated. See Materials and Methods for details.

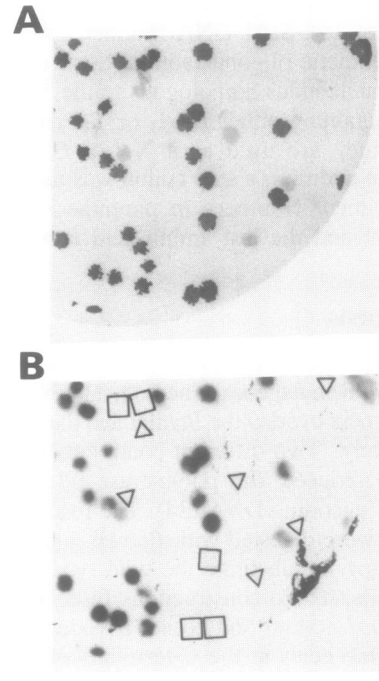

Figure 2. Detection of Sap1 DNA-binding activity and Sap1 expression using nitrocellulose replica filters. *E.coli* BL21 colonies expressing GST−Sap1 protein fusions were replica plated: (**A**) Identification of Sap1 expression by immunological analysis; (**B**) Autoradiography of the nitrocellulose filter used above and probed with a ^{32}p radio labelled SAS1 oligonucleotide. Squares indicate *E.coli* colonies which express GST−Sap1 protein fusion and are affected for SAS1 cognition. Triangles, indicate *E.coli* colonies which do not express GST−Sap1 protein fusion and consequently do not bind SAS1 oligonucleotide.

Sequence analysis of Sap1 mutants

Two different types of Sap1 mutants were sequenced and further studied. From library I, we looked for mutants having a single point mutation, that reduced Sap1 DNA-binding activity but still reacted with antibodies directed against the protein. Among 62 candidates sequenced, only 23 contained a single missense mutation. Due to the mutagenesis method used, some mutations were redundant and only 16 independent missense mutants were actually isolated (on the right of Fig. 3). From library II, we looked for presumed mutants having single or multiple point mutations and still bound to SAS1. Since these mutants still bind to the target, the mutated amino acids should indicate protein domains which are not involved in SAS1 recognition and the analysis should be complementary to the previous screening. Among 12 sequenced candidates, 7 contained between two and five missense mutations (on the left of Fig. 3). Mutagenesis using 0.1 mM $MnCl_2$ produced mostly transitions, while using 0.5 mM $MnCl_2$ produced roughly an equal frequency of transitions and transversions (data not shown).

On the right side of Figure 3 are shown the position of the single point mutations found in the Sap1 mutants which have a reduced SAS1-binding activity. Three clusters of mutations which may define three protein domains are distinguishable. Domains I, II, III contained the amino-acid residues from T(30) to L(51), T(77) to I(88) and L(112) to L(126), respectively. On the left side of Figure 3 are shown the position of mutations found in seven Sap1 mutants maintaining a wild type like DNA-binding

