Nucleic Acids Research, Vol. 20, No. 9 2271–2278

# The *swi4*$^+$ gene of *Schizosaccharomyces pombe* encodes a homologue of mismatch repair enzymes

Oliver Fleck, Holger Michael and Lutz Heim[1]
Institut für Genetik, Biozentrum, Technische Universität Braunschweig, Konstantin-Uhde-Strasse 5, W-3300 Braunschweig and [1]Max-Planck-Institut für Biophysikalische Chemie, Am Fassberg, W-3400 Göttingen, Germany

Received February 4, 1992; Revised and Accepted April 3, 1992 EMBL accession no. X61306

## ABSTRACT

**The *swi4*$^+$ gene of *Schizosaccharomyces pombe* is involved in termination of copy-synthesis during mating-type switching. The gene was cloned by functional complementation of a *swi4* mutant transformed with a genomic library. Determination of the nucleotide sequence revealed an open reading frame of 2979 nucleotides which is interrupted by a 68 bp long intron. The putative Swi4 protein shows homology to Duc-1 (human), Rep-3 (mouse), HexA (*Streptococcus pneumoniae*) and MutS (*Salmonella typhimurium*). The prokaryotic proteins are known as essential components involved in mismatch repair. A strain with a disrupted *swi4*$^+$ gene was constructed and analysed with respect to the switching process. As in *swi4* mutants duplications occur in the mating-type region of the *swi4* (null) strain, reducing the efficiency of switching.**

## INTRODUCTION

Genetic recombination is an essential cellular process providing genetic variation as well as being involved in certain DNA repair mechanisms. Special forms of recombination even contribute to cellular differentiation. An important example of a directed recombination event occurs during mating-type switching in the budding yeast *Saccharomyces cerevisiae* and in the fission yeast *Schizosaccharomyces pombe*. In homothallic ($h^{90}$) strains of *S. pombe*, the frequent MT switching is accomplished by copy transposition of MT information from one of the silent storage cassette genes, *mat2-P* (Plus information) or *mat3-M* (Minus information), to the expression site, *mat1* (1, 2). The cassettes are separated by the L and K regions, each about 16 kb long. Common to all cassettes are flanking regions of homology, H2 (135 bp, proximal) and H1 (59 bp, distal). *mat2* and *mat3* have a third homology box, H3 (57 bp), proximal to H2, which is absent in *mat1* (3). MT switching is initiated by a double-strand break (DSB) positioned at the border of *mat1* to the H1 box. Generation of the DSB is mediated by the nearby located *smt* (switching of mating type) signal (2, 4, 5). The DSB is persistent throughout the cell cycle and present in about 20% of the cells (2). By means of duplications and deletions in the $h^{90}$ MT region, various heterothallic MT configurations arise (6). In all these configurations, including $h^{90}$, the expression site is always located at the most proximal end of the respective MT region. For detailed reviews of the MT system of *S. pombe*, see Egel (7) and Klar (8).

The switching of mating type is regulated by at least eleven switching (*swi*) genes (9–11). All known *swi* mutations reduce but do not abolish MT switching in homothallic strains. The *swi* genes were allocated to three steps of the switching process (9). *swi1*$^+$, *swi3*$^+$ and *swi7*$^+$ are required for the formation of double-strand breaks (class Ia). *swi2*$^+$, *swi5*$^+$ and *swi6*$^+$ are involved in the utilization of the DSB for DNA transposition (class Ib), whereas *swi4*$^+$, *swi8*$^+$, *swi9*$^+$, *swi10*$^+$ and *rad22*$^+$ resolve the recombination structure by termination of the copy-synthesis (class II). The termination site is proposed to be at the 5′ flanking H2 box (4, 12). An important difference between classes I and II is that mutants of the latter frequently give rise to heterothallic segregants, due to duplications in the MT region. This has recently been shown in detail for one of the class II genes, *swi4* (12). The rearrangements are obviously caused during MT switching by the failure to terminate the copy-synthesis in the H2 box of the silent donor cassette. In this study we report the cloning, sequencing and gene disruption of the *swi4*$^+$ gene. The deduced amino acid sequence was compared with the EMBL and PIR databases. Homologies were found with two bacterial (HexA, MutS) and two mammalian (Duc-1, Rep-3) proteins. The prokaryotic proteins are known as mismatch repair enzymes (reviewed in 13, 14).

## MATERIALS AND METHODS

### Strains

The *S. pombe* strains used were originally derived from Leupold's stock culture; their genotypes are given in the text. *E. coli* strains: BJ5183 F$^-$ *recBC sbcB endoI gal met str thi biot hsd* (15); JM101 *supE thi* Δ(*lac-proAB*) F′ [*traD*36 *proAB*$^+$ *lacI*$^q$ *lacZ*ΔM15] (16).

### Media and techniques for *S. pombe*

Media and standard genetical procedures were the same as described elsewhere (17). Chromosomal DNA was isolated according to Wright *et al.* (18), plasmid DNA was isolated from yeast according to Beach *et al.* (19). Transformation of *S. pombe*

was carried out by the lithium acetate method (20) or the lithium chloride method (21). The mitotic stability of plasmids in transformants was measured according to Heyer *et al.* (22).

### Recombinant DNA techniques

Restriction enzyme digests, Southern and Northern hybridization, isolation of DNA fragments from low melting agarose gels and other recombinant techniques were carried out as described by Sambrook *et al.* (23).

### Gene library construction

To construct an *S. pombe* gene library, genomic DNA was isolated from the strain L972 $h^{-S}$ and partially digested with *Sau*3AI. Fragments of approximately 4 to 10 kb in size were eluted from a low melting point agarose gel and cloned into the single *Bam*HI site in the tetracycline-resistance gene of the yeast/*E. coli* shuttle vector pFL20 (15). This plasmid contains the *Saccharomyces cerevisiae URA3* gene which allows for selection in *S. pombe ura4*$^-$ strains. Furthermore, pFL20 contains the *stb* element, a DNA sequence that highly improves plasmid stability in *S. pombe ars*-containing vectors (22). A library of recombinant plasmids was generated by transforming *E. coli* (BJ5183) with the ligation mixture. From the resulting 60,000 clones, plasmid DNA was prepared to form the genomic library GBS3. The percentage of plasmids in this library, GBS3, carrying *S. pombe* DNA was subsequently determined to be 30.5%, with an average insert size of 7.6 kb.

