JOURNAL OF BACTERIOLOGY, Mar. 1994, p. 1702–1710
0021-9193/94/$04.00+0

Vol. 176, No. 6

A Fourth Secreted Aspartyl Proteinase Gene (*SAP4*) and a *CARE2* Repetitive Element Are Located Upstream of the *SAP1* Gene in *Candida albicans*

SHELLEY H. MIYASAKI, THEODORE C. WHITE, AND NINA AGABIAN*

Intercampus Program in Molecular Parasitology, Departments of Pharmaceutical Chemistry and Stomatology, and the Oral AIDS Center, University of California at San Francisco and Berkeley, San Francisco, California 94143

Received 3 November 1993/Accepted 12 January 1994

Candida albicans **secreted aspartyl proteinases (Sap), products of the *SAP* genes, which are presumed to act as virulence factors. In the *C. albicans* strain WO-1, the ability to secrete Sap1 is regulated with switch phenotype, another putative virulence factor. *Kpn*I restriction fragment length polymorphisms differentiate between several distinct *SAP1* alleles in laboratory and clinical strains. Both *SAP1* alleles from strain WO-1 along with their 5′- and 3′-flanking regions were cloned and sequenced, as were both alleles from another strain, SS. The 5′-flanking regions were remarkably similar in all four of the sequenced alleles over approximately 1,500 nucleotides. S1 analysis revealed that both alleles of WO-1 are transcribed. Characterization of the one allele from strain WO-1 identified a 284-nucleotide insertion flanked by 8-bp direct repeats that shows homology to the *CARE2* repetitive element and that is not present in the other alleles. Characterization of the *SAP1* alleles also identified a fourth *SAP* gene (*SAP4*) that includes an extended leader sequence. *SAP4* is positioned upstream, in tandem to *SAP1*, in all strains tested and may encode another closely related secreted aspartyl proteinase.**

Candida albicans is one of the most common causes of opportunistic infections among immunocompromised patients. Factors which may contribute to the conversion of this normally commensal organism to a more virulent form include a morphological transformation between yeast and hyphal cells (6) and the ability of a *Candida* species to express alternative colony and cell shapes, a phenomenon which has been referred to as phenotypic switching (reviewed by Soll [31]). The phenotypic switching system of strain WO-1 has been characterized in detail (30, 32) and is easily manipulated, making it an important laboratory tool for examining the relationship between switch phenotype and the expression of virulence characteristics. In strain WO-1, the cells switch reversibly and at high frequency between predominantly spherical and smooth cells and cells which are larger, ovoid, and rough in appearance. Each of these cell types gives rise to a distinct colony phenotype, white (W) and opaque (O), respectively, on solid medium. The switch phenotype also affects properties associated with virulence, such as cell permeability, adhesiveness, drug susceptibility, and sensitivity to neutrophils and oxidants (32). The ability to switch between morphologic forms therefore seems to be a reflection of more pervasive changes in the spectrum of proteins expressed by each phenotype, a feature which is particularly interesting in the context of virulence, as each distinct cell type may be able to respond differently to the host.

The secretion of proteolytic enzymes by *C. albicans* was first noted by Staib in 1965 (33), and since that time, a *C. albicans*-secreted aspartyl proteinase (Sap) activity has been linked with virulence through indirect observation. It has been reported that Sap activity is directly proportional to the virulence of several different *Candida* species, the more virulent species, such as *C. albicans*, producing more Sap than *Candida glabrata*, for example (26). Immunocytological studies have detected Sap antigens in the tissues of infected patients and animals, and Sap-deficient mutants of *C. albicans* are less able to adhere to host tissues and less virulent in experimental animal models (6, 8, 13, 15, 24). Sap activities purified from different *C. albicans* isolates also cleave a number of appropriate host substrates in vitro, including keratin, collagen, albumin, hemoglobin, immunoglobulin A (IgA), and secretory IgA (10, 12, 23, 25, 33). Thus, many possible roles have been proposed for Sap in virulence, including assisting in adhesion to and invasion of the host and in interfering directly with host defense mechanisms (4). The ability to express a Sap activity was recently associated with phenotypic switching in the white-opaque strain WO-1 (19), providing a possible connection between two putative virulence factors.

Determining the specific role of Sap in pathogenesis has been difficult for several reasons. In particular, studies characterizing Sap activity have been inconsistent, reporting differences in pH optima, inhibition profiles, and substrate specificities (25). These differences suggested that more than a single type of Sap might exist, a perception which has recently been confirmed through molecular genetic studies which have identified several different aspartyl proteinase genes in *C. albicans*: *SAP1* by Hube et al. (11), *SAP2* by Wright et al. (36), and *SAP3* by White et al. (35). The presence of multiple loci in the *C. albicans* genome which encode similar aspartyl proteinases suggests that reports of the induction of Sap expression by various substrates and under different growth conditions may in fact reflect the differential expression of each of these loci or of alleles within a given locus.

At this time, little is known about the differential expression of these three Sap proteinases under various conditions. Wright et al. (36) showed that *SAP1* and *SAP2* transcripts are produced by strain ATCC 10261 in medium containing yeast extract and glucose supplemented with bovine serum albumin

* Corresponding author. Mailing address: Department of Pharmaceutical Chemistry, University of California at San Francisco, Suite 150, 3333 California St., San Francisco, CA 94118. Phone: (415) 476-6845. Fax: (415) 476-0664. E-mail: twhite@cgl.ucsf.edu.

(BSA), although they showed that the *SAP2* mRNA was more abundant. Morrow et al. (19) showed that the *SAP1* gene is expressed exclusively in O cells and not in W or hyphal cells of strain WO-1. In addition, they showed that the *SAP1* gene is not expressed in budding or hyphal cells of strain 3153A or in cells of two clinical isolates. *SAP1* is expressed in O cells under a variety of conditions, including growth in defined medium and serum induction (19). Recently, they showed that transcription of *SAP1* and a second opaque-specific gene, *Op4*, is immediately inhibited by shifting the temperature from 25 to 37°C, a shift which results in the mass conversion of O cells to W cells (18). This inhibition of *SAP1* and *Op4* transcription can be reversed by lowering the temperature before the second semisynchronous cell cycle. Our Western blot (immunoblot) analysis of culture supernatants and subsequent N-terminal protein sequencing (35) confirmed that the Sap1 protein is O specific and identified a second O-specific Sap protein, Sap3. This Western analysis also showed that Sap2 is the predominant protein expressed in strain 3153A, in strain SS, and in W cells of strain WO-1. These Western analyses were conducted in proteinase-inducing medium to maximize the production of Sap proteinases. The correlation between the *SAP1* transcription results of Morrow et al. (19) and the Western blot results (35) suggests that the pattern of differential proteinase expression is not altered by growth in these different media.

Our studies were initiated with the *SAP1* gene locus, the only *SAP* locus which had been identified when these studies began. We found a correlation between restriction fragment length polymorphisms (RFLPs) in the *SAP1* alleles and the O-specific expression of *SAP1*. Comparisons of the Southern hybridization patterns of WO-1 and two other *C. albicans* strains revealed an RFLP pattern (designated 2.6) specific to WO-1 cells. The *SAP1* alleles and their flanking regions were cloned from WO-1 and from SS. These four alleles account for the major RFLP patterns of *SAP1* in laboratory and most clinical strains of *C. albicans*. S1 analysis demonstrated that both alleles of *SAP1* are transcribed in O cells. Analysis of the upstream flanking regions of the 2.6 allele identified a 284-bp insertion with homology to part of the *CARE2* element (14), bounded by an 8-bp repeat, suggesting that *CARE2* is indeed a transposable element. Sequence analysis of the upstream regions also identified a fourth *SAP* gene (*SAP4*) positioned in tandem with the *SAP1* gene.