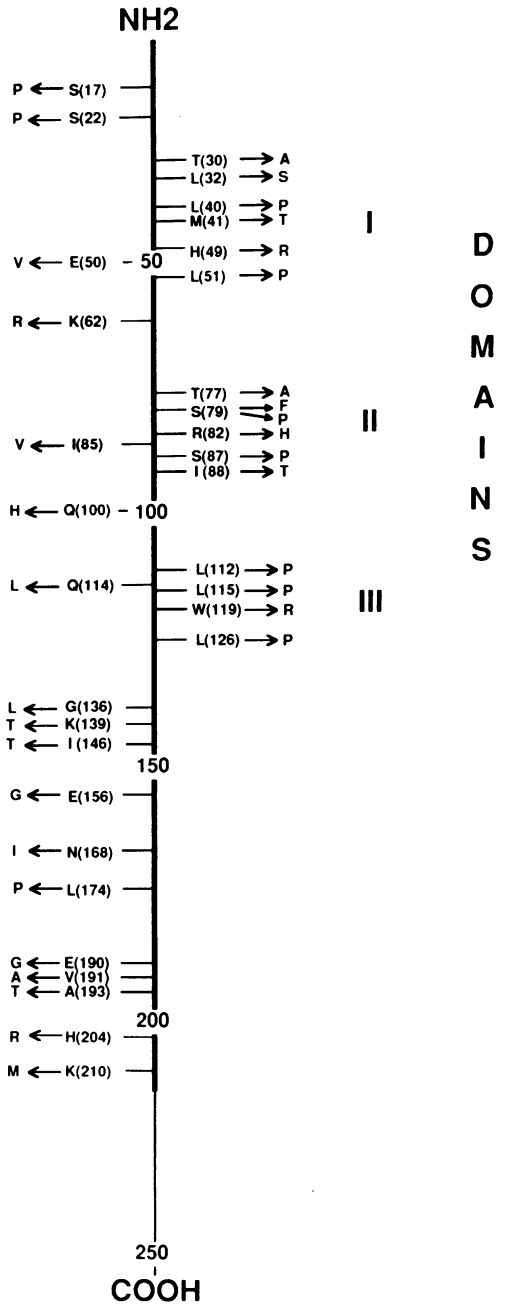

Figure 3. A collection of GST−Sap1 mutants. For simplicity only the Sap1 portion of the protein fusion is shown. Location of point mutations which reduced SAS1 recognition are indicated on the right side of the figure. The mutants are described by: wild-type residue/position/mutant residue. The mutations in *Sap1* which did not reduce SAS1 recognition are indicated on the left side of the figure (See Materials and Methods for sequences). Sap1 domains I, II and III are indicated.

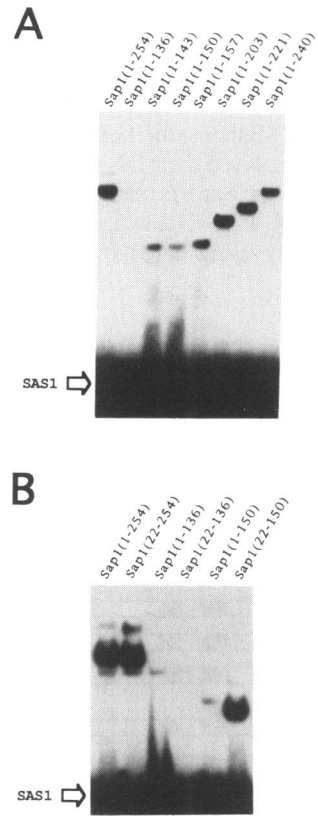

Figure 4. Electrophoretic mobility shift analysis of Sap1 deletion mutants. Crude *E.coli* protein extract containing truncated Sap1 fused to a 6xHis residues at the N-terminus were prepared and Sap1 concentration were estimated by SDS-PAGE followed by coomassie blue staining (data not shown). Crude extracts containing equal amounts of Sap1 were incubated with end labelled SAS1 double strand oligonucleotides and subjected to electrophoretic mobility analysis. The number following Sap1 in parenthesis indicates the amino acids retained. (**A**) Binding activity of mutants: Sap1(1−254), Sap1(1−136), Sap1(1−143), Sap1(1−150), Sap1(1−157), Sap1(1−203), Sap1(1−221), Sap1(1−240). (**B**) Binding activity of mutants: Sap1(1−254), Sap1(22−254), Sap1(1−136), Sap1(22−136), Sap1(1−150), Sap1(22−150).

activity. These mutations are dispersed all along the Sap1 sequence and are mostly non-overlapping with the first class of mutants.

Further analysis of the GST−Sap1 mutants was carried out by electrophoretic mobility shift assay. sixteen Sap1 mutants containing point mutation were expressed in *E.coli*, proteins extracts were prepared and gel shift analyses were performed using SAS1 oligonucleotide as a probe. The results indicated that 15 mutant proteins had a DNA-binding activity reduced to at least 10%, except for the mutant I(88)T which has a DNA-binding activity reduced only 3 fold (data not shown).

These data allowed us to defined three Sap1 protein domains, I, II and III which are required for SAS1 specific DNA-interaction. However, this analysis did not establish whether these regions are essential for direct DNA recognition or for potential oligomerization.

Domains I, II and III are essential but not sufficient for SAS1 binding

Our results, clearly showed that efficient binding to SAS1 sequence required, at least, the N-terminal part of Sap1, encoding for domains I, II and III. However, we do not know if the N-terminus of Sap1 is sufficient by itself for binding and to what extent the GST portion of the protein fusions may interfere with the DNA-binding activity of those mutants. To address these questions we fused a tail composed of six histidines residues (6xHis) at the N-terminus of various truncated forms of Sap1 (See Materials and Methods for construction), overexpressed them in *E.coli* and analysed their DNA-binding activity *in vitro*.

