Isolation and Characterization of Two *Saccharomyces cerevisiae* Genes Encoding Homologs of the Bacterial HexA and MutS Mismatch Repair Proteins

Robert A. G. Reenan and Richard D. Kolodner

Division of Cellular and Molecular Biology, The Dana-Farber Cancer Institute, Boston, Massachusetts 02115, and Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02115

Manuscript received March 19, 1992
Accepted for publication August 12, 1992

ABSTRACT

Homologs of the *Escherichia coli* (*mutL*, *S* and *uvrD*) and *Streptococcus pneumoniae* (*hexA*, *B*) genes involved in mismatch repair are known in several distantly related organisms. Degenerate oligonucleotide primers based on conserved regions of *E. coli* MutS protein and its homologs from *Salmonella typhimurium*, *S. pneumoniae* and human were used in the polymerase chain reaction (PCR) to amplify and clone *mutS/hexA* homologs from *Saccharomyces cerevisiae*. Two DNA sequences were amplified whose deduced amino acid sequences both shared a high degree of homology with MutS. These sequences were then used to clone the full-length genes from a yeast genomic library. Sequence analysis of the two *MSH* genes (*MSH* = *mutS* homolog), *MSH1* and *MSH2*, revealed open reading frames of 2877 bp and 2898 bp. The deduced amino acid sequences predict polypeptides of 109.3 kD and 109.1 kD, respectively. The overall amino acid sequence identity with the *E. coli* MutS protein is 28.6% for MSH1 and 25.2% for MSH2. Features previously found to be shared by MutS homologs, such as the nucleotide binding site and the helix-turn-helix DNA binding motif as well as other highly conserved regions whose function remain unknown, were also found in the two yeast homologs. Evidence presented in this and a companion study suggest that *MSH1* is involved in repair of mitochondrial DNA and that *MSH2* is involved in nuclear DNA repair.

REPLICATION fidelity is essential to ensure that genetic information is faithfully passed from one generation to the next. Errors not corrected by the proofreading mechanism of DNA polymerases must be repaired by another mechanism or mutations result. One repair system that has been well characterized at the biochemical level, the *mutHLS* system of *Escherichia coli*, removes misincorporated nucleotides in a manner dependent upon the methylation state of d(GATC) sites. The result of the removal of mismatched nucleotides from the nascent strand is frequently referred to as mutation avoidance (for review see MODRICH 1987).

An *in vitro* methyl-directed mismatch repair system has been characterized in *E. coli* (LU, CLARK and MODRICH 1983). This system was found to require the gene products of the *mutH*, *mutL*, *mutS* and *mutU(uvrD)* genes. In addition to the *mut* gene products, the *in vitro* repair reaction also requires SSB (single-stranded binding protein), DNA ligase, exonuclease I, DNA polymerase III holoenzyme, ATP, and the four deoxynucleoside triphosphates (MODRICH 1987). Specific functions have been assigned to several of the *mut* gene products: MutH protein is a d(GATC) specific endonuclease which makes a nick in the unmethylated strand at a hemimethylated GATC site and is required for strand discrimination (WELSH *et al.* 1987); MutS protein binds to DNA containing a mismatch (SU and MODRICH 1986), and has weak ATPase activity (HABER and WALKER 1991); and MutU protein (UvrD) is DNA helicase II and is thought to act in the excision stage of the reaction (MATSON and GEORGE 1987; MODRICH 1987). No precise function has been elucidated for the *mutL* gene product although the purified product increases the size of the nuclease protection pattern of MutS protein on mismatched DNA substrates (GRILLEY *et al.* 1989). Mutations in any of the *mut* genes result in a mutator phenotype (COX 1976; GLICKMAN, VAN DEN ELSEN and RADMAN 1978; GLICKMAN and RADMAN 1980). Studies have also shown that the *mutHLS* system functions *in vivo* to repair different mismatches with different efficiencies (KRAMER, KRAMER and FRITZ 1984).

Mismatch repair systems studied in other organisms share similarity to the *mut* system of *E. coli*. *Salmonella typhimurium* possesses homologs of all of the *mut* genes and these genes are capable of complementing corresponding mutations of *E. coli* (PANG *et al.* 1984). The gram-positive bacterium *Streptococcus pneumoniae* has the *hex* repair system which repairs heteroduplex DNA containing mismatches (CLAVERYS and LACKS 1986). Two loci have been identified which are necessary for this long-patch mismatch repair, *hexA* and

hexB. *hexA* is homologous to *mutS* and *hexB* is homologous to *mutL* (PRIEBE *et al.* 1988; PRUDHOMME *et al.* 1989). No *pneumococcal mutH* homolog has been discovered. The mechanism of strand discrimination remains unknown in *S. pneumoniae*, although single strand breaks in the nascent DNA strands have been proposed as a potential recognition signal which could explain the lack of a requirement for a *mutH* homolog in *S. pneumoniae* (CLAVERYS and LACKS 1986).