MATERIALS AND METHODS

Strains, media, and plasmids. The *C. albicans* strains used were strains 3153a (ATCC 28367) and WO-1 (30), both generously provided by David Soll (University of Iowa), and strain SS (3), provided by Remo Morelli (San Francisco State University). Cultures were maintained at 25°C on YEPD medium (10 g of yeast extract, 20 g of peptone, and 20 g of dextrose per liter) and transferred weekly. The WO-1 strain was maintained separately as both W and O phenotypes on YEPD plates containing phyloxine B (5 μg/ml). Cultures were grown in YCB-BSA medium (23.4 g of yeast carbon base, 2 g of yeast extract, and 4 g of BSA per liter, pH 5.0, filter sterilized) to induce proteinase transcription. The *Escherichia coli* Bluescript plasmid vector (Stratagene Inc., La Jolla, Calif.) was used to subclone the *SAP1* alleles and was transformed into *E. coli* XL1-Blue for sequencing.

SAPAS probe. The DNA probe used for screening genomic libraries and for performing Southern analyses was a 655-bp internal fragment of the *SAP1* gene, amplified from strain SS genomic DNA by PCR (27). Oligonucleotide primers derived from the *SAP1* sequence published by Hube et al. (11) were used to amplify the internal *SAP1* fragment. These primers were SM-1 (5′-CAA GCC ATC CCA GTT ACT TTA AA-3′) and SM-2 (5′-CCA GAA TCT AAA AGA ACA TCG ATA TTA CC-3′). This probe contains both active-site domains and extends from nucleotide 277 to nucleotide 932 of the sequence (11). This fragment will be referred to as SAPAS, for *SAP1* active-site probe.

Southern blot analysis. Southern analyses of restriction enzyme-digested genomic DNAs from strains WO-1, SS, and 3153a were performed as described by Sambrook et al. (28). Restriction enzymes used included *Eco*RI, *Bam*HI, and *Kpn*I (New England Biolabs, Beverly, Mass.). We also performed Southern analyses with *Eco*RI- and *Kpn*I-digested phage DNAs isolated from genomic DNA libraries of strains WO-1 and SS as described below. All Southern blots were probed with the ^{32}P-labeled SAPAS probe.

Cloning and DNA sequence analysis of the *SAP1* alleles from WO-1 and SS. To isolate the alleles corresponding to the *SAP1* gene from both the WO-1 and SS strains, we screened libraries of genomic DNA fragments digested with *Sau*IIIA from both strains in the λ phage vector FIX II (Stratagene Inc.) and size-selected libraries of genomic DNA fragments digested with *Kpn*I and with *Bam*HI from the WO-1 strain in the plasmid vector pBluescript.

At high stringency, the SAPAS probe hybridizes exclusively to the *SAP1* alleles. Hybridization at high stringency (65°C in 3× SSC buffer [20× SSC is 3 M NaCl, 0.3 M sodium citrate, pH 7.0] with a final wash in 0.1× SSC) and low stringency (50°C in 3× SSC buffer with all washes in 3× SSC) yielded two sets of clones. Recombinant DNA was prepared from these clones. This DNA was digested separately with *Eco*RI and *Kpn*I, Southern blotted, and hybridized with the SAPAS probe. Representative *Kpn*I and *Eco*RI fragments from each group were transferred to the Bluescript plasmid vector for restriction digest mapping and sequence analysis. The nucleotide sequence of the four alleles obtained from clones that hybridized at high stringency was determined by the dideoxy chain termination method (29) with synthetic oligonucleotide primers. All other recombinant DNA procedures were carried out according to previously published protocols (28). Sequence analysis was performed with programs from the Molecular Biology Information Resource (Baylor College of Medicine, Houston, Tex.).

RNA isolation. *C. albicans* cells were grown in 5 ml of either YEPD or YCB-BSA and harvested at mid-log phase. Cells were washed in ice water and suspended in 2.5 ml of RNA lysis buffer (0.5 M NaCl, 1% sodium dodecyl sulfate [SDS], 0.2 M Tris-Cl [pH 8], 10 mM EDTA). An equal volume of phenol and 1.5 volumes of glass beads were added, and the entire mixture was vortexed at high speed for 2 min. The mixture was placed on ice for 2 min and again vortexed for 2 min. After centrifugation at 11,950 × *g* for 10 min in a Sorvall SS-34 rotor at 4°C, the aqueous phases were collected and extracted with phenol. The mixture was again centrifuged (all remaining centrifugations and manipulations were done at room temperature), and the aqueous phase was extracted with phenol-chloroform-isoamyl alcohol (25:24:1, vol/vol). After a final centrifugation for 5 min, the aqueous phase was collected and extracted a final time with chloroform-isoamyl alcohol (24:1, vol/vol). The aqueous phase was ethanol precipitated overnight at −20°C. Determinations of the A_{260} were made to assess RNA concentration.

S1 nuclease protection analyses. S1 nuclease protection analyses of allelic expression (1) were performed as described before (2), using 50 μg of total RNA. The RNA was pelleted and dried with 50,000 cpm of a 2-kb probe labeled at the 5′ end

with ^{32}P by using polynucleotide kinase. The probe consisted of an *Eco*RI fragment composed of the first 870 bp encoding the *SAP1* gene and approximately 1.2 kb of upstream sequence. This mixture was suspended in 20 μl of S1 hybridization solution [80% formamide, 0.4 M NaCl, 1.0 mM EDTA, 40 mM piperazine-*N*,*N*′-bis(2-ethanesulfonic acid) (PIPES) buffer, pH 6.4] and hybridized while submerged overnight at 37°C. Then, 300 μl of S1 reaction mix (20 U of S1 nuclease per ml, 20 μg of denatured salmon sperm DNA per ml, 4.5 mM zinc acetate, 0.280 M NaCl, 30 mM sodium acetate [pH 4.5]) was added, and the reaction mixture was incubated at 37°C for 30 min. The reaction was terminated with the addition of 77 μl of S1 stop mix (50 mM EDTA, 2.5 mM ammonium acetate, 20 μg of *E. coli* tRNA), and the mixture was precipitated overnight in 2 volumes of 100% ethanol at −20°C. The reaction mixture was electrophoresed on a 7% polyacrylamide gel, dried, and autoradiographed.

RNA transcripts representing the 2.4 and 2.6 alleles of WO-1 were synthesized to provide positive controls for the S1 analysis. The pBluescript clones were linearized by digestion with *Hin*dIII. The linearized clones were used as templates in a T3 RNA polymerase reaction. The reaction mixture, containing 2.5 μl of 10× transcription buffer (400 mM Tris-Cl [pH 7.9], 60 mM $MgCl_2$, 20 mM spermidine, 100 mM dithiothreitol, 0.5 M NaCl), 1 μg of linearized DNA template, 1 μl of RNasin (40 U/liter), 10 U of T3 RNA polymerase, and H_2O to a total volume of 25 μl, was incubated at 37°C for 30 min. Then, 75 μl of H_2O was added, and the entire mixture was spun through a P-30 column (Bio-Rad, Richmond, Calif.). The RNA concentration was determined by the A_{260}. A total of 5 ng of transcript was used in the control S1 reactions.

RESULTS

Our interest in proteinase virulence factors in *C. albicans* originated from the reported observation that an aspartyl proteinase was secreted primarily by O cells (and not by W cells) in vivo at lesional sites in the kidneys, liver, and spleen of mice experimentally infected with WO-1 cells (21). This is consistent with the results of Morrow et al. (19), who found that *SAP1* is expressed exclusively in O cells in defined and serum-induced media, and with our results (35) that *SAP1* and *SAP3* are expressed exclusively in O cells in proteinase-inducing medium. This suggested that *SAP1*, then considered the principal Sap, might be associated either with switch phenotype or the expression of virulence characteristics in *C. albicans* infections. Accordingly, we initiated an analysis of aspartyl proteinase gene expression in the WO-1 switch phenotype and two unrelated strains: SS, originally isolated from a patient with systemic candidiasis, and 3153a, a strain which possesses a different phenotypic switching system.