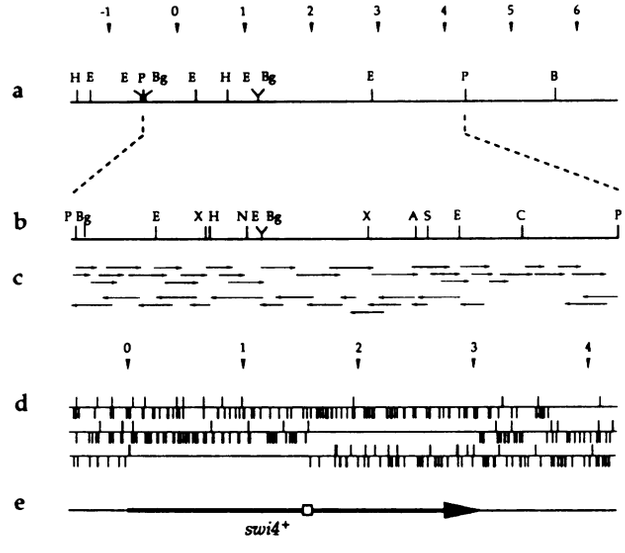


**Figure 1. a.** Restriction map of the pLH1 insert, which was isolated from the genomic library GBS3. Numbers above the map indicate length in kb. Position 0 is the site of the putative ATG codon of the *swi4*$^+$ gene. **b.** Subclone pLH2, the smallest subclone able to complement the chromosomal *swi4* mutation. Restriction enzyme sites in a and b: A, *Acc*I; B, *Bam*HI; Bg, *Bgl*II; C, *Cla*I; E, *Eco*RI; H, *Hin*dIII; N, *Nsi*I; P, *Pst*I; S, *Spe*I; X, *Xba*I. The restriction map is complete for those enzymes shown in panel a and b, respectively. **c.** Sequenced segments of the pLH2 insert. Each arrow represents a region of which the nucleotide sequence was determined in a single sequencing experiment. **d.** Open reading frames, presented in the same orientation as the *Pst*I fragment shown in panel b. No open reading frame was found in the reversed orientation. Potential ATG start sites are shown as upward bars, termination codons as downward bars. **e.** Predicted open reading frame of the *swi4*$^+$ gene (black arrow). The putative intron is indicated by an open box. See text for further details.

### Sequencing

DNA fragments suitable for sequencing were cloned either in M13mp18/mp19 or pUC19. Recombinant clones were incubated at various times with Exonuclease III to obtain nested deletion sets appropriate for sequencing. After ligation with T4 ligase, suitable samples were used for transfection/transformation of the *E. coli* strain JM101 and DNA was prepared from white plaques/colonies by standard procedures. Sequencing was carried out by the dideoxy method (24).

## RESULTS

### Cloning of the *swi4*$^+$ gene

Homothallic *swi4* mutants give rise to heterothallic $h^+$ segregants due to duplications in the MT region. The heterothallic colonies can be easily distinguished from $h^{90}$ *swi* mutants by treatment with iodine vapours. They give a characteristic yellow colour (iodine negative) on sporulation medium (MEA), whereas the $h^{90}$ *swi* mutants have a mottled phenotype. Homothallic wild-type strains turn homogeneously black (iodine positive) (25). These phenotypes reflect the different proportions of starch containing ascospores in the respective colonies.

The *swi4*$^+$ gene of *S. pombe* was cloned by functional complementation using the described iodine test. A genomic library, GBS3, was constructed as described in MATERIALS AND METHODS and an $h^{90}$ *swi4 ura4-D18* (LH83) mutant was transformed with this library. First *ura*$^+$ transformants were selected by plating on minimal medium. In a second step, these were replicated on MEA. After growth the colonies were treated with iodine vapours to search for those which turned homogeneously black. One single colony with this phenotype was found among 30,000 *ura*$^+$ transformants. This clone had obviously regained the ability to switch mating type efficiently. The plasmid, designated pLH1, was recovered from the transformant and subsequently amplified in *E. coli*.

The strain LH83 was retransformed with pLH1. All homothallic transformants grown on MEA were homogeneously iodine positive after treatment with iodine vapours. This result proves that we isolated a plasmid which can complement the *swi4* mutation; thus the possibility is excluded that in the original homogeneously black colony a reversion had occurred. Two of the transformants exhibited an especially intense iodine reaction. One explanation is that these transformants have the cloned gene

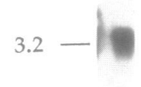


**Figure 2.** Northern blot of total RNA of wild-type strain L968 hybridized with a mixture of the 1.65 *Bgl*II and 1.75 kb *Eco*RI fragments containing the *swi4*$^+$ gene. Before hybridization, fragments were labeled with [$\alpha$-$^{32}$P] dCTP using the multiprime kit (Amersham).

stably integrated into the genome. To test the stability of the *ura*$^+$ prototrophy the two transformants were cultivated for about 20 generations in non-selective complex medium (YEL) and thereafter plated on YEA. 100 colonies of each were tested on selective minimal medium. All remained *ura*$^+$, indicating a chromosomal integration of the plasmid. Since in yeast integration of plasmid DNA in most cases occurs by homologous recombination, these transformants seem to have pLH1 integrated in the chromosomal *swi4* gene. We tested this assumption by mapping the plasmid borne *S. cerevisiae URA3* gene by means