In eucaryotes, mismatch repair systems remain less well defined. Mismatch correction has been demonstrated in cell-free extracts of *Saccharomyces cerevisiae* (MUSTER-NASSAL and KOLODNER 1986). Transformation of yeast with mismatch containing DNAs has directly detected mismatch repair *in vivo* and demonstrated that different mismatches are repaired with different efficiencies (BISHOP and KOLODNER 1986; BISHOP *et al.* 1987; KRAMER *et al.* 1989a; BISHOP, ANDERSON and KOLODNER 1989). The *PMS1* gene of *S. cerevisiae* has been cloned and is homologous to the *E. coli mutL* gene (KRAMER *et al.* 1989b). Mutations in *PMS1* cause a mutator phenotype and cause an increase in the ratio of post-meiotic segregation events to gene conversion events during meiotic recombination, consistent with a defect in mismatch repair (WILLIAMSON, GAME and FOGEL 1985). In addition, *pms1* mutations decrease the frequency of mismatch repair following transformation with heteroduplex DNA, decrease the repair of G-G mismatches during gene conversion at the *arg4*-nsp allele, and decrease mismatch repair-induced meiotic recombination (BISHOP *et al.* 1987; BISHOP, ANDERSON and KOLODNER 1989; KRAMER *et al.* 1989a; LICHTEN *et al.* 1990; BORTS *et al.* 1990). The *RADH/HPR5/SRS2* gene of *S. cerevisiae* is homologous to the *mutU(uvrD)* gene of *E. coli* (ABOUSSEKHRA *et al.* 1989; RONG *et al.* 1991). *radH/hpr5/srs2* mutants have increased sensitivity to killing by UV irradiation and certain *hpr5/radH/srs2* alleles bestow a hyper-gene conversion phenotype, although a direct role in mismatch repair has not been demonstrated (RONG *et al.* 1991).

Mismatch repair in higher eucaryotes has been demonstrated in cell-free extracts of *Xenopus laevis* and human cells (VARLET, RADMAN and BROOKS 1990; HOLMES, CLARK and MODRICH 1990; THOMAS, ROBERTS and KUNKEL 1991). A system for transfection of SV40 DNA molecules containing mismatches into simian cells has been used to demonstrate general mismatch correction *in vivo* (BROWN and JIRICNY 1988). This system has also revealed the presence of a specific mismatch repair pathway capable of restoring G/C pairs lost via deamination of 5-methylcytosine residues (BROWN and JIRICNY 1987). Several proteins from mammalian cells specifically bind to certain mismatches, although their role in mismatch repair has not been demonstrated (JIRICNY *et al.* 1988; Stephenson and KARRAN 1989). Molecular analysis of the DHFR loci in both human and mouse has revealed an adjacant gene whose deduced amino acid sequence has a high degree of homology to *E. coli mutS* (FUJII and SHIMADA 1989; LINTON *et al.* 1989). No other evidence exists as to the potential function of these homologs in mismatch repair.

To characterize mismatch repair in *S. cerevisiae* more thoroughly, we decided to determine whether yeast possesses a homolog of the *E. coli* MutS and *S. pneumoniae* HexA proteins. We report here the cloning and sequence analysis of two yeast genes, *MSH1* and *MSH2* (*mutS homolog*) whose deduced amino acid sequences are highly homologous to *mutS* and *hexA*. We were surprised to find two *mutS* homologs in yeast and propose a possible differential function based on analysis of the sequence data. A companion study (REENAN and KOLODNER 1992) describes the phenotypes and genetics of insertion mutations in the two *MSH* genes and further supports the interpretation of function based on the sequence data.

MATERIALS AND METHODS

Enzymes and chemicals: Restriction enzymes were from New England Biolabs (Beverly, Massachusetts). T4 DNA ligase was prepared by R.D.K. using a method similar to that of TAIT, RODRIGUES and WEST (1980). The Klenow fragment of DNA polymerase I and a random primed DNA labeling kit were obtained from Boehringer Mannheim (Indianapolis, Indiana). *Taq* DNA polymerase was purchased from Perkin Elmer-Cetus (Norwalk, Connecticut). Sequenase DNA sequencing kits were from U. S. Biochemical Corp. (Cleveland, Ohio). [α-^{32}P]dATP used in random primed labeling and [α-^{35}S]dATP used in DNA sequencing were from Amersham (Arlington Heights, Illinois).

Oligonucleotides: Oligonucleotides were synthesized on an Applied Biosystems 380A DNA synthesizer using phosphoramidite chemistry and deprotected using standard methods. Degenerate oligonucleotides for polymerase chain reactions (PCR) were further purified by electrophoresis through a 15% denaturing acrylamide gel followed by purification on a Waters (Milford, Massachusetts) Sep/Pak column as per the manufacturers' instructions.

Strains and media: The *S. cerevisiae* strain NKY858 (*MATα ura3 lys2 leu2::hisG ho::LYS2 his4x*) used in this study for the isolation of genomic DNA is derived from SK1 and was the gift of NANCY KLECKNER (Harvard University, Cambridge, Massachusetts). Methods for the construction and manipulation of this strain have been described elsewhere (TISHKOFF, JOHNSON and KOLODNER 1991; CAO, ALANI and KLECKNER 1990). *E. coli* strain HB101 (BOYER and ROULAND-DUSSOIX 1969) was the host for the YCP50 library (ROSE *et al.* 1987). *E. coli* strain RK1400 (SYMINGTON, FOGARTY and KOLODNER 1983) was used as the host for all other plasmids. *E. coli* JM101 was the host for recombinant M13 phage (MESSING 1983). All *E. coli* strains were grown in L broth (LB) with appropriate antibiotics. Strains used for M13 infections were grown in 2xYT (MESSING 1983). M13 phage, the YCP50 library and all plasmids were from our laboratory collection.