RFLP mapping distinguishes three alleles of *SAP1*. By using the amino acid sequence of *SAP1* reported by Hube et al. (11), a PCR-derived probe (SAPAS) for the *SAP1* gene was synthesized as described in Materials and Methods. Southern blots containing genomic DNA isolated from WO-1, 3153a, and SS cells and digested with the restriction endonuclease *Kpn*I detected two Southern patterns specific to the SAPAS probe (Fig. 1). The first hybridization pattern, common to both the SS and 3153a strains, consisted of *Kpn*I fragments of 2.8, 2.4, and 0.4 kb; the second pattern was found only in the WO-1 strain and consisted of *Kpn*I fragments of 2.6, 2.4, and 0.4 kb. Three distinct *SAP1* alleles could be identified from this Southern analysis by the presence or absence of *Kpn*I sites (Fig. 2A). For convenience, these three classes of alleles will be referred to by the sizes of their largest *Kpn*I fragment as the 2.4 allele from WO-1, SS, and 3153a, the 2.6 allele from WO-1, and the 2.8 allele from SS and 3153a. The same three alleles were consistently identified in clinical isolates of *C. albicans* (34). In Fig. 1, the differences in hybridization signals between the alleles are due to the length of the probe, which hybridizes to each *Kpn*I fragment (Fig. 2A). Thus, the 2.8-kb *Kpn*I fragment hybridizes more intensely in Fig. 1 because the SAPAS probe hybridizes to a longer region of the 2.8-kb *Kpn*I fragment than it does to the 2.6- or 2.4-kb *Kpn*I fragments. The allelic pattern for strain WO-1 was the same for DNA isolated from either W or O cells (34). Further restriction analyses with *Hin*dIII, *Eco*RI, *Eco*RV, *Bgl*II, *Bam*HI, *Cla*I, *Kpn*I, and *Xmn*I (34) established that all fragments represent polymorphisms of the single-copy *SAP1* gene in *C. albicans*.

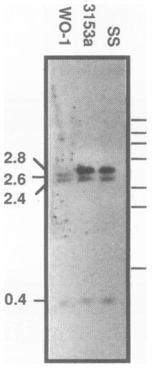

FIG. 1. RFLP mapping distinguishes three alleles in three laboratory strains. Southern analysis of *Kpn*I-digested genomic DNAs from strains SS, 3153a, and WO-1 hybridized with the ^{32}P-labeled SAPAS probe. Genomic DNAs (10 μg) from the three strains were digested with *Kpn*I, electrophoresed on a 0.8% agarose gel, and blotted to nitrocellulose. The blot was hybridized at high stringency. The genomic DNA used in each lane is indicated above the lane. Numbers to the left of the panel indicate sizes of the genomic fragments in kilobases. Phage λ size marker positions are indicated by lines on the right of the panel. The size of each band (from the top) is 23.1, 9.4, 6.7, 4.3, 2.3, 2.0, or 0.5 kb. The intensity of hybridization of each band is related to the length of the fragment that hybridizes to the SAPAS probe (see Fig. 2A).

DNA sequence analysis of *SAP1* alleles and flanking regions. To understand the molecular signals that might be operating in the differential transcription of the *SAP1* gene in each of the laboratory strains, genomic DNA libraries prepared from either strain WO-1 or SS were screened with the SAPAS probe at high stringency. Several clones encoding sequences homologous to SAPAS were isolated from each strain and characterized in detail. The SAP regions of λ clones which correspond to the 2.4 and 2.6 alleles of WO-1 and the 2.4 and 2.8 alleles of SS were sequenced. Tables 1 and 2 summarize the differences among these four alleles of *SAP1*. Within the coding regions of the *SAP1* alleles (Table 1), there are very few sequence differences. Of these, only the Tyr-to-Asn change at position 391 in the 2.6 allele of WO-1 represents a substitution that alters the nature of the amino acid residue; the remaining differences result mostly in silent or conservative changes. Approximately 1,600 bases of the 2.6 allele nucleotide sequence upstream of the gene are shown in Fig. 3. Table 2 compares this sequence with that upstream of the remaining three alleles as well as approximately 140 bases of 3′-flanking sequence. We have numbered the nucleotides of the *SAP1* gene and flanking regions according to conventional standards, with the A of the initiation AUG numbered 1. For ease of comparison, we have included the numbers (where available)

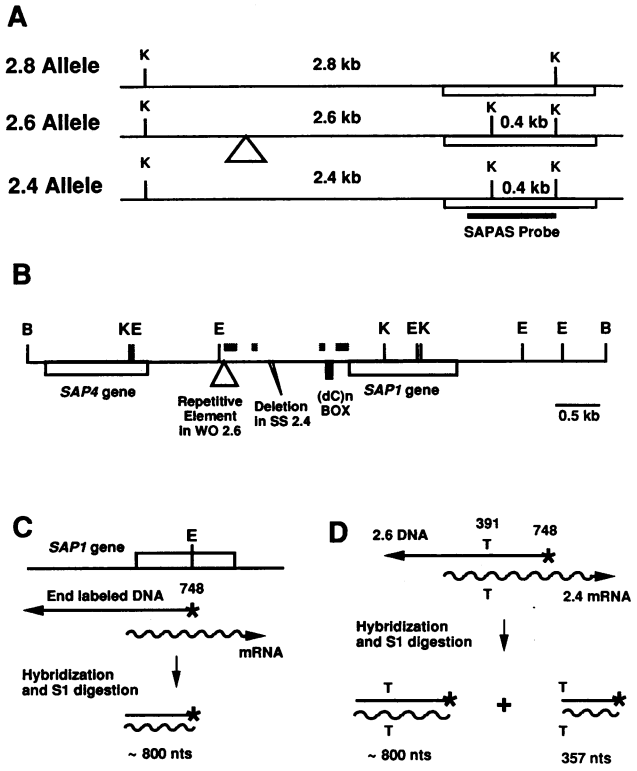

FIG. 2. Restriction maps and summary of major allelic differences. (A) Schematic drawing of the *Kpn*I (K) restriction fragments which are used to identify the three *SAP1* alleles. The position of the SAPAS probe relative to these fragments is shown as a bar below the three allele maps. The 2.6 allele contains an insert located approximately 1,255 nucleotides upstream of the gene. (B) Major insertions and deletions in the *SAP1* alleles and a limited restriction map of the region which includes *SAP1* and *SAP4* (to scale). The bar labeled (dC)n box shows the position of the series of deoxycytidines, which vary in number between alleles. DNA secondary structures in the 5′-flanking regions of *SAP1* include two short inverted repeats (−1024 to −1044 and −234 to −255) and two large inverted repeats (−13 to −147 and −1218 to −1590). All inverted repeat pairs are shown by shaded boxes. B, *Bam*HI; E, *Eco*RI; K, *Kpn*I. (C) Schematic outline of the S1 analysis. Cloned DNA was digested with *Eco*RI (E in figure at position 748 in the *SAP1* gene) and 5′ end labeled with polynucleotide kinase. Total RNA, which contains *SAP1* mRNA, was hybridized to the labeled DNA. After hybridization, the DNA-RNA hybrids were digested with S1 nuclease, which degrades single-stranded DNA and RNA, leaving the DNA-RNA hybrid. The resulting end-labeled DNA fragment (~800 nucleotides [nts] in length), which is detected on denaturing polyacrylamide gels, represents the distance from the internal label at the *Eco*RI site to the start of transcription (the 5′ end of the mRNA). Arrowheads on DNA and RNA indicate the 5′-to-3′ direction of the strand. (D) Schematic drawing of an S1 experiment with mismatches. End-labeled DNA from a clone of the 2.6 allele contains a T at position 391 (based on positions in Table 1). RNA from allele 2.4 contains a T at position 391. When hybridized, the DNA-RNA hybrid contains a mismatch at position 391, since the two T residues cannot form a base pair. This mismatch is detected by S1, although at a low frequency. Therefore, most of the DNA-RNA hybrids are ~800 nucleotides long, as expected. However, a small proportion will be cleaved at the T-T mismatch by S1 nuclease, resulting in a labeled fragment of 357 nucleotides. Similar diagrams can be constructed for DNA from the 2.4 allele hybridized with RNA from the 2.6 allele and for each of the nucleotide positions in which the 2.4 and 2.6 alleles differ in sequence (see Table 1).