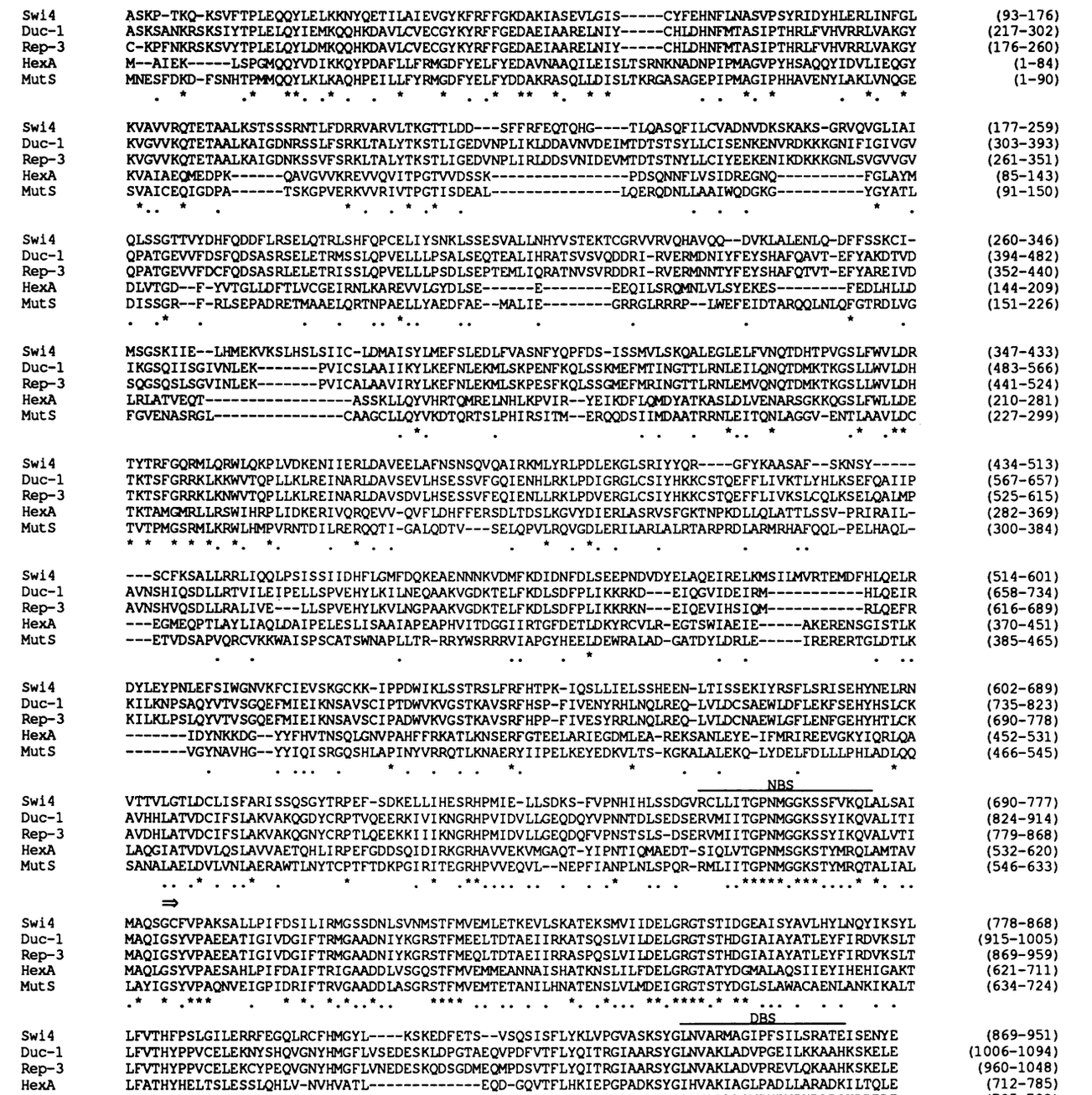


**Figure 3.** Alignment of the predicted amino acid sequence of the *S. pombe swi4*$^+$ gene with the amino acid sequences of Duc-1 (human), Rep-3 (originally reported as Rep-1) (mouse), MutS (*S. typhimurium*) and HexA (*S. pneumoniae*) (27–30). The sequences were obtained from the PIR data bank. The alignment was produced by the CLUSTAL program (31) followed by manual adjustment. Gaps are introduced to obtain an optimal alignment. The MutS sequence contains some corrections as published by Linton *et al.* (28), the Rep-3 sequence includes some corrections furnished by G.F. Crouse (32). Only regions which are clearly homologous are shown. Swi4: 93–951 (859 out of 993), Duc-1: 217–1094 (878 out of 1137), Rep-3: 176–1048 (873 out of 1091), HexA: 1–785 (out of 844), MutS: 1–798 (out of 854). Identical amino acids in the same position in all proteins are marked by asterisks, conservative substitutions by dots. Using the amino acid similarity matrix of Dayhoff (33), conservative substitutions are defined if all pairs of residues have a similarity score greater than or equal to 9. Arrows mark the highly conserved region near the COOH termini. The percent values of identity for this region are listed in Table 1b. Putative binding motifs (see Fig. 4) are designated with NBS (nucleotide-binding site) and with DBS (DNA binding-site).

of tetrad analyses. The *swi4* gene is located on chromosome I, closely linked to *his6* (1.5 cM). Therefore we crossed one transformant (*h*$^{90}$ *ura4-D18 swi4 int::URA3*) with an *h*$^{90}$ *his6 swi4-1 ura4-D18* strain. All 15 dissected tetrads showed the markers in parental combination: 2 *his6*$^{+}$ *swi4*$^{+}$ *ura*$^{+}$: 2 *his6*$^{-}$ *swi4*$^{-}$ *ura*$^{-}$. Thus the integrated *URA3* gene is closely linked to *his6* and therefore it is most likely that pLH1 contains the wild-type *swi4*$^{+}$ gene and not a suppressor.

Restriction analyses of the 8.2 kb insert of pLH1 were made and subclones were constructed to delimit the region harbouring the *swi4*$^{+}$ gene. pLH2 with a 4.75 kb *Pst*I insert was found as the smallest subclone with an intact *swi4*$^{+}$ gene (Fig. 1a, b).