Plasmids: Plasmids were constructed using standard procedures (SAMBROOK, FRITSCH and MANIATIS 1989). Small scale plasmid preparations were performed by the boiling

method of HOLMES and QUIGLEY (1981). Large scale plasmid preparations were prepared by a modification of the Triton-lysis method with subsequent purification of form-I plasmid DNA by centrifugation in CsCl-ethidium bromide density gradients (SAMBROOK, FRITSCH and MANIATIS 1989) DNA for double-stranded DNA sequencing was purified using two cycles of CsCl-EtBr density gradient centrifugation. Preparation of single-stranded M13 DNA for sequencing was essentially by the polyethylene glycol precipitation method (MESSING 1983). *E. coli* transformation procedures used were based on a standard Mg-Ca transformation procedure (WENSINK *et al.* 1974).

PCR amplification products of the *MSH1* and *MSH2* genes were inserted into the *Bam*HI site of M13mp19 to generate M13mp19-39 and M13mp19-45, respectively. These inserts will be referred to as ms351-I and ms351-II for convenience. pI-A5 (containing *MSH1*) contains a *Sau*3A partial digest fragment from chromosome *VIII* of *S. cerevisiae* inserted into the *Bam*HI site of YCP50. pII-2 (containing *MSH2*) contains a *Sau*3A fragment from chromosome *XV* of *S. cerevisiae* inserted into the *Bam*HI site of YCP50. These two plasmids and their less well characterized overlapping clones (see Figure 3, a and b) were recovered from the library constructed by ROSE *et al.* (1987).

PCR techniques: Based upon protein sequence comparisons, the following three regions of protein sequence were selected and used to design the indicated degenerate oligonucleotides: (1) F(A/V)THY, 5′-CTGGATCC(G/A)TA(G/A)TG(G/A/T/C)GT(G/A/T/C) (G/A)C(G/A)AA-3′; (2) TGPNM, 5′-CTGGATCCAC(G/A/T/C)GG (G/A/T/C)CC(G/A/T/C)AA(T/C)ATG-3′; and (3) H(P/A)V(V/I/M)(D/E), 5′-CTGGATCCA(T/C)(G/C)C(G/A/T/C) GT (GA/T/C)(G/A)T(G/A/T/C)GA-3′. The sequence CTGGATCC at the 5′ end of each oligonucleotide is a *Bam*HI restriction enzyme cleavage site added to facilitate cloning of the amplification product. PCR was performed in 50-μl volumes containing 10 mM Tris, pH 8.3, 3 mM $MgCl_2$, 50 mM KCl, 0.01% gelatin, 1.0 unit of *Taq* DNA polymerase, 25 pmol of each degenerate primer and 1 μg of yeast chromosomal DNA. The cycle for amplification using these degenerate oligonucleotides was as follows: (1) *denaturation* 1 min, 94°; (2) *annealing* 2 min, 55°; (3) *polymerization* 20 sec, 72°. The reaction was continued for 30 cycles. PCR amplification products for cloning were digested with *Bam*HI and passed over a Sephadex G-50 column run in 10 mM EDTA pH 8.0 to remove linkers and primers.

Colony hybridizations: Colonies were grown overnight on LB plates, lifted off onto Genescreen (Du Pont) and autoclaved at 120° for 2 min. The filters were washed in 40 mM $NaHPO_4$ buffer, pH 7.2, at 65° until all cellular debris was removed. Hybridization reactions contained: 0.5 M $NaHPO_4$ buffer, pH 7.2, 0.5% w/v bovine serum albumn, 1 mM EDTA, 5% sodium dodecyl sulfate (SDS) and 0.5 μg (10^8 cpm/μg) of ^{32}P-labeled probe made from the M13mp19 containing the appropriate 351-bp PCR product insert by the random priming method of FEINBERG and VOGELSTEIN (1983). Hybridization was allowed to proceed overnight at 60° followed by four 30-min washes with 40 mM $NaHPO_4$ buffer, pH 7.2, 1 mM EDTA and 1% SDS at 65°. Filters were exposed to x-ray film to detect the hybridizing colonies.

Southern hybridization analysis: DNA was transferred from agarose gels to Genescreen membrane (Du Pont) in 25 mM $NaHPO_4$ buffer, pH 6.5, and UV cross-linked to the membrane (CHURCH and GILBERT 1984). Hybridization was performed as described above except washes were done for 30 minutes with a solution containing 2 × SSC and 1% SDS at 65° with constant agitation. The hybridizing DNA bands were then detected by autoradiography.

DNA sequencing: Single-stranded M13 and double-stranded plasmid DNAs were sequenced by the dideoxy-chain termination method using Sequenase and the protocols supplied by the manufacturer. Double-stranded sequencing templates were prepared as follows: covalently closed circular template DNA was denatured in 0.2 M NaOH, 0.2 mM EDTA for 30 min at 37°. The mixture was neutralized with 0.1 volume of 3 M sodium acetate, pH 4.5, the DNA precipitated with 4 volumes of ethanol and resuspended in 5 mM Tris, pH 7.5, 0.5 mM EDTA. The Mn^{2+} sequencing buffer supplied by the manufacturer was used to determine DNA sequences close to the primer. The DNA sequences reported here have been submitted to GenBank under accession numbers M84169 for *MSH1* and M84170 for *MSH2*.

Sequence analysis: Homology searches and alignments were performed using the Eugene program (Lark Sequencing Technologies, Ltd., Houston, Texas) run on a Sun Microsystems Sparkstation 1. Sequence alignment of the various MutS homologs was performed by subdividing the sequence into smaller blocks of homology. The anchor points of these smaller domains were chosen based on the Lawrence homology search (LAWRENCE and GOLDMAN 1988), which defines homology domains between peptide sequences. The Dayhoff cost matrix of the Lawrence homology search was used which reports a minimum homology domain of 10 residues with a minimum acceptable standard deviation from chance of 3.0. Once regions of sequence were anchored by homology domains, the Altschul program (ALTSCHUL and ERICKSON 1986) was used to compute a globally optimal alignment using the SS2 algorithm. Both the Dayhoff and the genetic distance cost matrices were used with the Altschul program (ALTSCHUL and ERICKSON 1986). The penalty for gap opening was either 1.5 or 2.0 and the incremental penalty for each null in the gap was 1.0.