We first estimated the Sap1 protein concentration in the crude protein extract by SDS-PAGE (data not shown) and analysed their DNA-binding activity using SAS1 oligonucleotide as a probe.

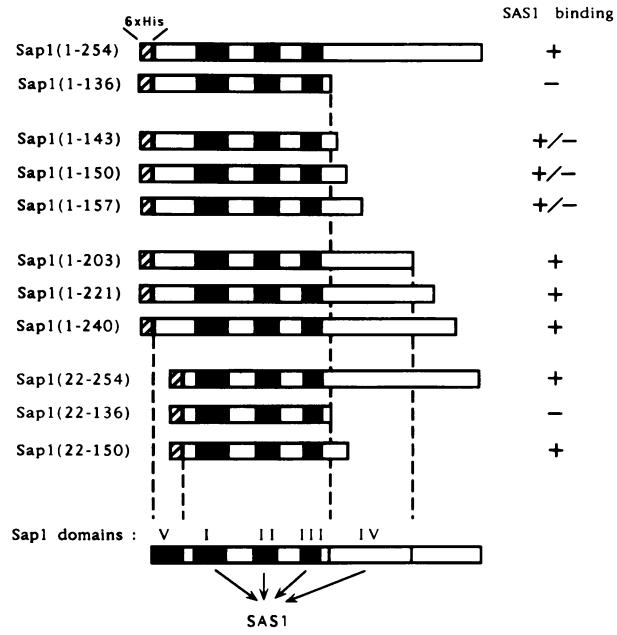

Figure 5. DNA-binding activity for the serial deletion mutants of *Sap1*. The Sap1 DNA-binding activity for SAS1 is indicated. Sap1 domains I, II, III (See Fig. 3), IV and V are shown. The arrows indicated Sap1 domains required for specific binding.

Figure 4A shows that the 6xHis tail fused at the N-terminus of full length Sap1, Sap1(1−254), did not interfere with SAS1 binding. However, the Sap1 deletion mutant, Sap1(1−136), which still contained domains I, II and III, did not bind SAS1. The addition of 7 residues at the C-terminus in mutant Sap1(1−143) restored binding to SAS1, although with a weaker affinity than the full length protein. The same result was obtained with the mutants Sap1(1−150) and Sap1(1−157). The wild type Sap1 DNA-binding activity was restored for mutants Sap1(1−203) and longer derivatives.

As shown in Figure 4B, we also analysed the SAS1-binding activity of Sap1 mutants containing a deletion of the first 22 residues, since this segment of the protein did not contain negative mutations. Sap1(22−254) binds SAS1 with an equivalent affinity as the full length protein. A similar result was obtained with Sap1(22−203) (data not shown) Intriguingly, when the same deletion was made in Sap1(1−150), which weakly bound SAS1, a strong SAS1-binding activity was restored.

These results demonstrate that domains I, II, and III are required but are not sufficient for SAS1 specific DNA-binding. For efficient binding two additional domains, IV and V, were identified (see Fig. 5). Deletion of C-terminal sequences of Sap1 that gave rise to a polypeptide, such as Sap1(1−203), did not affect the DNA-binding activity. Further, deletion between residues 203 and 136, which defined domain IV, reduced and finally abolished binding to SAS1. Domain V, was defined by the deletion of the first 22 amino-acids of Sap1. The presence of domain V reduced SAS1 recognition only when domain IV was partially deleted.

Domain IV overlaps with an oligomerization motif

Several Sap1 protein domains responsible for sequence-specific DNA-binding were defined above. Since it is possible that these domains were responsible for either contacting nucleic acid or