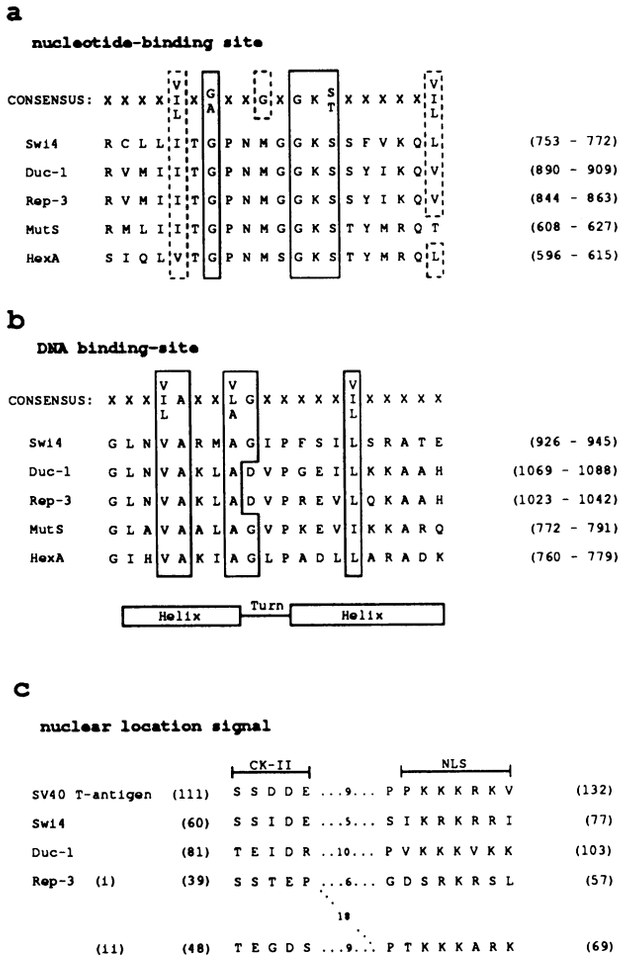


**Figure 4. a.** Homologies in Swi4, Duc-1, Rep-3, MutS and HexA to the consensus sequence of a type A nucleotide-binding site. The nucleotide-binding site was first described by Walker *et al.* (35), enlarged data from Haber *et al.* (29). Continous boxes sign strong conservation of amino acids (≥98%) as compiled by Gill *et al.* (34), interrupted boxes mark conserved amino acids (≥60%), X's in the consensus indicate variable residues. **b.** Homology of Swi4, Duc-1, Rep-3, MutS and HexA to the consensus sequence of a DNA binding-site (36). Boxed residues are conserved (≥75%) as compiled by Pabo and Sauer (36). **c.** Sequence homology of Swi4, Duc-1 and Rep-3 to the nuclear location signal (NLS) and to the casein kinase II (CK-II) site of the SV40 T antigen (38, 39, 45). Numbers between the potential CK-II and NLS indicate the numbers of intervening amino acids. Numbers in brackets give the position of either the first residue of the potential CK-II site (left) or the last residue of the NLS (right). For Rep-3, two potential CK-II sites and two NLSs are present.

## Nucleotide sequence of the *swi4*$^{+}$ gene

The pLH2 insert was subcloned in M13 and pUC vectors for sequencing (see MATERIALS AND METHODS). The nucleotide sequence was determined by the dideoxy method (24) according to the strategy presented in Fig. 1c. The sequence was determined from both strands with the exception of about 650 bp outside the putative *swi4*$^{+}$ ORF (see below) between the right *Eco*RI and *Pst*I sites. The putative open reading frame of the *swi4*$^{+}$ gene is shown in Fig. 1d starting at position 0. A second ORF might be present downstream of the *swi4*$^{+}$ gene (Fig. 1d; around position 4 kb). The reading frame of *swi4*$^{+}$ is obviously disrupted by an intron including the nucleotide positions 1498 to 1565. This intron seems to be 68 bp long and has the 5′ splice site GTAAG, the branch site CTGAA and the 3′ splice site AAG. These sequences are in good agreement with the *S. pombe* consensus sequences GTANG (5′ splice site), CTPuAN (branch site), NAG (3′ splice site) (26).

To distinguish between the possibilities of either one large gene of approximately 3 kb or two smaller genes (cf. Fig. 1d), a Northern hybridization was performed. Total RNA of the *h*$^{90}$ wild-type strain was probed with a mixture of the 1.65 kb *Bgl*II and the 1.75 kb *Eco*RI fragments. These two fragments comprise both sides of the putative intron, but not the additional reading frame, located downstream of *swi4*$^{+}$ (see Fig. 1d). As shown in Fig. 2, one single band of approximately 3.2 kb was observed, indicating one large transcript and thus the presence of the predicted intron. Therefore, the *swi4*$^{+}$ gene terminates at nt 3048 by a TGA stop codon and comprises a coding region of 2979 bp.

## Predicted amino acid sequence of *swi4*$^{+}$

Translation of the *swi4*$^{+}$ coding region gives rise to a polypeptide consisting of 993 residues with a calculated $M_r$ of 113 kD and a pI value of 7.9. The predicted Swi4 protein was compared with proteins in the Protein Identification Resource (PIR) and the EMBL Data Bank. We found significant homologies to the human Duc-1 (27), the mouse Rep-3, originally

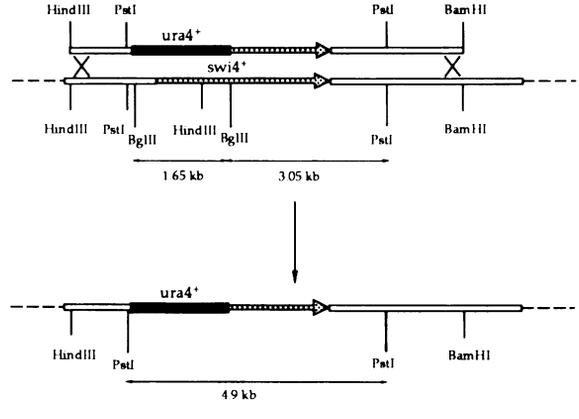


**Figure 5.** Gene disruption of the *swi4*$^{+}$ gene with *ura4*$^{+}$ and insertion of the resulting 7.2 kb *Hin*dIII/*Bam*HI fragment in the chromosome via homologous recombination. The gene disruption was proved by genetical crosses (see text) and Southern hybridization using the 4.75 kb *Pst*I fragment as probe. In *Pst*I/*Bgl*II digests the wild-type strain and the *swi4-1* mutant yielded 1.65 kb and 3.05 kb bands whereas the *swi4* (null) mutants gave only one single 4.9 kb band (data not shown).

designated as Rep-1 (28) and also to MutS of *Salmonella typhimurium* (29) and HexA of *Streptococcus pneumoniae* (30). The two bacterial proteins, HexA and MutS, have previously been identified as essential components of their respective DNA mismatch-repair systems (reviewed in 13, 14). In Fig. 3, the alignment of the five amino acid sequences is presented. The $NH_2$ terminal region of the three eukaryotic proteins, Duc-1 (aa 1–216), Rep-3 (aa 1–175) and Swi4 (aa 1–92) and approximately 50 amino acids at the COOH ends of all five proteins are not shown, because no clear homologies exist in these tracts among all. The aligned sequences share significant homology over their entire lengths, with a few less conserved stretches. The numbers scheduled in Table 1a correspond to percent values of identical residues within these regions. A region near the COOH termini is the best conserved part of the proteins (Table 1b). Inside this region, a type A nucleotide-binding site and a DNA-binding motif has been found in the two bacterial proteins MutS and HexA (29, 34–36). For the nucleotide-binding site of the MutS protein ATPase activity was recently proved (37). These motifs are also present in Duc-1, Rep-3 and the deduced sequence of Swi4 (Fig. 4a, b). In position 1 of the helix-turn-helix motif of the putative DNA binding-site, a glycine residue is present in all five proteins. Since glycine is a strong helix breaker, the left helix of Fig. 4b should begin at position 2 and is therefore one amino acid shorter than the motif which is known for the DNA-binding site of other proteins (36).