The amino-terminal 21 amino acids of MSH1 were analyzed in detail to identify features associated with mitochondrial targeting sequences. The presence of sequences with the potential to form amphiphilic helices was determined using the analysis of VON HEIJNE (1986). Estimations of hydrophobic moment, maximal hydrophobicity and surface seeking potential %surf and surf(E) were performed using the methods of EISENBERG, WEISS and TERWILLIGER (1984) and EISENBERG *et al.* (1984). The normalized consensus scale (EISENBERG, WEISS and TERWILLIGER 1984) was used in all calculations of hydrophobicity as follows: R = −2.53, K = −1.50, D = −0.90, Q = −0.85, *n* = −0.78, E = −0.74, H = −0.40, S = −0.78, T = −0.05, *P* = 0.12, Y = 0.26, C = 0.29, G = 0.48, A = 0.62, M = 0.64, W = 0.81, L = 1.06, V = 1.08, F = 1.19, I = 1.38.

RESULTS

PCR amplification of *mutS* homologs from *Saccharomyces cerevisiae*: Degenerate oligonucleotide primers were designed based upon protein sequence comparison of *mutS* homologs from two gram-negative bacteria, a gram-positive bacterium and human. Primers 1 and 2 lie on the boundary of a highly conserved region of 117 amino acids that is perfectly aligned between the four sequences (see Figure 1). Primers 1 and 2 yielded a product from *S. cerevisiae* genomic DNA that corresponded in size to the pre-

FIGURE 1.—Comparison of bacterial, mouse and human MutS homologs within the putative nucleotide binding site. The alignment of the nucleotide binding site domains of the various homologs is shown. The location of the sequences which were chosen for the synthesis of degenerate oligonucleotide primers 1 and 2 are marked with arrows. The sequences are designated as follows: coli, *E. coli* MutS (KEN STACY, unpublished results; also GenBank Accession number M64730); typh, *S. typhimurium* MutS (HABER *et al.* 1988); pneu, *S. pneumoniae* HexA (PRIEBE *et al.* 1988); mice, predicted product of the mouse *Rep-1* gene (LINTON *et al.* 1989 and personal communication from GRAY CROUSE); uman, predicted product of the human DUC-1 cDNA (FUJII and SHIMADA 1989). Boxed sequences represent identity with *E. coli* MutS.

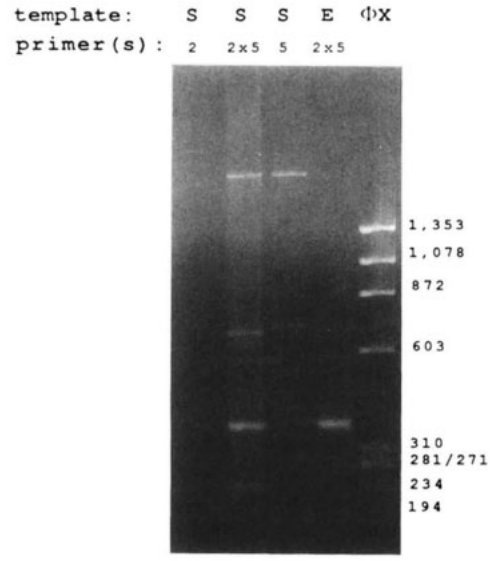


FIGURE 2.—The PCR amplification product from *S. cerevisiae* is identical in size with the amplification product from *E. coli*. *S. cerevisiae* genomic DNA (S) or *E. coli* genomic DNA (E) was used as template in PCR and the DNA products were analyzed by agarose gel electrophoresis. Reactions contain either primer 1 (5), primer 2 (2) or both primers 1 and 2 (2x5) as described under MATERIALS AND METHODS. The marker lane contains *Hae*III fragments of phiX174 RF DNA.

dicted 351 bp *E. coli* amplification product (Figure 2). Several of the minor and one of the major contaminating amplification products were the result of spurious priming amplification by one individual primer (Figure 2, control lanes). Primer 3 in combination with primer 1 did not result in significant amplification under any conditions.

Sequence analysis of the cloned products: Sequence analysis was performed using eight independent single-stranded M13 clones containing the cloned 351-bp fragment. Four of the clones contained one DNA sequence and the other four clones contained a different DNA sequence, whose predicted amino-acid sequence in both cases was highly homologous to the *mutS* homologs. The two different DNA sequences shared 56% identity and 66% similarity when conservative, third position changes were included. The predicted amino acid sequences shared 51% identity and 71% similarity when conservative amino acid substitutions were considered (see Figure 5). The two different 351 bp inserts were designated ms351-I and ms351-II.

Determination of the number of potentially clonable sequences: There could have been other clonable sequences in the 351-bp band that were not included in the eight sequenced clones. The sequence data from the two cloned fragments were used to devise a restriction digestion strategy which would selectively eliminate one or both fragments from the 351-bp band and thus allow detection of any remaining DNA. Two unique 6-bp cutting restriction enzymes were chosen for each fragment (*Hind*III and *Xba*I for ms351-I and *Mbo* II and *Eco*RI for ms351-II) and all combinations of restriction digestions were performed (data not shown). Single digests of the PCR amplification products showed that the ms351-I amplification product was 5–10-fold more abundant than the ms351-II product. In all four cases where double restriction digests should have cleaved both 351-bp fragments, no detectable DNA remained at the position of the 351-bp fragment. This analysis suggests that the primer set used yields only two major amplification products and that both had been cloned.