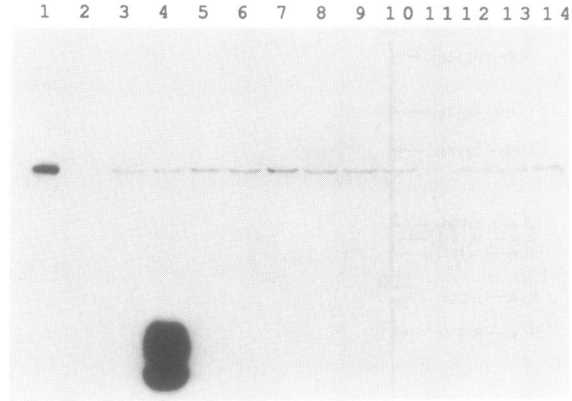

Figure 6. Oligomerization analysis of GST−Sap1 mutants with Sap1 protein. Autoradiograph of a 12% SDS-PAGE showing the *in vitro* labelled Sap1 product bound to GST−Sap1 mutants and eluted from glutathion beads. Lane 1, the ^{35}S *in vitro* translated Sap1 product. Lanes 2−14 the labelled Sap1 protein was incubated with *E.coli* protein extract containing the GST protein (lane 2) or the GST−Sap1 wild-type in absence or in presence of ^{32}P labelled SAS1 DNA (Lane 3 and 4, respectively). Lanes 5, 6 and 7 are the GST−Sap1 containing point mutation L(40)P, R(82)H and L(115)P, respectively. The others 13 GST−Sap1 containing point mutation were also tested (data not shown). Lanes 8−14 are the seven GST−Sap1 mutants (from +1 to +7, respectively) having a similar SAS1 binding activity than the wild-type protein (See Materials and Methods for sequence).

oligomer formation, we tested the ability of the GST−Sap1 mutants to interact with *in vitro* translated ^{35}S-labelled Sap1 protein. In this assay, we first incubated GST−Sap1 mutants with the ^{35}S-Sap1 protein, then we loaded the mixture onto a glutathione-Sepharose column. Proteins bound were eluted and subjected to SDS-PAGE (See Materials and Methods). Figure 6 shows that ^{35}S-Sap1 coeluted with the GST−Sap1 protein fusion but not with the GST protein from the glutathione-Sepharose indicating a potential oligomerization of Sap1 in solution (compare lanes 2 and 3). In addition these data indicated that all the bacterially expressed GST/Sap1 mutant proteins are stable and soluble. The presence of SAS1 oligonucleotide did not further stimulate the association between the proteins (lane 4). Among the GST−Sap1 mutants only one mutant, called Sap1(+5), which contained four point mutations S(17)P, S(22)P, L(174)P and H(204)R, did not interact with the labelled Sap1 protein (lane 11). However, this mutant still bound SAS1, as strongly as the wild type protein, and did not have the phenotype expected for a mutant deficient in oligomerization.

Such behaviour may occur if mutant GST-Sap1(+5) forms very stable oligomers or has a new oligomerization properties incompatible with normal Sap1. To test these hypotheses, we transferred the mutations onto 6xHis type protein fusion, either onto the full length Sap1(1−254), or on a construction which encoded only for the C-terminal part of Sap1, Sap1(134−254). These proteins were expressed in *E.coli* and were purified through a Nickel (Ni-Ta) column (see Materials and Methods) and then analyzed by mobility-shift assays, using SAS1 DNA as a probe. We reasoned that if the C-terminal part of Sap1 encodes for an oligomerization motif, then when mixed with the full length Sap1 protein, prior to the incubation with the SAS1 probe, it may abolish DNA-binding. Indeed, the C-terminal wild type and mutated peptides interfered with the formation of the Sap1(1−254)/SAS1 complex (Fig. 7, lanes 1−10). However, only the mutated C-terminus and not the wild type polypeptide interfered with the formation of the Sap1(+5)/SAS1 complex (Fig. 7, lanes 11−20). The DNA-binding inhibition required prior incubation at 37°C of the two polypeptides, no effect was seen when the two polypeptides were incubated at 4°C (data not shown). No protein degradation occurred during incubation at 37°C (data not shown). Taken together, these results indicate that the two mutations (L(174)P and H(204)R) still present in the C-terminal part of Sap1(+5) promote a more stable self-oligomerization in solution than the wild type protein, but can still oligomerize with the wild type protein.