The five amino acid sequences contain domains of significant homologies distributed over their entire lengths, only diverging at the $NH_2$ and COOH ends. However, the eukaryotic proteins Duc-1, Swi4 and Rep-3 have a short domain of mainly basic residues in their otherwise quite different $NH_2$ termini. These domains show similarity to a motif of the SV40 T antigen (Fig. 4c), which is known as a nuclear location signal (NLS) (38, 39). Based on the homologous alignment the $NH_2$ termini of the two bacterial proteins begin about 150 residues past this domain and lack the corresponding basic domain. Potential NLS with similiar sequences have been found for instance in various steroid receptors of higher eucaryotes (reviewed in 40), the polyoma virus T antigen (41), and the *dorsal* protein of *Drosophila melanogaster* (42, 43). Moreover, it was recently reported that a fusion protein containing the NLS of the SV40 T antigen is translocated into the nucleus of tobacco cells (44), indicating a widespread and well conserved transport mechanism in eukaryotes. In many proteins harbouring nuclear location signals, putative casein kinase II (CK-II) sites are located nearby the NLSs. It was recently shown for the CK-II of the SV40 T antigen that phosphorylation at the $S^{111}/S^{112}$ amino acids regulate the protein transport into the nucleus, while the specifity of the transport is determined by the NLS (45). As in the SV40 T antigen, sequences conforming to the consensus CK-II recognition site S/T X X D/E (46) are also present near the potential NLSs of Swi4, Duc-1 and Rep-3 (Fig. 4c).

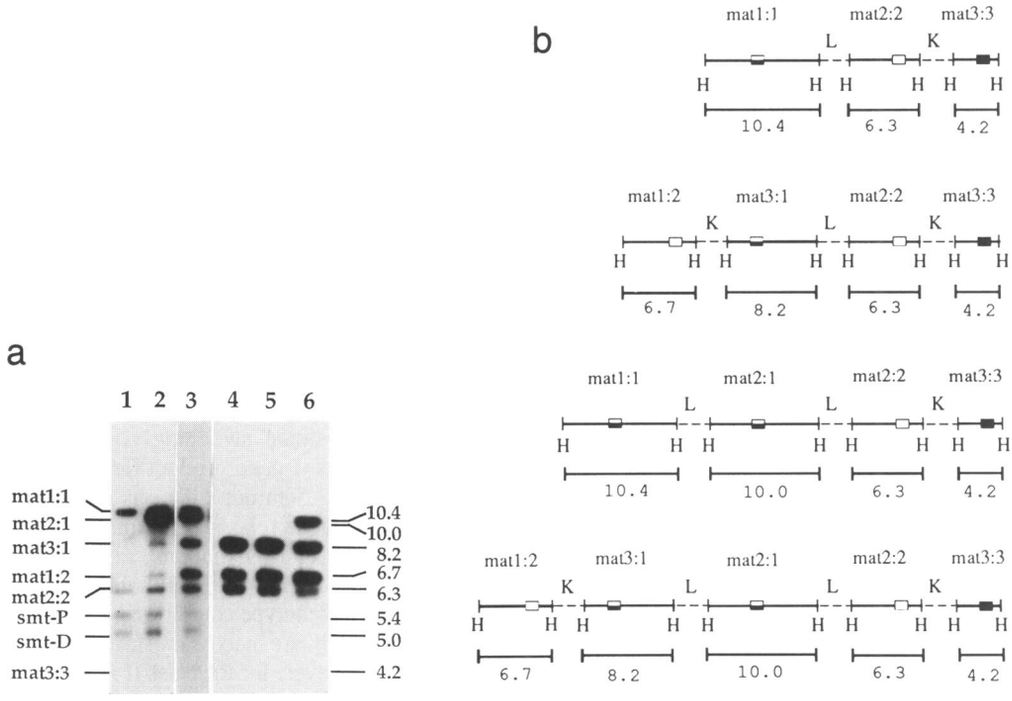


**Figure 6. a.** Southern blot analyses of *Hin*dIII digested DNA, hybridized with the 10.4 kb *mat1–P* fragment. Lane 1: $h^{90}$ wild type; lane 2: $h^{90}$ *swi4*$^-$; lane 3: $h^{90}$ *swi4* (null); lane 4: $h^{+N}$ wild type, lanes 5 and 6: $h^+$ segregants of the $h^{90}$ *swi4* (null) strain. Numbers on the right side indicate the sizes of *Hin*dIII fragments (in kb). Due to the low homology of the 4.2 kb *mat3* (=*mat3:3*) fragment with the hybridization probe, a definite band is detectable only after overexposure. Symbols on the left side indicate the respective cassette types, specified according to the nomenclature of Beach and Klar (6): The first number of a given cassette indicates the origin of the proximal flanking sequence, the second number the origin of the distal flanking sequence, both relating to the standard $h^{90}$ MT region (shown in β, first panel). Therefore, *mat1*, *mat2* and *mat3* can be designated also as *mat1:1*, *mat2:2* and *mat3:3*, respectively. The cassette *mat3:1*, for example, has the proximal flanking sequence K as *mat1:1*, but the distal flanking sequence L as *mat2:2*. **b.** Interpretation of the autoradiographs, supported by previous data about the *swi4-1* mutant (12). Several MT configurations, obviously present in $h^{90}$ *swi4* and $h^{90}$ *swi4* (null) cultures, are shown. *Hin*dIII fragments thereof hybridizing with the 10.4 kb *mat1–P* fragment are given below each configuration. The sizes of *Hin*dIII fragments allow to identify the flanking sequences of cassettes and thereby the relative position of each cassette inside the MT configurations. Cassettes are shown as white (with *P* information) or black boxes (with *M* information). Cassettes with *smt* signals can either contain *P* or *M*. First panel: $h^{90}$ MT region; second panel: $h^{+N}$ MT region; third panel: $h^{90'}$ MT region, which contains an additional *mat2:1* cassette as compared to the standard $h^{90}$ configuration. Fourth panel: $h^{+N'}$ configuration, which differs from the standard $h^{+N}$ by an additional *mat2:1* cassette.

### Gene disruption of *swi4*$^+$

In all known $h^{90}$ *swi* mutants, MT switching is reduced but not completely abolished. This is evidenced by the fact that MT switching of $h^{90}$ *swi* mutants leads to colonies which are mottled after treatment with iodine vapour. The reduced switching could either be carried out by the mutated gene product (leaky) or by other proteins (non leaky). To differentiate between these possibilities we disrupted the wild-type gene and analyzed the resulting strains.