Genomic restriction map and full-length genomic clones: The two different 351 bp clones were uniformly ^{32}P-labeled and used to probe Southern blots of various restriction digested *S. cerevisiae* genomic DNAs and a YCP50-genomic library from *S. cerevisiae*. The probes detected single-copy DNA and allowed the construction of restriction maps of the genomic loci (see Figure 3, a and b). Four clones were obtained for the ms351-I probe and two for the ms351-II probe by screening approximately 15,000 colonies. The full length genes will be referred to as *MSH1* and *MSH2* (*mutS* homolog). The restriction maps of the clones are shown in Figure 3. pI-A5 and pII-2 were chosen for further analysis since they most likely contained the entire coding region. This assumption is based upon the size of the bacterial genes and the assumed location of the 351-bp conserved region close to the carboxyl terminus of the gene. High resolution restriction mapping and later PCR analysis of chromosomal DNA performed using pairs of sequencing primers demonstrated that the sequenced *MSH* loci present in pI-A5 and pII-2 corresponded to the sequences present in the genomic DNA. These two plasmids were used as hybridization probes against a library of ordered cloned *S. cerevisiae* chromosomal DNA frag-

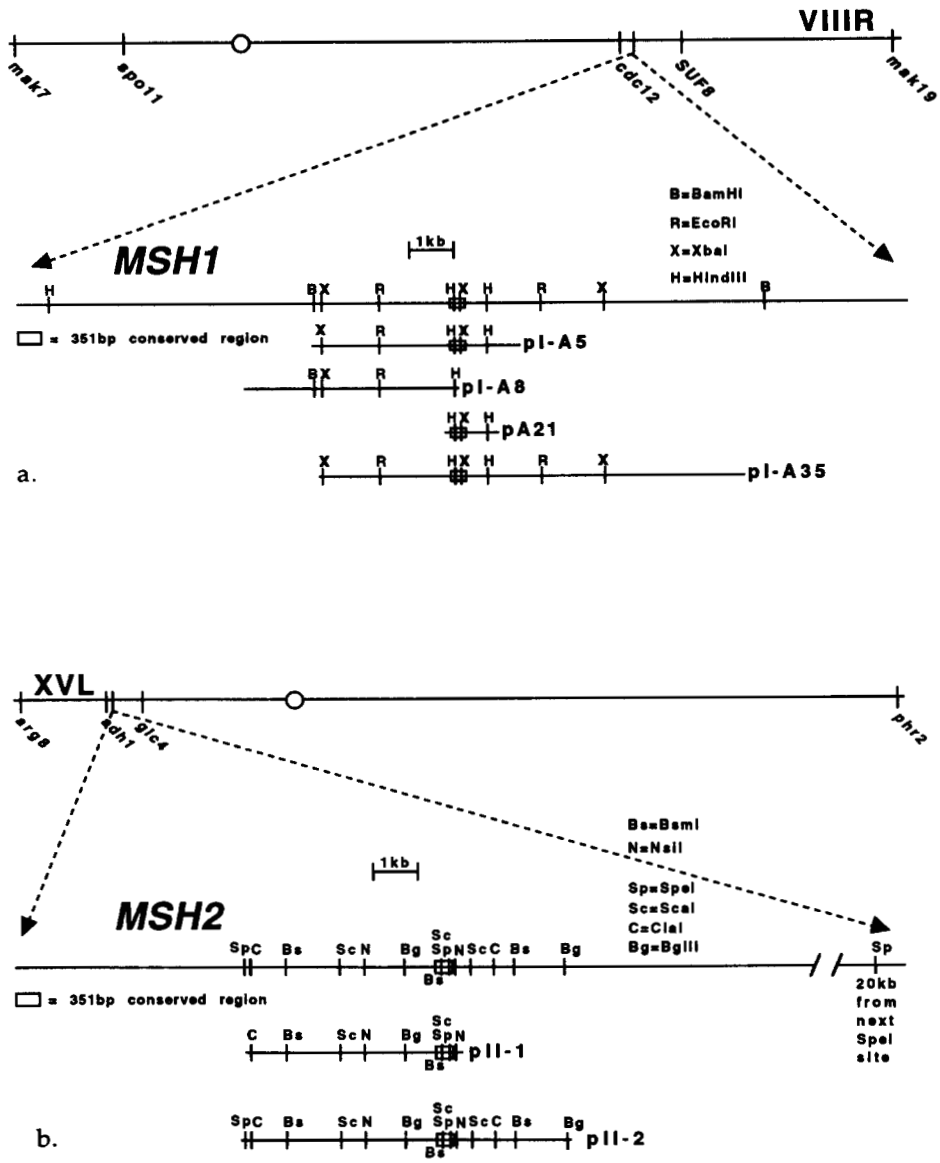
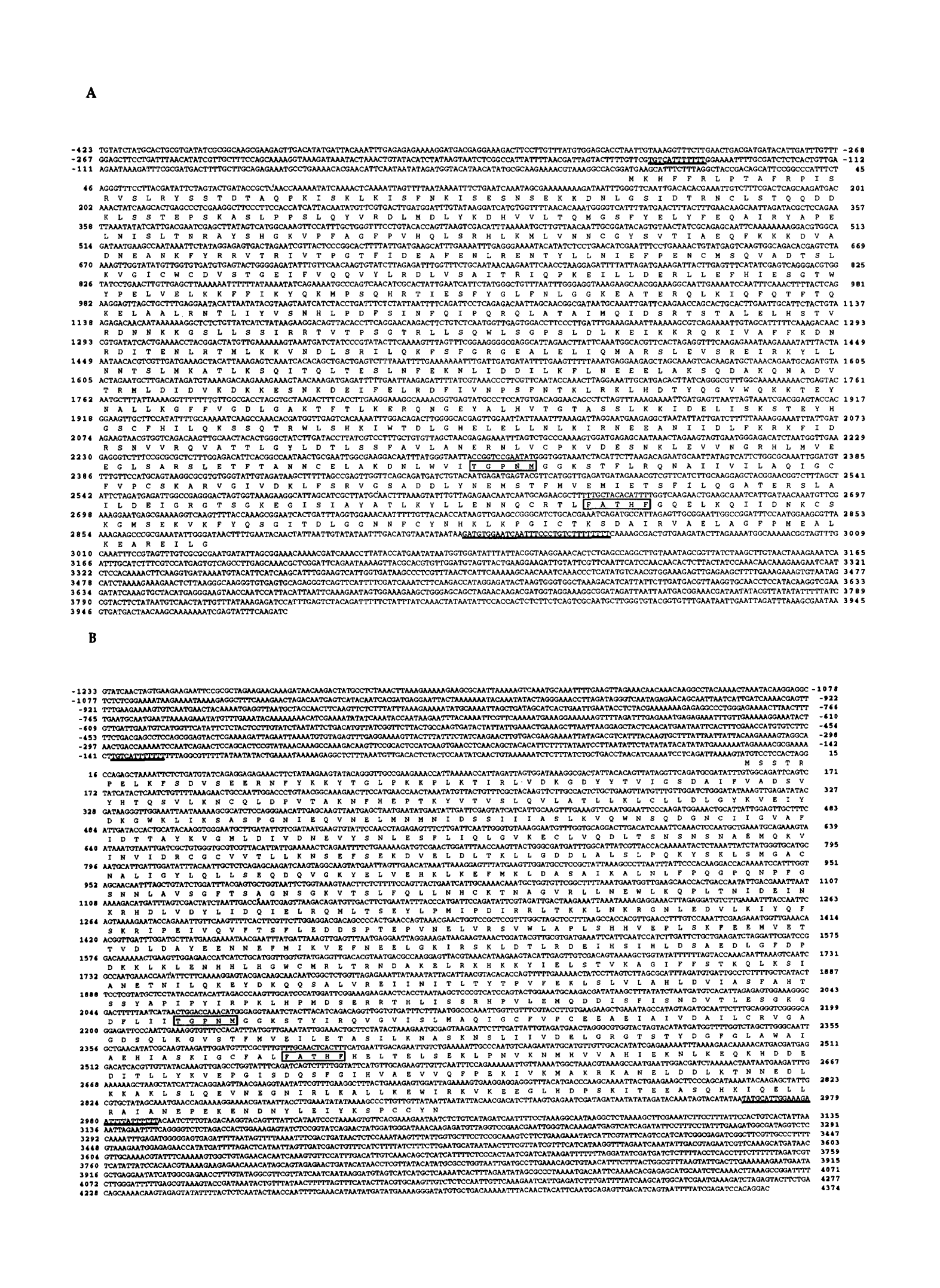