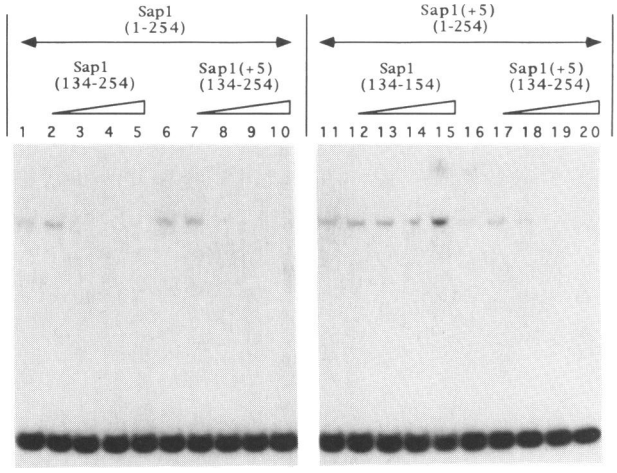

Figure 7. Electrophoretic mobility shift interference using wild-type and mutate. C-terminal polypeptides. Crude *E.coli* protein extract containing full length Sap1(1−254) wild type or mutant (+5) were preincubated 10 min. at 37°C with either Sap1(134−254) wild-type or mutant (+5) purified polypeptides and then mixed with ^{32}P labelled SAS1 oligonucleotides (See Materials and Methods for details). Protein concentrations were estimated by SDS-PAGE followed by coomassie blue staining. Sap1(1−254)/Sap1(134−254) ratio varied from 1/1, 1/3, 1/9 to 1/27.

Domain IV is involved in Sap1 dimerization

To determine the multimerization state of the Sap1 protein in solution, the 6xHis-Sap1(1−254) protein produced in *E.coli* was purified through a Nickel column, and cross-linked with the bifunctional protein cross-linker Dithiobis (DSP), prior to analysis on SDS-PAGE. When the Sap1 protein was subjected to increasing concentrations of DSP, the band which corresponds to the 32 kd monomer Sap1 was converted into new bands with apparent molecular size ranges between 55−70 kd (Fig. 8, lanes 1−5). At high concentrations of DSP, when 100% of the monomer was cross-linked an additional faint band appeared with a apparent molecular size of 150−200 kd. From their migration positions, we believe that these bands correspond to Sap1 dimer and multimer. A similar result was obtained with the Sap1(1−203) (data not shown).

The same experiments were repeated with the N-terminal part of Sap1, Sap1(1−136), which lacks domain IV. When this truncated protein was treated with a DSP concentration capable of converting 100% of the monomeric Sap1 into a dimer, no additional major bands were found, suggesting that the N-terminal part of Sap1 is a monomer in solution. At the higher cross-linker concentration a faint smear became visible and might be attributed to non-specific cross-link between Sap1(1−136) monomers. The increasing mobility of the treated monomers indicated the efficient linking of the DSP to Sap1(1−136) (Fig. 8, lanes 6−10). In conclusion the position and the number of complexes seen after chemical cross-linking argue that, in solution and at high concentration, Sap1 exists as a dimer.

In subsequent experiments we expressed and purified from *E.coli* the 6xHis-Sap1(134−203) and 6xHis-Sap1(134−254) polypeptides, (containing domain IV) and their multimeric states

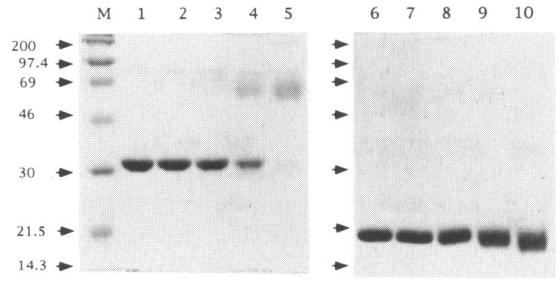

Figure 8. SDS-PAGE analysis of cross-linked full length and N-terminus Sap1 polypeptides. Purified Sap1(1−254) (lanes 1−5) and Sap1(1−136) (lanes 6−10) were cross-linked *in vitro* with increasing concentration of DSP: 0, 0.1, 0.3, 1 and 3 mM (lanes 1 and 6, 2 and 7, 3 and 8, 4 and 9, 5 and 10, respectively). The products were resolved on 12% SDS-PAGE and stained by Coomassie blue. The molecular size markers are shown one the left (kDa).

were analysed by SDS-PAGE under non-reducing conditions. Sap1(134−203) migrated as two polypeptides with apparent molecular weights of 9,000 and 23,000 (bands a and c in Fig. 9). Sap1(134−254) migrated also as two polypeptides with apparent molecular weights of 15,000 and 30,000 (bands b and e in Fig. 9). When both polypeptides were incubated together prior migration an additional band appeared with an apparent molecular weight of 27,000 (band d in Fig. 9). The additional bands were caused by slight cross contamination during loading. When the reaction mixtures were treated with the bifunctional protein cross-linker DSP prior to analysis on SDS-PAGE, similar results were obtained. These data do not determine precisely the stoichiometry between these polypeptides. However, with the exception of the slightly abnormal position of the band c, it is reasonable to conclude that domain IV exist as a dimer in solution.