TABLE 1. Differences between coding regions in alleles from three strains of *C. albicans*

Position[a]	Nucleotide					Coding change
	10231	WO-1 2.6 allele	WO-1 2.4 allele	SS 2.8 allele	SS 2.4 allele	
21 (147)	C	C	C	T	C	Silent
24 (150)	C	T	T	T	T	Silent
157 (283)	A	C	C	A	C	Ile/Leu
159 (285)	C	C	C	T	C	
210 (336)	T	T	A	T	T	Silent
383 (509)	G	C	C	C	C	Ser/Thr
384 (510)	C	C	C	T	C	
390 (516)	T	C	C	C	C	Silent
391 (517)	T	A	T	T	T	Tyr/Asn
444 (570)	T	C	C	T	T	Silent
480 (606)	C	T	C	C	C	Silent
663 (789)	C	T	T	T	T	Silent
882 (1008)	T	C	C	C	T	Silent
956 (1082)	C	C	T	C	C	Ala/Val
981 (1107)	A	G	G	G	G	Silent

[a] Numbering system based on sequence of WO-1 2.6 allele; corresponds to numbering in Fig. 3. Numbers in parentheses represent the corresponding nucleotides in the sequence reported by Hube et al. for ATCC 10231 (11).

of these nucleotides as they correspond to the *SAP1* sequence as published by Hube et al. (11).

It is remarkable that there are very few significant differences in the untranscribed 5′- and 3′-flanking regions between WO-1 and SS. The 3′-flanking regions are essentially identical over 140 nucleotides. In the 5′-flanking regions, with the exception of the insertion at positions −1539 to −1255 of the 2.6 allele of WO-1, there are no nucleotide changes that are found uniquely in the WO-1 alleles which might encode *cis*-acting regulatory elements for *SAP1* expression in the WO-1 and not the SS strain, in which *SAP1* expression has not been detected (35). In the putative promoter regions, at positions −166 to −174, there are a variable number of C residues, 9 and 11 in the WO-1 2.6 and 2.4 alleles, respectively, and 23 (CT-21C) and 26 in the SS 2.8 and 2.4 alleles, respectively. This duplication of a C tract 5′ of the SS alleles is the only region proximal to the promoter that may be significant and of interest for functional testing for its effect upon transcription.

The 284-bp insert found only in the 2.6 allele of WO-1 is located 1,255 bp upstream of the start of translation. A GenBank search with this insert sequence revealed that over a span of 200 bp, there was 94% homology with a repetitive *C. albicans* DNA sequence, *CARE2*, whose function is unknown (14). The 284-bp insert upstream of the 2.6 allele is flanked at both ends by an 8-bp direct repeat of the presumed target sequence at the site of insertion.

Each of the four 5′-flanking sequences was used to search the transcription factor data base (9) with the FASTA program (20); no possible binding sites were identified for known

TABLE 2. Sequence differences between flanking regions of alleles from three strains of *C. albicans*

Position[a]	10231	WO-1 2.6 allele	WO-1 2.4 allele	SS 2.8 allele	SS 2.4 allele
5′-Flanking region					
−1562		A	TT	T	TT
−1553		C	G	G	G
−1539 to −1255[b]		Insert	—	—	—
−1251		T	C	C	C
−1212		T	C	C	T
−1107		G	A	A	A
−1102		C	A	A	A
−858 to −875[b]		—	—	—	Deleted
−588		A	A	G	A
−489		A	G	G	G
−166 to −174[c]		9C	11C	CT-21C	26C
−133		T	T	G	T
−110 (17)	G	G	G	G	A
−23 (104)	T	C	C	C	C
3′-Flanking region					
1270 (1396)	T	T	T	C	C
1272 (1398)	T	C	C	C	C

[a] Numbering system is based on sequence of WO-1 2.6 allele and corresponds to numbering in Fig. 3. Numbers in parentheses represent the corresponding nucleotides in the sequence reported by Hube et al. for strain 10231 (11). The sequence of strain 10231 as reported by Hube et al. (11) extends only to position −126.

[b] Dashes represent allelic sequences which do not contain insertions or deletions.

[c] Values represent poly(C) tracts of the indicated lengths. CT-21C represents a CT dinucleotide followed by 21 C nucleotides.

transcription factors over the entire 5′-flanking region. Figure 2B presents a limited restriction map and compilation of the major differences in the approximately 2.6 kb of sequence determined for each of the four alleles and their flanking regions.

Transcription of both 2.4 and 2.6 alleles of strain WO-1 in proteinase-inducing medium. After comparison of the upstream sequences of the four different *SAP1* alleles, the flanking region of the 2.6 allele was the most distinct. We therefore asked whether the differences between the 5′-flanking regions of the 2.6 and 2.4 alleles of strain WO-1 might affect the transcription of *SAP1* in these cells. We took advantage of several base pair differences in the coding regions of the 2.4 and 2.6 alleles of WO-1 (see Table 1) to use S1 analysis (diagrammed in Fig. 2C and D) to determine which *SAP1* alleles of strain WO-1 are transcribed for *SAP1* gene expression.

In this experiment, DNA probes representing the 2.6 and 2.4 alleles of WO-1 were labeled with ^{32}P at an *Eco*RI site within the coding region of the *SAP1* gene (Fig. 2C). These DNA probes extend into the 5′-flanking regions of each allele. Each probe was hybridized to total RNA from O cells grown in the proteinase-inducing medium YCB-BSA. Hybridization of the probes to synthetic transcripts of the 2.6 and 2.4 alleles of WO-1 provided controls. The hybridization mixtures were digested with S1 nuclease under conditions in which only DNA-RNA hybrids survive. The length of the largest resulting DNA fragment represents the distance from the 5′ end of the RNA (the start of transcription) to the internal label (see Fig. 2C). The result is shown in Fig. 4 as the intense band of approximately 800 nucleotides in lanes 3 and 7. The bands near the top of the control lanes (lanes 1, 2, 5, and 6) most likely represent breathing of the DNA-RNA hybrid under S1 conditions and/or nuclease degradation of the control transcripts.