FIGURE 3.—Genomic restriction maps, YCP50 genomic library clones and physical map locations of the *MSH1* and *MSH2* genes. The genomic restriction maps of the *MSH1* and *MSH2* genes were generated by Southern analysis using the ms351-I and ms351-II products as probe. The 351-bp conserved region is indicated in the figure by an open box. Genomic YCP50 clones were isolated as described under MATERIALS AND METHODS. Physical map locations of the *MSH1* and *MSH2* genes were determined by hybridization of their respective YCP50 genomic clones, pI-A5 and pII-2 to an ordered cosmid library (L. RILES and M. OLSEN, Washington University, St. Louis, Missouri).

ments which allowed unambiguous determination of the chromosomal locations of the genes (L. RILES and M. OLSEN, Washington University, St. Louis, Missouri, personal communication). The *MSH1* gene is located on chromosome *VIIIR* between *cdc12* and *SUF8*, near *cdc12*, and the *MSH2* gene is located on chromosome *XVL* between *adh1* and *glc4*, near *adh1* (Figure 3).

The sequence of the *MSH1* gene: The sequence (Figure 4A) revealed a large open reading frame of 2877 bp capable of encoding a polypeptide of M_r 109,322. The sequence of 5′ flanking DNA revealed a 73% A-T rich region which extended from −98 to −180 relative to the putative start site (CHEN and STRUHL 1988). Within this region is a 12-bp sequence (TGTCATTTTTTT) at position −139 of *MSH1* that is identical to the 12-bp sequence at position −127 of *MSH2* (see Figure 4 and description of *MSH2* sequence below). A potential TATA element was found at −41 bp (WOBBE and STRUHL 1990). The sequence also revealed that the *MSH1* gene may be a poorly expressed gene based on the sequence context of the putative ATG (HAMILTON, WATANABE and DE BOER 1987) and the low codon-adaptation-index (C.A.I.) of 0.120 (SHARP *et al.* 1988). Downstream of the putative stop codon of *MSH1* are several potential polyadenylation sites (AATAA) (FITZGERALD and SHENK 1981). The *MSH1* gene possesses a transcriptional termination signal (ZARET and SHERMAN 1982) about 70 bp downstream of the putative stop codon (see Figure 4A).