In order to disrupt *swi4*$^+$, the plasmid pLH1 was digested with *Bgl*II to delete the 1.65 kb *Bgl*II fragment. This fragment contains about 1100 nt of the coding region and putative promoter sequences. At this site the 1.8 kb *Hin*dIII fragment containing the *ura4*$^+$ marker was inserted via blunt ends. A haploid $h^{90}$ *ura4-D18* strain was transformed with the 7.2 kb *Hin*dIII/*Bam*HI fragment containing the *swi4* gene disrupted with *ura4*$^+$ (Fig. 5). Since in *ura4-D18* mutants the *ura4*$^+$ gene is completely deleted (47), gene replacement by the linear *Hin*dIII/*Bam*HI fragment should occur exclusively at the *swi4*$^+$ locus (Fig. 5). Seven *ura*$^+$ transformants were isolated. All transformants ($h^{90}$ *swi4::ura4*$^+$ *ura4-D18*) were crossed with an $h^{90}$ *his6 swi4-1 ura4* strain to test genetical linkage between the insert borne *ura4*$^+$ marker and *his6*. Of the crosses 58 tetrads were dissected. All asci are of the parental ditype: 2 *his6*$^+$ *ura4*$^+$ : 2 *his6*$^-$ *ura4*$^-$. This result demonstrates a strong linkage of the markers and thus a correct gene replacement. The gene disruption was additionally confirmed by Southern analyses (data not shown).

### The strains with a disrupted *swi4*$^+$ gene generate duplications in their MT region

For *swi4* and the other class II *swi* mutants it was postulated that the copy-synthesis is frequently not terminated at the H2 box of the silent donor cassette (either *mat2* or *mat3*) but continues to the H2 box of the next proximal located cassette (9, 12). After termination two cassettes and their intervening sequence (either K or L) are present instead of *mat1*.

To test the influence of the disruption on MT switching the strains with the disrupted *swi4* gene were plated on MEA. After growth the colonies were treated with iodine vapours. Mottled and iodine negative colonies were found. Almost all of the latter were $h^+$. The $h^{90}$ *swi4* (null) strains thus have the same phenotype as $h^{90}$ *swi4*$^-$ mutants with respect to the iodine reaction (9, 12). The occurrence of duplications in the MT region was tested by Southern hybridization. Genomic DNA samples of the $h^{90}$ *swi4*$^-$ mutant, of one $h^{90}$ *swi4* (null) strain and in addition of six $h^+$ segregants thereof were digested with *Hin*dIII and probed with the 10.4 kb *Hin*dIII fragment containing the *mat1-P* cassette (Fig. 6a). As controls, $h^{90}$ and $h^{+N}$ wild-type strains are shown in lanes 1 and 4 of Fig. 6a. The $h^{90}$ wild-type strain (lane 1) yielded the typical bands of 10.4 kb (*mat1*), 6.3 kb (*mat2*), 4.2 kb (*mat3*) and in addition of 5.0 kb and 5.4 kb length. The latter two are subfragments of the 10.4 kb *mat1* fragment (termed *smt-D* and *smt-P*, respectively) generated by double-strand breaks at the *smt* signal (2). The $h^{90}$ configuration is shown in Fig. 6b, panel 1. In contrast, the $h^{90}$ *swi4* (null) strain (lane 3) and the $h^{90}$ *swi4* mutant (lane 2) have additional bands of 8.2 kb (*mat3:1*), 10.0 kb (*mat2:1*) and 6.7 kb (*mat1:2*). The designation of the latter cassettes is explained in Fig. 6. The *mat3:1* and *mat1:2* bands belong to $h^+$ segregants which are present in the cultures. The $h^+$ segregants themselves have either an $h^{+N}$ MT region (one example is given in lane 5) or additional *mat2:1* bands (lane 6). The MT configuration of $h^{+N}$ is shown in Fig. 6b, panel 2, the one with an additional *mat2:1* cassette ($h^{+N'}$) in panel 4. Since the expression locus *mat1:2* does not contain an *smt* signal, MT switching is abolished in the heterothallic segregants. *smt* signals are present at *mat3:1* and *mat2:1*, but they are rarely used for initiation of the copy-transposition (12). Thus in contrast to homothallic configurations (with *mat1*) further duplications are generated only rarely in $h^+$.

In summary, in the $h^{90}$ strains with a disrupted *swi4* gene, duplications occur in the same manner as in the $h^{90}$ *swi4-1* mutant (12). Both strains consist of mixed cultures in which cells with different MT configurations are present. Four examples are given in Fig. 6b. The data show that the duplications are due to the loss of the Swi4 function instead of being caused by an altered Swi4 protein with a malfunction. Since the haploid strains with a disrupted *swi4*$^+$ gene are viable, *swi4*$^+$ is not an essential gene.

**Table 1.** Comparisons of the homologous amino acid sequences of Swi4, Duc-1, Rep-3, HexA and MutS.

**a.** Percentage values of identical residues of the aligned regions as shown in Fig. 3. Gaps, introduced for optimal alignment, are counted as one position.

| | Rep-3 | Swi4 | HexA | MutS |
|---|---|---|---|---|
| Duc-1 | 86.6 | 39.0 | 28.0 | 28.4 |
| Rep-3 | – | 37.5 | 28.5 | 29.1 |
| Swi4 | | – | 25.0 | 26.2 |
| HexA | | | – | 37.1 |

**b.** Percentage values of identical residues between the regions near the COOH termini (marked in Fig. 3 by arrows).

| | Rep-3 | Swi4 | HexA | MutS |
|---|---|---|---|---|
| Duc-1 | 89.9 | 47.9 | 45.1 | 46.9 |
| Rep-3 | – | 46.8 | 43.1 | 46.5 |
| Swi4 | | – | 41.5 | 40.7 |
| HexA | | | – | 48.6 |

## DISCUSSION

In the $h^{90}$ *swi* mutants MT switching still occurs, albeit with a lower frequency as evidenced by the mottled colonies. The reduction in residual switching could either be caused by a partially active *swi* gene product (leakiness) or by the products of other genes. Schmidt (48) has shown that strains with a disrupted *swi1* gene (class Ia) are mottled and exhibit a reduced level of DSBs like *swi1* mutants, i.e. they still show some MT switching. The *swi1* mutations therefore seem not to be leaky. Their mottled phenotype could best be explained by the low level of DSBs, which are necessary for the initiation step of MT switching. However, for the class II mutants the situation is more complicated. The termination errors could also be caused by a malfunction of the mutated gene products.