S1 nuclease can also attack and cleave at mismatches in the RNA-DNA hybrid (see Fig. 2D), although cleavage occurs at a low frequency. The length of these shorter DNA fragments

```
-1593  GAATTCTCATTCGAGTTTTT CTTTTTAGTTTATTTTTTTC CGAATTATATTTTGTTTGTA TACACGGAATATTCGTGAGA ATTAGTCCTAGTACACGATA
-1493  GGTGATAATTTGTAATAAGC AAAAAACAGAGTGAATGAAG TTTACGCCCACTAAAAGTGT CTCCTGAAACTCGGTCCCTA TATAATAAATATCCTAAATA
-1393  TCTTTCTTTTCTCTTCTTTT TATACTAATTCCTGGCTCTC GATGAGTGCAAATTTTTCAC CAATGTTGACCTAATACGCA CGACTTATATCAAGGTGTTA
-1293  TTTCCCTGACACAAACGTCG CATAAACCAACATATTTTGC TGTAAAAACTCTTAAAACGA AAACCATCAAAAAGAAACAA GTAATTACACCAATCACCGC
-1193  GTACATTTCTTTATGATTAT ACAACTACTTTCCCCAATAT GTTGTCAAGTATTGTTACTA TAATTAAAATATGAGCTACA TGAAAGGATTTCTATTGTCT
-1093  TGGTGGGTGTTATGAATCAA GTGCTGAGGTGATTTAATTT GCATCTAACATTAGTACCCC ATTACTAATATACAACTTTA AATGAAAAAGGAGGTTATTT
 -993  TACACTTATTTCAAGTTTAA AAAGAAGTGGGGATTGAAGA GGAATTTGATATTATACTTT TGCCACAATATTCTCAATTA TGATTTGGTTACGGAAAATA
 -893  AAGAAGATGGCCCTTAATCG TGAAAGCCACTTTAATTTGA CAAAAGACAACAAGCAACAG TATAAGAATCTCGTTGTAGT TCAATAATTACAATAGAAAA
 -793  ATGTGGCTTTCAAAAATAGA AACAGCAATTTTAAACAAAC GTTTTCTGTGTATGTCAATA AGTACAACTATAATAGCTAA ATGTTCATTATACTATAATA
 -693  TTTCATGATGATAACAATAG AAGGTTTGGAACATTGATGA ACAAAGTACTATTTTTAGTC GTGTATTTTTGCGCGCATTT TCAAATCTTGATTCTGTTGC
 -593  ATTCAATGTTACATTGGAAA TCTTTATCTCCATTTTCTGT TCCATACTTCTCTTGAAGAA CTTGTCGCTAAAAGATTTGA TTCAAAGATATACCTTTAAT
 -493  TTCGATCACATATAACTACA GATTGTATTTAAGTAACTAC TGTCCTCAAAAGATGGATTA TCAAATAATGGTAGTTCCTA TTTTTAGTTTTGGTTTTAGT
 -393  ACTGTGGACATGATTGAAAT GGAATTATACTGATAAAAAG TGGTTCTTGACACAAACTGA ATGAAAGCAATATTCAATAA TTTCACATCATAACCATTAT
 -293  CAACAACTATACGGCACCAT GTTATTCAAAAAAACTAACC AGTTATCGCGTTATAACTGG GAGGGGAGAATGTAAAATTA ATCAAATTGTGCTATCTTTT
 -193  AAAAATGGTACACCTCCTTC CCCCCCCCTATAGCTTTTGC TGGTTGATGCCATACTCAAA TGGATAATATTCTGATGGAT AATGTTATATCTTGAAAAAC
  -93  ATATAAATATGGGAGTTGGA TCTATAACTTTATTGAAATA AATCATATTTAATCCAACAA TCAATCAATTCACTCTTCCA TTTCTAACAAACAATG
```

FIG. 3. Sequence of the 5′-flanking region of the WO-1 2.6 allele. The sequence ends at the ATG (in large letters) at the start of the coding region. The underlined deoxycytidines present at positions −166 to −174 vary in number among the different alleles. The underlined sequence beginning at positions −857 to −875 is deleted only from the SS 2.4 allele. The 284-bp insert present in only the WO-1 2.6 allele is underlined and spans positions −1568 to −1255. The double lines above the sequence indicate the 8-bp direct repeats (ATATTTTG) which flank the 284-bp insert. The completed sequences for all four alleles of *SAP1* have been deposited in GenBank under accession numbers L12449 through L12452.

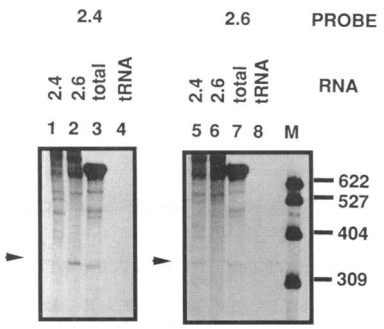

FIG. 4. Both the 2.4 and the 2.6 *SAP1* alleles of strain WO-1 are transcribed in RNAs from O cells grown in YCB-BSA. Total RNA was hybridized with a ^{32}P-end-labeled DNA 2-kb probe representing either the 2.4 or the 2.6 allele of WO-1 cells. In control experiments, transcripts of the two alleles synthesized by using T3 polymerase from cloned fragments were hybridized with the labeled DNAs. After incubation with S1 nuclease, the reaction mixtures were electrophoresed on a 7% polyacrylamide–urea gel; the gel was dried and exposed to X-ray film. Lanes 1 to 4 contain ^{32}P-labeled DNA from the 2.4 allele. Lanes 5 to 8 contain ^{32}P-labeled DNA from the 2.6 allele. Lanes 1 and 5 were hybridized to unlabeled T3 transcripts representing the 2.4 allele. Lanes 2 and 6 were hybridized to unlabeled T3 transcripts representing the 2.6 allele. Lanes 3 and 7 were hybridized to total RNA from O cells grown in YCB-BSA. Lanes 4 and 8 were hybridized to *E. coli* tRNA as a control. Lane M represents pBR322 marker DNA digested with *Msp*I and labeled with ^{32}P; the sizes of the marker bands (in base pairs) are listed on the right. Arrows point out the mismatches at position 357, indicating that both alleles are transcribed in total RNA from O cells grown in YCB-BSA. The slight size difference between bands in the total RNA lanes and the control transcript lanes is due to the different amounts of RNA loaded into each lane (50 μg of total RNA versus 50 ng of unlabeled control transcript).

(Fig. 4, lanes 3 and 7) represents the distance from the mismatch to the internal label. From DNA sequence analysis (Table 1), there are three mismatches between the 2.4 and 2.6 alleles of strain WO-1 upstream of the labeled *Eco*RI site (position 748). These include positions 210, 391, and 480, which are expected to generate labeled fragments of 538, 357, and 268 nucleotides, respectively, as the labeled *Eco*RI site is present at position 748. All three mismatches are present on one strand of RNA encoded by the other allele, suggesting that the shortest fragment should be the only fragment observed. However, S1 nuclease cleaves 1-bp mismatches at very low frequencies. Therefore, under ideal conditions, bands representing all three fragments should be present. However, the sequence surrounding the mismatch and the S1 conditions influence the recognition of 1-bp mismatches. The expected 538- and 268-nucleotide bands were not observed, even in the control transcript lanes. However, the 357-nucleotide band was observed (Fig. 4, arrows, lanes 2 and 5) when the 2.4 allele DNA probe was hybridized with the 2.6 allele T3 transcript and when the 2.6 allele DNA probe was hybridized with the 2.4 allele T3 transcript. The presence of bands of the same size in lanes 3 and 7 of Fig. 4 demonstrates that mRNA from both the 2.4 and 2.6 alleles of WO-1 is present in total RNA, and therefore, both alleles are transcribed. There is a slight size difference between the bands seen in the total RNA lanes (lanes 3 and 7) and the control transcript lanes (lanes 2 and 5). This difference is a common problem and is due to the different amounts of RNA loaded into each lane (50 μg of total RNA versus 50 ng of unlabeled control transcript), which result in a slight alteration in the mobility of the banding pattern. The difference cannot be corrected by running carrier RNA in the control lanes (34).

Identification and sequencing of the *SAP4* gene. During the subcloning and sequencing of fragments upstream of the *SAP1* gene, we identified a fourth *SAP* gene (see restriction map [Fig. 2B] and complete sequence [Fig. 5]). A comparison of the predicted protein product of the *SAP4* gene with the other *SAP* gene products (Fig. 6A) shows that the mature proteins are highly conserved in size and in sequence, especially around the two portions of the active site. The hydrophobic signal sequences in the precursor proteins of the four gene products are also highly conserved (first 20 amino acids). The most divergent region of the four genes is the region 3′ of the signal sequences and immediately 5′ of the start of the mature proteins; *SAP4* has a 24-amino-acid insertion, which includes two additional Lys-Arg dipeptide residues which may be involved in processing of this extended precursor.

A phylogenetic tree (Fig. 6B) shows that Sap1 and Sap2 are the most closely related Sap proteins and that the *SAP4* gene product is the most divergent Sap identified to date. Southern analysis shows that the *SAP4* gene is present upstream of the *SAP1* gene in all laboratory and clinical strains tested (34). To date, Northern (RNA blot) analysis has not detected the expression of *SAP4* in any cell line or switch system under any of a variety of growth conditions (34).