We believe that the dimers resist SDS denaturation in absence of reducing agent by formation of interchain S-S bonds of cysteines residues at position 140 and 142 (See 21 for sequence).

We next tried to confirm that Sap1 interacts indeed with SAS1 as a dimer. To do this, we adopted an approach developed by Hope and Struhl (31). Two forms of 6xHis-Sap1 with different lengths, Sap1(1−254) and Sap1(1−203), were isolated from *E.coli* and analysed by mobility shift assay individually or in combination. When incubated individually, Sap1(1−254) and Sap1(1−203), formed one major protein−DNA complex each, with different mobilities (Fig. 10). When both proteins were incubated together at 37°C (not at 4°C) prior to binding, one additional complex with an intermediate mobility was seen (Fig. 10). These results confirmed the previous data and support the conclusion that Sap1 forms a dimer when bound to SAS1.

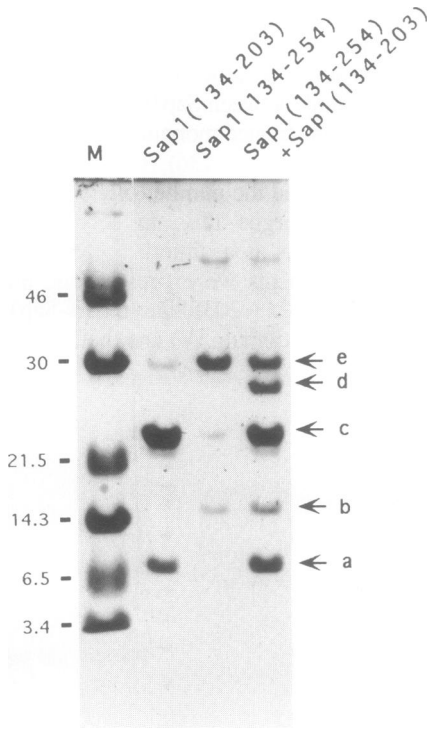

Figure 9. SDS-PAGE analysis of Sap1 dimerization domain. Purified Sap1(134−254) and Sap1(134−203) were incubated at 37°C, individually or in combination, without reducting agent, resolved on 12% SDS-PAGE and stained by Coomassie blue. Monomeric and multimeric polypeptides species are indicated with arrows and molecular size markers are shown one the left (kDa).

DISCUSSION

In a search for random point mutations in Sap1 which affect specific DNA-recognition we were able to identify three clusters of point mutations as potential domains I, II and III. These three domains are contiguous and are localized in between residues T30 and L126 of Sap1. However, the N-terminal part of Sap1, Sap1(1−136), is unable to bind the SAS1 mating type switching control element indicating that although domains I, II and III are essential for recognition they are not sufficient. Since the wild type Sap1/SAS1 DNA-binding activity is restored with Sap1(1−203), a fourth domain (IV), encompassing 67 residues from amino acids 136 to 203 is essential for SAS1 DNA-binding and that residues from 203 to 254 are dispensable. Intermediate deletions into domain IV gave rise to polypeptides with reduced SAS1 DNA-binding activities. The failure to isolate nonfunctional point mutations in this domain may be the result of our screening strategy, which searches for strong mutant phenotypes, and also of the intrinsic properties of the dimerization domain, which does not prevent interaction with DNA when partially deleted. Finally, the first 22 amino acids of Sap1 repressed SAS1 specific recognition only when domain IV was partially deleted. In conclusion, SAS1 specific binding required Sap1 domains I, II, III, and IV (see Fig. 5), which are contained between residues 23 and 203.