In this study, we have shown that strains with a disrupted *swi4*$^+$ gene have the same mottled phenotype as *swi4* mutants and likewise exhibit duplications in the MT region (Fig. 6). A malfunction of an aberrant *swi4* gene product responsible for the duplications can therefore be excluded and a partial function of the *swi4* gene product is neither essential for cell viability nor responsible for the incorrect termination step.

The *swi4*$^+$ gene product shows a high degree of homology to Duc-1 (human), Rep-3 (mouse), MutS (*Salmonella typhimurium*),

and HexA (*Streptococcus pneumoniae*) (Fig. 3). The two bacterial proteins are known to be mismatch repair enzymes. The Mut system of *Salmonella typhimurium* is very similar to the Mut system of *E. coli*; i.e. cloned *Salmonella mut* genes can complement respective mutations in *E. coli* (49, 50). In both the Mut and the Hex repair systems, a DNA segment of the target strand greater than 1 kb is removed (51, 52). It was demonstrated for the MutS protein of *Salmonella typhimurium* that it binds to single-stranded DNA and to heteroduplex DNA with high affinity (50, 53).

The aligned part of the Swi4 amino acid sequence comprises approximately 86% of the total protein and is over its entire length clearly homologous with all other shown sequences. The two mammalian proteins Duc-1 and Rep-3 share in this region a homology of 86.5% (amino acid identity, see Table 1a), which is the highest value found for all comparisons. The homology of these two mammalian sequences with Swi4, the third eukaryotic sequence, is about 38% (Duc-1 39%, Rep-3 37.4%). This value is clearly higher than the one found between all three eukaryotic sequences and the eubacterial HexA (25−28%) and MutS sequences (26−29%). The *S. pombe* sequence shows a slightly lower percent value (difference about 3%) than the two mammalian ones in the comparisons with the two eubacterial sequences. This difference might be due to a faster evolutionary rate of the protein from the unicellular eukaryote as compared to the proteins of the higher eukaryotes. The two eubacterial proteins, MutS and HexA, are much more similiar to each other (identity value of 37.1%) than to any of the eukaryotic proteins (25−29%).

In all five compared proteins a region of approximately 250 residues near the COOH terminus is highly conserved (Table 1b). This part of the protein has therefore been chosen for a separate comparison in order to get a clearer idea of the relationship between the different sequences.

Except for the very high 89.5% homology for the two mammalian proteins (Duc-1 and Rep-3), all pairwise comparisons are surprisingly within a range of 41−49% identical amino acid positions. Even the degree of homology between the two bacterial and the three eukaryotic proteins (41−47%) is almost the same as the one within each group (except for the higher value between Duc-1 and Rep-3). This suggests that this part of the protein contains sequences, which are functionally indispensible, and are therefore extremly conserved. One hint for the function of this region is the presence of nucleotide- and DNA binding motifs in each of the sequences (Fig. 4). A third highly conserved sequence motif 'DELGRGTSTXDG' (positions 838−849 in Swi4) has been found. This motif might represent an additional to date not characterized functional domain.

Although direct evidence concerning the function of the Swi4 protein has not yet been directly determined, it is known that *swi4*$^+$ plays a role in termination of copy-synthesis during MT switching (9, 12). Since Swi4 is required for efficient MT switching and probably for mismatch correction, this means that the protein must normally be translocated from the cytoplasm into the nucleus. However, since the Swi4 protein is 113 kD, exceeding the critical size of 60 kD, it is too large for a simple diffusion into the nucleus (54). Therefore, the Swi4 protein should be actively transported into the nucleus. This could be mediated by a short stretch of basic amino acids that corresponds to known or predicted nuclear location signals (38−43) (Fig. 4c). Such a motif is also present in the two eukaryotic homologues of Swi4, Duc-1 and Rep-3, but, as expected, absent in the prokaryotic proteins MutS and HexA. In addition, Swi4, Duc-1 and Rep-3 contain a sequence near their potential NLS, which is coincident with the consensus of a CK−II recognition site (46) (Fig. 4c).

Based on the homology data presented here, the idea suggests itself that Swi4 recognizes and binds to heteroduplex DNA as it was reported for MutS (37, 53). This is easily conceivable for the correction of mismatched DNA, although there is no direct evidence for the action of Swi4 in mismatch repair until now. However, by assuming the same role for Swi4 in MT switching, it is reasonable to predict a heteroduplex DNA in this mechanism too. In recent models it was postulated that the copy-synthesis terminates in the proximal flanking H2 box (4, 12). Swi4 may bind at this site by recognizing an intramolecular secondary structure which contains some unpaired bases (i.e. mismatches). Analyses of the nucleotide sequence of H2 revealed indeed the possibility for such a structure (data not shown). Subsequently, binding of Swi4 could stimulate strand incision, carried out by another Swi gene product, and thereby termination of the copy-synthesis.

As MT switching in yeast is a site specific recombination process which might require a mismatch repair enzyme, it is very likely that Swi4 is such an enzyme. Both the characterization of *swi4* mutants and disrupted strains, as well as the homology of Swi4 with other DNA mismatch repair enzymes support this idea. Subject of further analyses is to clarify if Swi4 is indeed involved in general mismatch repair or exclusively in the MT switching process.

## ACKNOWLEDGEMENTS

We are indebted to Henner Brinkmann for valuable discussion, especially in protein data analyses, and thank him, Herbert Gutz, Warren Kibbe and Henning Schmidt for critical reading of the manuscript. We also thank Susanne Heiser for excellent technical assistance, Thomas Heinemeyer, William F. Martin and John G. Sgouros for help in protein homology searches, and Ulrike Brandt for kindly providing the *S. pombe* RNA sample.