DISCUSSION

Restriction endonuclease and DNA sequence analysis of the three *C. albicans* strains tested here revealed an RFLP pattern which differentiated among three different allelic forms of *SAP1*. The two allelic forms 2.4 and 2.8 differ in a single nucleotide change which results in the presence or absence of a *Kpn*I site within the gene. The third allelic form, 2.6, is the result of a transposable element insertion 1,255 nucleotides upstream of the gene.

The observation that WO-1 cells have a unique *SAP1* allele, 2.6, coupled with the demonstration that *SAP1* is expressed only in O cells which contain this allele (19, 35), led us to explore the possibility that the pattern of Sap1 expression in WO-1 cells might be related specifically to a feature of the 2.6 allele or its flanking regions. While this hypothesis was proven incorrect, the cloning, sequencing, and comparison of these four *SAP1* alleles and their flanking regions from the WO-1 and SS strains of *C. albicans* have resulted in many interesting observations.

The regions encoding the *SAP1* gene alleles and their 5′- and 3′-flanking regions are highly homologous. The Tyr-to-Asn change at position 391 of the 2.6 allele aside, the structural gene sequences are essentially identical except for minor conservative or silent mutations in sequence. The sequence similarity of all four coding regions of the *SAP1* alleles establishes that these are each true alleles of *SAP1* and not pseudogenes or other closely related aspartyl proteinases. Despite the high degree of homology between these alleles, the DNA sequence information obtained allowed us to design probes which differentiate *SAP1* alleles directly by S1 endonuclease protection analysis. Using these probes, we were able to show that both the 2.6 and 2.4 alleles of *SAP1* are transcribed in O cells grown in proteinase-inducing medium.

Among the sequences 5′ of the ATG for approximately 200 nucleotides, there are several potential TATAA-like sequence elements which may serve as promoters (Fig. 3). Nothing is known about spacing or sequence requirements for functional TATAA elements in *C. albicans*. However, in the related yeast *Saccharomyces cerevisiae*, TATAA elements in the regulatory

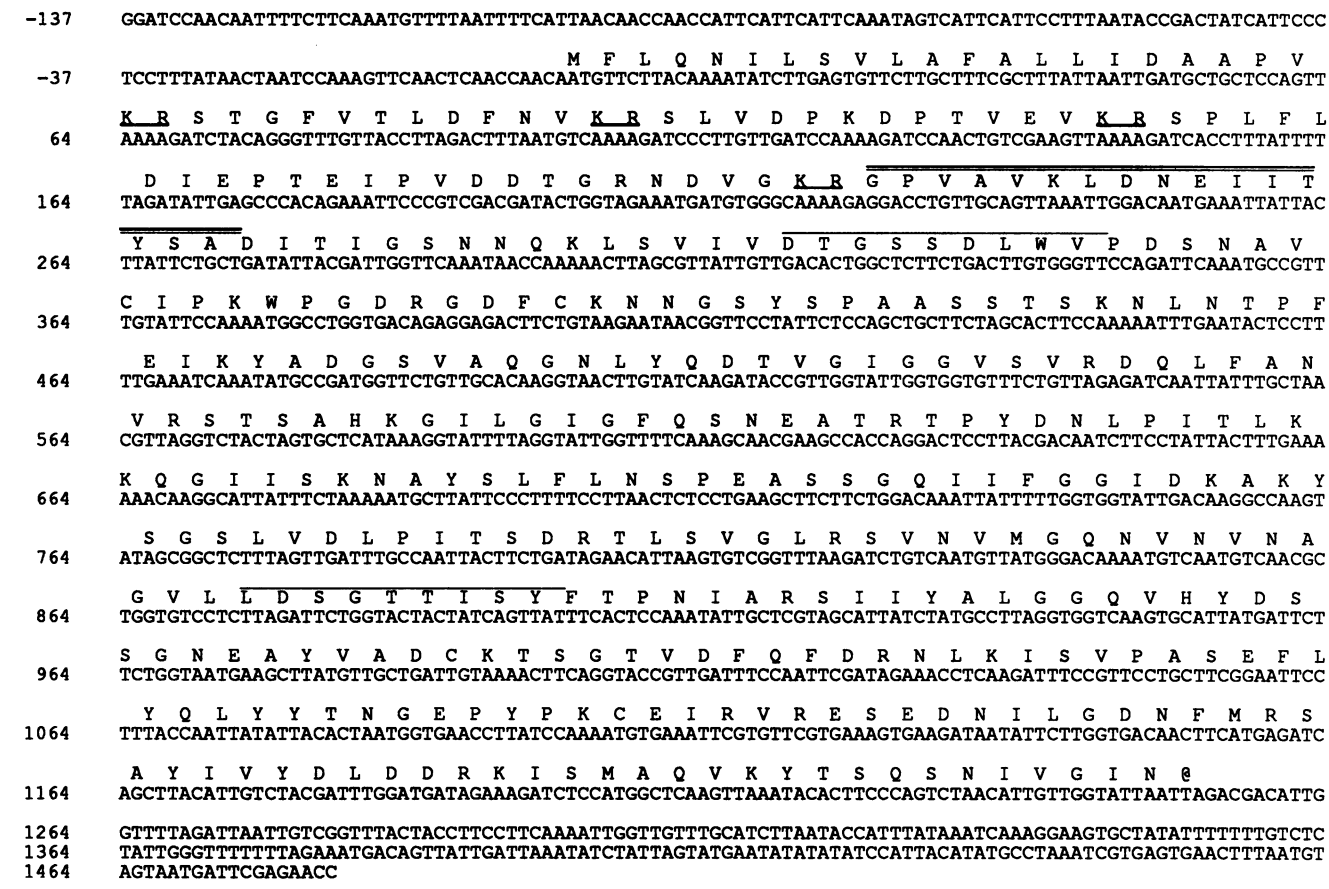

FIG. 5. Sequence of the *SAP4* gene and deduced amino acid sequence of the protein. The sequence begins at a *Bam*HI site shown in Fig. 2B. The putative N terminus of the mature protein, which is highly conserved with the N termini of the other *SAP* proteins, has a double line above the sequence. The two sections of the active site of the proteinase have a single line above the sequence. The four Lys-Arg dipeptides in the precursor peptide are underlined. The entire sequence is 1,618 nucleotides in length. The GenBank accession number is L25388.

regions of genes are commonly found anywhere between 40 and 120 bp upstream of the mRNA initiation site. We note two potential TATAA-like elements at −92 and −71. Intergenic sequences in the *C. albicans* genome are A-T rich, and thus identification of putative promoter sequences requires functional analysis. In the RNA 5′ to the AUG, the *SAP1* sequence 5′-AAACAAUGUUU-3′ has some homology with the consensus *S. cerevisiae* translation initiator sequence 5′-$\overset{A}{Y}A\overset{A}{U}$AAAU GUCU-3′, where Y is C or U. Two small differences, a C for the A or U at position −3 and a U replacing the C at +5, should not be enough to cause a major alteration in translational efficiency, since studies of mutations both 5′ and 3′ to the AUG of the *HIS4* gene in *S. cerevisiae* showed no more than a twofold decrease in translation caused by the change in context (5).

Both the 2.6 and 2.4 alleles are transcribed in WO-1, while no *SAP1* expression is found in strain SS (35). Since the differences between the WO-1 alleles and the SS alleles are overlapping, we surmise that differences in the upstream flanking regions are not important in the differential transcription of *SAP1* in these two strains. The only striking difference in sequence is found at position −166, where each of the alleles contains a poly(C) tract varying from 9 to 26 nucleotides in length. Those in the SS strain are over twice as long as those in WO-1 and represent the only candidate region where any significant differences in potential *cis*-acting regulatory sequences might exist.