Sap1 DNA-binding region

The surprising distribution of the sixteen point mutations in GST/Sap1 reducing SAS1 recognition strongly suggested the existence of three protein domains. Also, these domains are separated by only few functional mutation while a lot of the

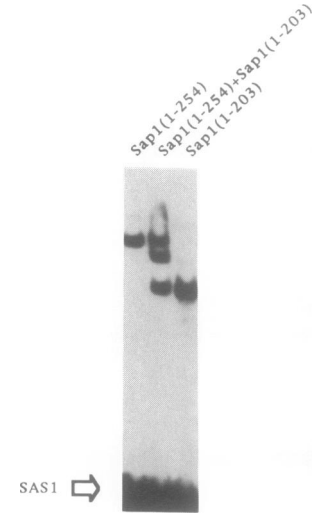

Figure 10. Electrophoretic mobility shift assay for dimer formation. Sap1(1−254) and Sap1(1−203) partially purified were preincubated alone or mixed for 10 min. at 37°C before incubation with the ^{32}P labelled SAS1 oligonucleotides for analysis by mobility-shift assay. (lane 1) Sap1(1−254), (lane 2) Sap1(1−254) and Sap1(1−203), (lane 3) Sap1(1−203). Are also indicated: (SAS1) free probe; (LL) Sap1(1−254) homodimer; (SS) Sap1(1−203) homodimer; (LS) Sap1(1−254)/Sap1(1−203) heterodimer.

dysfunctional mutations are severe amino acid replacements (7/16 create new proline residues). However, we like to propose that domains I, II and III specifically mediate at least part of the DNA sequence contacts. One might speculate that domain I, which is rich in lysine and arginine residues, may be involved in protein/nucleic acid contacts and that domains II and III may contribute to the three dimensional folding of this region. This prediction led us to reexamine the potential homology between this region with that of other DNA-binding proteins. No similarity was observed, suggesting that Sap1 constitutes a novel class of DNA-binding protein. Studies to determine the 3 dimensional structure of the Sap1/DNA complex are in progress.

The dimerization domain

For many DNA binding proteins, oligomerization is an essential part of the binding interaction and also an important means by which binding is regulated. We have investigated the oligomerization state of Sap1 in solution and when bound to SAS1. Using a GST−Sap1 protein fusion and an *in vitro* labelled Sap1 protein we have shown that Sap1 oligomerizes in solution. In addition, none of the sixteen point mutations which reduced DNA-binding activity reduced Sap1 oligomerization, suggesting that another domain is responsible for oligomerization.

Chemical cross-linking experiments indicated that Sap1 produced in *E.coli* formed complexes with an apparent molecular weight that is twice that of the monomer, consistent with a dimeric association of Sap1. In addition, we have localized the protein region responsible for the dimerization between residues 136 and 203. By using long and short derivatives, we demonstrated that three complex species are formed when incubated together with SAS1 DNA (31). Such complexes were not seen when the shorter polypeptide is deleted for domain IV. These data showed that domain IV encodes a dimerization motif, essential for SAS1 specific DNA-binding. Several lines of evidence indicated that the C-terminal end of Sap1 (residues 204−254) is not involved in DNA-binding and oligomerization. First, the DNA-binding activity of full length Sap1 and Sap1(1−203) are identical (Fig.4A). Second, when the same concentation of both polypeptides were mixed together they formed homo- and heterodimers on DNA with the same proportion (Fig. 10).

We previously described (21) that a region from residues 143 to 173 displays a strong potential to form an amphipatic α-helical structure when analysed with the Chou and Fasman (32) algorithm. The hydrophobic surface may form the contact between paired helices. One example of this association is the 'leucine zipper' which has been shown to be essential for the dimeric association of several transcription factors (18). The observation that domain IV alone can dimerized and formed an NH_2-terminal disulfide linkage (Fig. 9) indicates that the helices are parallel. However such a covalent bond seems to occur less easily with the full length Sap1 protein (Fig. 8) In addition, the secondary structure prediction also indicated that this α-helix (helix1) is broken at the residue L174, and another α-helix structure (helix2) is predicted until residue Q200. In this context, it is interesting to recall that mutant Sap1(+5) containing the mutation L(174)P located in between helix1 and 2, dimerized more avidly in solution than the wild type protein. This mutation may stabilise the α-helical conformations by acting as a 'terminator' or 'initiator' (33, 34). It is also possible that this mutation favoured helix2 to bend back and interact with helix1 (35).

Intriguingly, when domain IV is present, the N-terminal end of the protein do not influence Sap1 DNA-binding activity, while reduced DNA-binding activity imposed by the partial removal of domain IV can be restored by deletion of the 22 first residues (Fig. 4B). Its remain possible that this intriguing behaviour reflects a conformational change of the protein structure or of the oligomerization properties. Finally we show that the dimerization domain by itself behaves as a dominant inhibitor of DNA-binding probably by making inactive heterodimers.