Sequence comparison of the *MSH1* ORF with those of the bacterial and human *mutS* homologs showed striking similarity (Figure 5). The predicted MSH1 protein sequence shared 28.6% identity to the *mutS* gene product of *E. coli* and 45.5% similarity when conservative amino acid substitutions were considered. Several noteworthy features were found to be



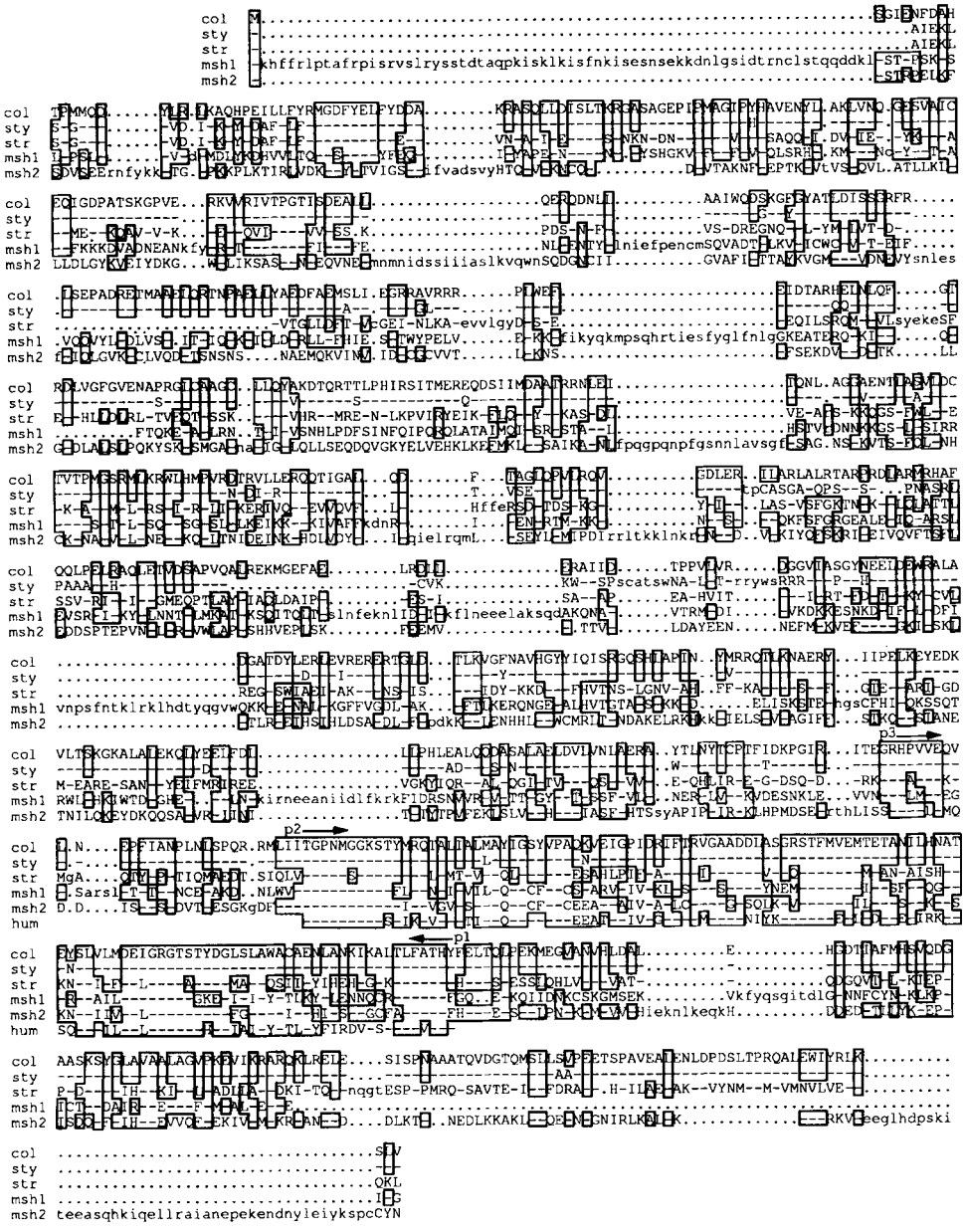


FIGURE 5.—Comparison of the predicted protein sequences of the *MSH1* and *MSH2* genes with bacterial and human MutS homologs. Sequences were aligned as described under MATERIALS AND METHODS. The positions of primers 1, 2 and 3 are indicated above the relevant sequence. Boxed regions represent either MSH protein identity with *E. coli* MutS (as indicated by –) or identity of one MSH protein and another non-*E. coli* MutS homolog (as indicated by lettering within a boxed region). Sequence abbreviations are as follows: col, *E. coli* MutS; sty, *S. typhimurium* MutS; str, *S. pneumoniae* HexA; hum, protein sequence predicted from the human DUC-1 cDNA. Other abbreviations used are: (.) a gap in the sequence; (–) identical amino acid; lower case, sequence present in one protein where there is a gap in all of the other proteins.