Approximately 1.1 kb of DNA from a different strain of *C. albicans* has been sequenced and designated *CARE2* (*C. albicans* repetitive element). Regions of this sequence appear to be mobile, as determined from alternating genomic hybridization patterns of the element in different isolates (14). Analysis of the DNA sequence upstream of the 2.6 allele revealed a 284-bp insertion at positions −1539 to −1255. The center of this 284-bp sequence is highly homologous to a region of the 1.1 kb of DNA characterized as *CARE2*. However, large regions at the ends of the 284-bp sequence have diverged from the *CARE2* sequence. The 284-bp sequence is flanked by 8-bp direct repeats. The eight nucleotides of direct repeats are present at the same position in each of the other alleles sequenced, indicating that this is a typical duplication of a target sequence characteristic of transposable elements. A sequence arrangement of several hundred nucleotides bounded by short direct repeats is reminiscent of delta insertion elements in *S. cerevisiae*. The delta elements of *S. cerevisiae* are found at both ends of the TY1 transposons; during transposition, when a TY1 element is excised by recombination, a delta element is often left behind, bounded by direct repeats of the target sequence (22). From these observations, it would appear that the 284-bp sequence found upstream of

A

Precursor Peptides:

```
sap1  MFLKNIFIALAIALLVDASP..AKRSPGFVTLDFDVIK.....................T....PV.NATG.QEG....K.VKR
sap2  ------------------T-ttT--A---A---S-V-.....................-pkaf--T-..-.---....KTS--
sap3  ---------------A--T-ttFNN-----A-N-----.....................-hk.....-V--P-GEintnVN---
sap4  ---Q--LSV--F---I--A-..V--T-------N-.-rslvdpkdptvevkrsplfldiep-ei..--DD.--......rnDVG--
```

Mature Proteins:

```
sap1  QAIPVTLNNEHVSYAADITIGSNKQKFNVIVDTGSSDLWVPDASVTCDKPRPGQSADFCKGKGIYTPKSSTTSQNLGSPFYIGYGDGSSS
sap2  --V----H--Q-T------V---N--L---------------VN-D-QVTYSD-T-----Q--T-D-SG-SA--D-NT--K---------
sap3  -TV--K-I--Q----S---V------LT-VI-----------SQ-S-QAG.Q--DPN---NE-T-S-S--SS----N---S-E----TT-
sap4  GPVA-K-D--IIT-S--------N--LS--------------SNAV-IPKW--DRG----NN-S-S-AA-S--K--NT--E-K-A---VQ

sap1  SQGTLYKDTVGFGGASITKQVFADITKTSIPQGILGIGYKTNEA.AGDYDNVPVTLKNQGVIAKNAYSLYLNSPNAATGQIIFGGVDKAK
sap2  ---------------V--KN--L--VDS---D-----V--------.G-S---------K----------------D------------N--
sap3  ----W----I----I-----Q---V-S--VD----------H--.E-N------------I-S----------RQ-TS---------N--
sap4  A--N--Q----I--V-VRD-L--NVRS--AHK------FQS---tRTP---L-I---K--I-S------F----E-SS-------I----

sap1  YSGSLIAVPVTSDRELRITLNSLKAVGKNIN.GNIDVLLDSGTTITYLQQDVAQDIIDAFQAELKLDGQGHTFYVTDCQTSGTVDFNFDN
sap2  -------L----------S-G-VEVS--T--tD-V----------------L-DQ--K--NGK-TQ-SN-NS--EV--NL--D-V---SK
sap3  ---T---L-----N----H--TV-VA-QS--.ADV---------------G--DQV-S--NGQETY-AN-NL--LV--NL--S---A--K
sap4  -----VDL-I----T-SVG-R-VNVM-Q-V-.V-AG---------S-FTPNI-RS--Y-LGGQVHY-SS-NEA--A--K-------Q--R

sap1  NAKISVPASEFTAPLSYANGQPYPKCQLLLGISDANILGDNFLRSAYLVYDLDDDKISLAQVKYTSASNIAALT
sap2  -----------A-S-QGDD----D-----FDVN--------------I------NE------------S-S---
sap3  ---------------YTED--V-DQ----F-T--Y------------I------NE---------T--------
sap4  -L---------LYQ-Y-T--E-----EIRVRE-ED-------M----I------R---M-------Q---VGIN
```

B

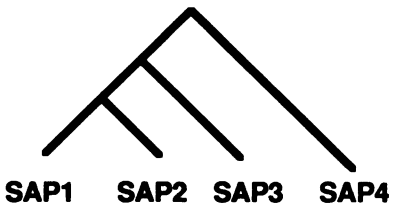

FIG. 6. Comparison of the deduced protein sequences of the four *SAP* genes. (A) Alignment of protein sequences. A dash represents a residue identical to that in the *SAP1* sequence on the top. A dot indicates a space in the sequence introduced to maximally align the sequences. Lys-Arg dipeptides are underlined in the sequences of the precursor peptides. The sections of the active site in the mature protein have a line above the sequence. (B) Phylogenetic tree of the mature portions of the four Sap proteins constructed by using the Feng and Doolittle program (7), based on the protein sequences. A similar tree is constructed from an alignment of the DNA sequences (34). The lengths of the lines are proportional to the estimated genetic distances.

the 2.6 allele represents an analogous solo delta-type sequence in *C. albicans*. In addition to strain WO-1, three clinical isolates have been found to contain the 2.6 allele polymorphism (34). In one isolate, oligonucleotides from the 284-bp *CARE2* sequence hybridize with the 2.6-kb *Kpn*I fragment from this clinical strain, indicating that a similar insertion occurs upstream of *SAP1* in an independent *C. albicans* isolate (34). The occurrence of this insertion at the site of these virulence factors raises the possibility that, like other transposable elements, a *CARE2*-like element might contribute to the regulation of expression of *SAP1*.

Assuming that the alteration in the poly(C) tract at positions −166 to −174 is not involved in the differential transcription of *SAP1*, what then can we surmise about the potential regulation of *SAP1* expression in strains WO-1, 3153a, and SS? Since the 5′- and 3′-flanking regions are essentially identical in two different *C. albicans* isolates (WO-1 and SS), one of which expresses the gene (WO-1 [19, 35]) and the other of which does not (35), one must conclude that there are different *trans*-acting elements provided by WO-1 and SS which interact with similar *cis*-acting elements upstream of *SAP1*. These could be either transcriptional activators, being expressed only in O cells, or, conversely, transcriptional repressors expressed in W cells and in other strains. In addition to the WO switching system, at least three different switching systems have been described for *C. albicans* (31). One interpretation of the *SAP1* expression studies (19, 35) is that SS and 3153a cells remained in a switch phase which was not permissive for *SAP1* expression, similar to the W cell phenotype. Presumably, if a phenotypic switch could be induced in strains 3153a and SS under comparable conditions, then the *SAP1* alleles would be transcribed.

The tandem linkage of the *SAP1* and *SAP4* genes may be the result of a tandem duplication of a *SAP* gene on that chromosome. Close inspection of the four protein sequences in Fig. 6A and the phylogenetic tree (Fig. 6B) shows that *SAP4* is the most divergent of the four proteins, suggesting that *SAP4* is not a recent duplication of *SAP1*.

Recently, the existence of another *SAP* gene was inferred from a hybridization signal to chromosome 3 on a chromosome blot (16). Signals on chromosomes 6 and R were attributed to the presence of *SAP1* and *SAP2* on their respective chromosomes. The presence of *SAP4* linked in tandem with *SAP1* raises the possibility that still other *SAP* genes are linked in tandem with the already identified *SAP1*, *SAP2*, and *SAP3* genes. Our preliminary hybridizations and sequencing of cloned flanking regions have not identified any further *SAP* genes. However, at least one more *SAP* gene remains to be identified—the gene which encodes a protein containing the N-terminal sequence GTVQTSLINE which was identified by Morrison et al. (17).