ACKNOWLEDGEMENTS

We thank members of the Oncogenis Virus Unit for discussions and comments on the manuscript, Moshe Yaniv and Simon Whiteside for critically reading the manuscript. M.G. was supported by a fellowship from the Ministere 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. Leupold, U. (1950) *Ser. Physiol.*, **24**, 381−480.
2. Miyata, H. and Miyata, M. (1981) *J. Gen. appl. Microbiol.*, **27**, 365−371.
3. Egel, R. and Eie, B. (1987) *Curr. Genet.*, **12**, 429−433.
4. Klar, A. J. S. (1990) *EMBO J.*, **9**, 1407−1415.
5. Egel, R. and Gutz, H. (1981) *Curr. Genet.*, **3**, 5−12.
6. Beach, D. H. (1983) *Nature*, **305**, 682−688.
7. Beach, D. H. and Klar, A. J. S. (1984) *EMBO J.*, **3**, 603−610.
8. Kelly, M., Burke, J., Smith, M., Klar, A. J. S. and Beach, D. (1988) *EMBO J.*, **7**, 1537−1547.
9. Egel, R., Beach, K. and Klar, A. J. S. (1984) *Proc. Natl. Acad. Sci.* USA, **81**, 3481−3485.
10. Klar, A. J. S. and Miglio, L. M. (1986) *Cell*, **46**, 725−731.
11. Nielson, O. and Egel, R. (1989) *EMBO J.*, **8**, 269−276.
12. Klar, A. J. S. (1987) *Nature*, **362**, 466−470
13. Egel, R. (1984) *Curr. Genet.*, **8**, 205−210.
14. Klar, A.J.S *et al.* (1991) *Genetics*, **127**, 489−496.
15. Gutz, H. and Fecke, H. C. (1979) *Hoppe-Seyler's Z. Physiol. Chem.*, **360**, 274.
16. Arcangioli, B. and Klar, A. J. S. (1991) *EMBO J.*, **10**, 3025−3032.
17. Harrison, S. C. (1991) *Nature*, **353**, 715−719.
18. Landschulz, W. H., Johnson, P. F. and McKnight, S. L. (1988). *Science.* **240**, 1759−1764.'
19. Soger, P. K. and Nelson,H. C. (1989) *Cell*, **59**, 807−813.
20. Perisic, O., Xiao, H. and Lis, J. T.(1989) *Cell*, **59**, 797−806.
21. Arcangioli, B., Copeland, T. D. and Klar, A., (1994), *Mol. cell. Biol.* **14**, 2058, 2065.
22. Smith, D. B. and Johnson, K. S. (1988),*Gene*, **67**, 31−40.
23. Rosenberg, A. H., Lade, B. N., Chui, D., Lin, S., Dunn, J. J. and Studier, F. W. (1987), *Gene*, **56**, 125−135.
24. Leung, D. W., Chen, E. Y. and Goeddel, D. V. (1989) *Technique*, **1**, 11−15.
25. Sanger, F., Nieklen, S. and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA*, **101**, 5463−5467.
26. Maniatis, T., Fritsch, E. F. and Sambrook, J. (1982) *In Molecular Cloning* Cold Spring Harbor Laboratory, New York.
27. Studier F. W. and Moffatt, B. A. (1986), *J. Mol. Biol.*, **189**, 113−130.
28. Hoffmann, A. and Roeder, R.G. (1991), *Nucl. Acids. Res.*, **19**, 6337.
29. Arcangioli, B. and Lescure, B. (1985) *EMBO J.*, **4**, 2627−2633.
30. Yang-Yen, H., Chambard, J., Sun, Y., Smeal, T., Schmidt, T., Drouin, J., and Karin, M. (1990) *Cell*, **62**, 1205−1215.
31. Hope, I. A. and Struhl, K. (1987) *EMBO J.*, **6**, 2781−2784.
32. Chou, P.Y.,and Fasman, G. D. (1978), *Adv. Enzymol*, **47**, 145−148.
33. Presta, L., G. and Rose, G., D. (1988) *Science*, **240**, 1632−1641.
34. Cohen, C. and Parry, D. A. (1990), *Proteins*, **7**, 1−15.
35. Carr, C.M. and Kim, P.S. (1993), *Cell*, **73**, 823−832.