## REFERENCES

1. Egel,R. and Gutz,H. (1981) *Curr. Genet.*, **3**, 5−12.
2. Beach,D.H. (1983) *Nature*, **305**, 682−687.
3. Kelly,M., Burke,J., Smith,M., Klar,A.J.S. and Beach,D. (1988) *EMBO J.*, **7**, 1537−1547.
4. Nielsen,O. and Egel,R. (1989) *EMBO J.*, **8**, 269−276.
5. Klar,A.J.S., Bonaduce,M.J. and Cafferkey,R. (1991) *Genetics*, **127**, 489−496.
6. Beach,D.H. and Klar,A.J.S. (1984) *EMBO J.*, **3**, 603−610.
7. Egel,R. (1989) In Nasim,A., Young,P. and Johnson,B.F. (eds.), Molecular Biology of the Fission Yeast. Academic Press, San Diego, pp. 31−73.
8. Klar,A.J.S. (1989) In Berg,D.E. and Howe,M.W. (eds.), Mobile DNA. American Society for Microbiology, Washington, D.C., pp. 671−691.
9. Egel,R., Beach,D.H. and Klar,A.J.S. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 3481−3485.
10. Gutz,H. and Schmidt,H. (1985) *Curr. Genet.*, **9**, 325−331.
11. Schmidt,H., Kapitza-Fecke,P., Stephen,E.R. and Gutz,H. (1989) *Curr. Genet.*, **16**, 89−94.
12. Fleck,O., Heim,L. and Gutz,H. (1990) *Curr. Genet.*, **18**, 501−509.
13. Modrich,P. (1987) *Annu. Rev. Biochem.*, **56**, 435−466.
14. Meselson,M. (1988) In Low,K.B. (ed.), The Recombination of Genetic Material. Academic Press, San Diego, pp. 91−113.
15. Losson,R. and Lacroute,F. (1983) *Cell*, **32**, 371−377.
16. Messing,J. (1979) *Recomb. DNA Tech. Bull.*, **2(2)**, 43−48.
17. Gutz,H., Heslot,H., Leupold,U. and Loprieno,N. (1974) In King, R.C. (ed.), Handbook of Genetics. Plenum Press, New York, London, Vol. 1, pp. 395−446.

18. Wright,A.P.H., Maundrell,K. and Shall,S. (1986) *Curr. Genet.*, **10,** 503–508.
19. Beach,D., Piper,M. and Nurse,P. (1982) *Mol. Gen. Genet.*, **187,** 326–329.
20. Ito,H., Fukuda,Y., Murata,K. and Kimura,A. (1983) *J. Bacteriol.*, **153,** 163–168.
21. Bröker,M. (1987) *BioTechniques*, **5,** 516–517.
22. Heyer,W.-D., Sipiczki,M. and Kohli,J. (1986) *Mol. Cell. Biol.*, **6,** 80–89.
23. Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) Molecular Cloning–A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.
24. Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74,** 5463–5467.
25. Leupold,U. (1955) *Schweiz. Z. allg. Pathol. Bakteriol.*, **18,** 1141–1146.
26. Russell,P. (1989) In Nasim,A., Young,P. and Johnson,B.F. (eds.), Molecular Biology of the Fission Yeast. Academic Press, San Diego, pp. 244–271.
27. Fujii,H. and Shimada,T. (1989) *J. Biol. Chem.*, **264,** 10057–10064.
28. Linton,J.P., Yen,J.-Y., Selby,E., Chen,Z., Chinsky,J.M., Liu,K., Kellems,R.E. and Crouse,G.F. (1989) *Mol. Cell. Biol.*, **9,** 3058–3072.
29. Haber,L.T., Pang,P.P., Sobell,D.I., Mankonch,J.A. and Walker,G.C. (1988) *J. Bacteriol.*, **170,** 197–202.
30. Priebe,S.D., Hadi,S.M., Greenberg,B. and Lacks,S.A. (1988) *J. Bacteriol.*, **170,** 190–196.
31. Higgins,D.G. and Sharp,P.M. (1988) *Gene*, **73,** 237–244.
32. Crouse,G.F. (unpublished). GenBank accession no. M80360.
33. Dayhoff,M.O. (1978) In Dayhoff,M.O. (ed.), Atlas of Protein Sequence and Structure. National Biomedical Research Foundation, Washington DC, pp. 345–358.
34. Gill,D.R., Hatfull,G.F. and Salmand,G.P.C. (1986) *Mol. Gen. Genet.*, **205,** 134–145.
35. Walker,J.E., Saraste,M., Runswick,M.J. and Gay,N.J. (1982) *EMBO J.*, **1,** 945–951.
36. Pabo,C.O. and Sauer,R.T. (1984) *Annu. Rev. Biochem.*, **53,** 293–321.
37. Haber,L.T. and Walker,G.C. (1991) *EMBO J.*, **10,** 2707–2715.
38. Kalderon,D., Richardson,W.D., Markham,A.F. and Smith,A.E. (1984a) *Nature*, **311,** 33–38.
39. Kalderon,D., Roberts,W.D., Richardson,W.D. and Smith,A.E. (1984b) *Cell*, **39,** 499–509.
40. Guiochon-Mantel,A., Loosfelt,H., Lescop,P., Sar,S., Atger,M., Perrot-Applanat,M. and Milgrom,E. (1989) *Cell*, **57,** 1147–1154.
41. Richardson,W.D., Roberts,B.L. and Smith,A.E. (1986) *Cell*, **44,** 77–85.
42. Steward,R. (1987) *Science*, **238,** 692–694.
43. Steward,R. (1989) *Cell*, **59,** 1179–1188.
44. Lassner,M.W., Jones,A., Daubert,S. and Comai,L. (1991) *Plant Mol. Biol.*, **17,** 229–234.
45. Rihs,H.-P., Jans,D.A., Fan,H. and Peters,R. (1991) *EMBO J.*, **10,** 633–639.
46. Krebs,E.G., Eisenmann,R.N., Kuenzel, E.A., Litchfield,D.W., Lozeman,F.J., Lüscher,B. and Sommercorn,J. (1988) *Cold Spring Harbor Symp. Quant. Biol.*, **53,** 77–84.
47. Grimm,C., Kohli,J., Murray,J. and Maundrell,K. (1988) *Mol. Gen. Genet.*, **215,** 81–86.
48. Schmidt,H. (1987) *Mol. Gen. Genet.*, **210,** 485–489.
49. Pang,P.P. and Walker,G.C. (1983) *J. Bacteriol.*, **153,** 1172–1179.
50. Pang,P.P., Lundberg,A.S. and Walker,G.C. (1985) *J. Bacteriol.*, **163,** 1007–1015.
51. Wagner,R.,Jr. and Meselson,M. (1976) *Proc. Natl. Acad. Sci. USA*, **73,** 4135–4139.
52. Lacks,S. (1966) *Genetics*, **53,** 207–235.
53. Su,S.-S. and Modrich,P. (1986) *Proc. Natl. Acad. Sci. USA*, **83,** 5057–5061.
54. Dingwall,C. and Laskey,R.A. (1986) *Annu. Rev. Cell Biol.*, **2,** 367–390.

## Note added in proof

The sequence of an additional homologue of Swi4 was determined recently: YMBP of *Saccharomyces cerevisiae* (Valle *et al.* 1991, Yeast **7**: 981–988)