ACKNOWLEDGMENTS

We thank David Soll, University of Iowa, for many helpful discussions and for providing strains WO-1 and 3153a; Remo Morelli, San Francisco State University, for providing strain SS; our colleagues in the laboratory for many helpful comments; and Evangeline Leash for editing the manuscript.

This work was supported in part by NIH grant 1RO1AI-33317, NIH grant PO1-DE-07946, and John D. and Catherine T. MacArthur Foundation Award 84-85 to N.A.; by University of California Universitywide AIDS Research Program grant K92-SF-011 from the State of California to T.C.W.; and by NIH Dentist-Scientist Award 5K15DE00334 to S.H.M.

REFERENCES

1. **Berk, A. J., and P. A. Sharp.** 1977. Sizing and mapping of early adenovirus mRNAs by gel electrophoresis of S1 endonuclease digested hybrids. Cell **12**:721–732.
2. **Bernards, A., L. H. T. van der Ploeg, A. C. C. Frosch, P. Borst, J. C. Boothroyd, S. Coleman, and G. A. M. Cross.** 1981. Activation of trypanosome surface glycoprotein genes involves a duplication-transposition leading to an altered 3′ end. Cell **27**:497–505.
3. **Blagoyevich, P. G.** 1979. Master's thesis. San Francisco State University, San Francisco, Calif.
4. **Borg, M., and R. Ruchel.** 1988. Expression of extracellular acid proteinase by proteolytic *Candida* spp. during experimental infection of oral mucosa. Infect. Immun. **56**:626–631.
5. **Cigan, A. M., E. K. Pabich, and T. F. Donahue.** 1988. Mutational analysis of the *HIS4* translational initiator region in *Saccharomyces cerevisiae*. Mol. Cell. Biol. **8**:2964–2975.
6. **Cutler, J. E.** 1991. Putative virulence factors of *Candida albicans*. Annu. Rev. Microbiol. **45**:187–218.
7. **Feng, D.-F., and R. F. Doolittle.** 1987. Progressive sequence alignment as a prerequisite to correct phylogenetic trees. J. Mol. Evol. **25**:351–360.
8. **Ghannoum, M., and K. Abu Elteen.** 1986. Correlative relationship between proteinase production, adherence, and pathogenicity of various strains of *Candida albicans*. Infect. Immun. **24**:407–413.
9. **Ghosh, D.** 1992. TFD: the transcription factor database. Nucleic Acids Res. **20**(Suppl.):2091–2093.
10. **Hattori, M., K. Yoshiura, M. Negi, and H. Ogawa.** 1981. Keratinolytic proteinase produced by *Candida albicans*. Sabouraudia **22**:175–183.
11. **Hube, B., C. J. Turver, F. C. Odds, H. Eiffert, G. J. Boulnois, H. Kochel, and R. Ruchel.** 1991. Sequence of the *Candida albicans* gene encoding the secretory aspartate proteinase. Med. Vet. Mycol. **29**:129–132.
12. **Kaminishi, H., Y. Hagihara, S. Hayashi, and T. Cho.** 1986. Isolation and characteristics of collagenolytic enzyme produced by *Candida albicans*. Infect. Immun. **53**:312–316.
13. **Kwon-Chung, D. J., D. Lehman, C. Good, and P. T. Magee.** 1985. Genetic evidence for role of extracellular proteinase in virulence of *Candida albicans*. Infect. Immun. **49**:571–575.
14. **Lasker, B. A., L. S. Page, T. J. Lott, and G. S. Kobayashi.** 1992. Isolation, characterization and sequencing of *Candida albicans* repetitive element 2. Gene **116**:51–57.
15. **Macdonald, F., and F. C. Odds.** 1980. Inducible proteinase of *Candida albicans* in diagnostic serology and in the pathogenesis of systemic candidosis. J. Med. Microbiol. **13**:423–435.
16. **Magee, B. B., B. Hube, R. J. Wright, P. J. Sullivan, and P. T. Magee.** 1993. The genes encoding the secreted aspartyl proteinases of *Candida albicans* constitute a family with at least three members. Infect. Immun. **61**:3240–3243.
17. **Morrison, C. J., S. F. Hurst, S. L. Bragg, R. J. Kuykendall, H. Diaz, J. Pohl, and E. Reiss.** 1993. Heterogeneity of the purified extracellular aspartyl proteinase from *Candida albicans*: characterization with monoclonal antibodies and N-terminal amino acid sequence analysis. Infect. Immun. **61**:2030–2036.
18. **Morrow, B., T. Srikantha, J. Anderson, and D. R. Soll.** 1993. Coordinate regulation of two opaque-phase-specific genes during white-opaque switching in *Candida albicans*. Infect. Immun. **61**: 1823–1828.
19. **Morrow, B., T. Srikantha, and D. R. Soll.** 1992. Transcription of the gene for a pepsinogen, *PEP1*, is regulated by white-opaque switching in *Candida albicans*. Mol. Cell. Biol. **12**:2997–3005.
20. **Pearson, W. R., and D. J. Lipman.** 1988. Improved tools for biological sequence comparison. Proc. Natl. Acad. Sci. USA **85**:2444–2448.
21. **Ray, T. L., C. D. Payne, and B. J. Morrow.** 1991. *Candida albicans* acid proteinase: characterization and role in candidiasis. Adv. Exp. Med. Biol. **306**:173–183.
22. **Roeder, G. S., and G. R. Fink.** 1983. Transposable elements in yeast, p. 299–328. *In* J. A. Shapiro (ed.), Mobile genetic elements. Academic Press, New York.
23. **Ruchel, R.** 1981. Properties of a purified proteinase from the yeast *Candida albicans*. Biochim. Biophys. Acta **659**:99–113.
24. **Ruchel, R., F. DeBernardis, T. L. Ray, P. A. Sullivan, and G. T. Cole.** 1992. *Candida* acid proteinases. Med. Vet. Mycol. **30**(Suppl. 1):123–132.
25. **Ruchel, R., R. Tegeler, and M. Trost.** 1982. A comparison of secretory proteinases from different strains of *Candida albicans*. Sabouraudia **20**:233–244.
26. **Ruchel, R., K. Uhlemann, and B. Boning.** 1983. Secretion of acid proteinases by different species of the genus Candida. Zentralbl. Bakteriol. Hyg. I Abt. Orig. A **255**:537–548.
27. **Saiki, R. K., D. H. Gelfand, S. Stoffel, S. J. Sharf, R. Higuchi, G. T. Horn, K. B. Mullis, and H. A. Erlich.** 1988. Primer-directed enzymatic amplification of DNA with thermostable DNA polymerase. Science **239**:487–494.
28. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
29. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
30. **Slutsky, B., M. Staebell, J. Anderson, L. Risen, M. Pfaller, and D. Soll.** 1987. "White-opaque transition," a second high-frequency switching system in *Candida albicans*. J. Bacteriol. **169**:189–197.
31. **Soll, D. R.** 1992. High-frequency switching in *Candida albicans*. Clin. Microbiol. Rev. **5**:183–203.
32. **Soll, D. R., B. Morrow, and T. Srikantha.** 1993. High frequency phenotypic switching in *Candida albicans*. Trends Genet. **9**:61–65.
33. **Staib, F.** 1965. Serum-proteins as nitrogen source for yeast-like fungi. Sabouraudia **4**:187–193.
34. **White, T. C., S. H. Miyasaki, and N. Agabian.** Unpublished observations.
35. **White, T. C., S. H. Miyasaki, and N. Agabian.** 1993. Three distinct secreted aspartyl proteinases in *Candida albicans*. J. Bacteriol. **175**:6126–6133.
36. **Wright, R. J., A. Carne, A. D. Hieber, I. L. Lamont, G. W. Emerson, and P. A. Sullivan.** 1992. A second gene for a secreted aspartate proteinase in *Candida albicans*. J. Bacteriol. **174**:7848–7853.