conserved between *MSH1* and the other *mutS* homologs, including *MSH2* (see below) and these features are summarized in Figure 6. First, the predicted MSH1 and MSH2 proteins each possess a putative Mg^{2+} ATP-binding site (Figures 5 and 6). The sequence comparison with other ATP-binding sites shows the highly conserved GKS/T sequence responsible for binding the phosphoryl moiety of ATP (GORBALENYA and KOONIN 1990). MSH1 also contains the invariant DE sequence responsible for chelating the Mg^{2+} ion of Mg^{2+} ATP. Second, a putative helix-turn-helix DNA binding motif (PABO and SAUER 1984) common to the carboxyl termini of the other *mutS* homologs was identified in MSH1 (Figure 6). However, the motif is located only 3 amino acids from the carboxyl terminus. This location for the helix-turn-helix differs from all other known *mutS* homologs, which have at least 50 amino acids between the end of the helix-turn-helix and the carboxyl termini (see Figure 5). Third, another conserved region was found in the amino terminus (MGDFY to RIVTPG corresponding to bases 304–567 on the *MSH1* DNA sequence and amino acids 101–189 on the protein sequence, see Figures 4A and 5). This region showed a high degree of conservation with the bacterial but not the human or MSH2 sequences. Overall, the MSH1

FIGURE 4.—Complete sequence of the *MSH1* and *MSH2* genes of *S. cerevisiae*. The sequences of the *MSH1* (A) and *MSH2* (B) genes are presented along with 5′ and 3′ flanking sequences. The boxed protein sequences represent the sequences to which primers 1 and 2 were directed. The underlined sequence to the 5′ of the AUG of both genes indicates the shared TGTCATTTTTTT sequence. The underlined sequences 3′ to the putative stop codons are potential transcriptional termination sites. The adenine of the putative start codon, in both cases, is designated as position number 1.

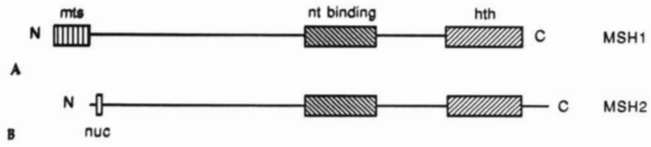


FIGURE 6.—Illustration of structural features predicted to be present in MSH1 and MSH2 proteins. (A) MSH1 protein. Features illustrated are: mitochondrial targeting sequence located between amino acid positions 1 and 21 (mts); nucleotide binding fold GKS located at amino acid position 776 and DE located at amino acid position 850 (nt binding); helix-turn-helix AIRVAELAGFPME-ALKEARE located at amino acid position 937–956 (hth). (B) MSH2 protein. Features illustrated are: nuclear targeting sequences FY**KK** YTGLP**K** and L**PKK** PL**K** TIR located at amino acid positions 19 and 26, respectively (nuc); nucleotide binding fold GKS located at amino acid position 693 and DE located at amino acid position 767 (nt binding); helix-turn-helix GIHVAEVVQFPEKIVKMAKR located at amino acid position 855–874 (hth).

protein is more homologous, especially in the amino terminus, to the bacterial *mutS* homologs than to the predicted gene products of the *MSH2* gene or the human gene.

Amino-terminal non-homology in the MSH1 protein: The predicted MSH1 protein has a region of 68 amino acids at the amino terminus that is not present in the other proteins (Figures 5 and 6). Analysis of the 21 amino-terminal residues revealed an exceptionally good mitochondrial targeting sequence by several criteria. First, studies have shown that many mitochondrial targeting sequences possess the potential to form amphiphilic helices (VON HEIJNE 1986) and such a potential amphiphilic helix is present in the first 21 amino acids of MSH1. Second, the calculated hydrophobic moment, μ_H,and the maximum hydrophobicity of the hydrophobic face of MSH1 (amino acids 4–21) are among the highest observed for amphiphilic helices known to function as mitochondrial targeting sequences and, among the MutS homologs, only MSH1 has an amino terminus that qualifies as a surface-seeking peptide by these criteria (VON HEIJNE 1986). Third, analysis by the method of EISENBERG, WEISS and TERWILLIGER (1984) and EISENBERG *et al.* (1984) showed that only the MSH1 amino terminus qualified as a surface seeking peptide, although it should be noted that many mitochondrial targeting sequences fail to satisfy the Eisenberg criteria (VON HEIJNE 1986). This analysis suggests that the protein encoded by *MSH1* is imported into mitochondria and functions there.

The sequence of the *MSH2* gene: Sequence analysis of the *MSH2* genomic clone revealed a large open reading frame of 2898 bp capable of encoding a polypeptide of M_r 109,073. The sequence of 5′ flanking DNA revealed an 82% A-T rich region extending from −93 to −198 relative to the putative start site. Within this region is the 12-bp identity with the upstream region of *MSH1* mentioned above (at −127, see Figure 4B). Potential TATA elements were located at −45 and −54. The *MSH2* sequence also revealed that it may be a moderately expressed gene. The sequence context of the putative ATG is near the consensus for highly expressed genes and the codon-adaptation index (0.177) is indicative of a moderately expressed gene. As with the *MSH1* gene, potential polyadenylation sites are present downstream of the *MSH2* stop codon. A match to the consensus transcriptional termination signal is found at nearly the same position downstream of the stop codon as in the *MSH1* gene. Comparison of the sequence predicted for the MSH2 protein with the bacterial and human MutS homologs revealed striking similarity (Figure 5). MSH2 shared 25.2% identity with *E. coli* MutS and 42.7% similarity when conservative amino acid substitutions were considered. A comparison between MSH1 and MSH2 revealed that they are 17.2% identical and 31.5% similar when conservative amino acid substitutions are considered. Thus, the MSH1 and MSH2 proteins are less homologous to each other than either is to its bacterial counterparts.

As with the predicted MSH1 protein, the predicted MSH2 protein also has shared structural motifs with the other MutS homologs. First, the Mg^{2+} ATP-binding site shares the invariant sequences GKS/T and DE with all of the MutS homologs including MSH1 as well as highly conserved blocks of homology throughout the region (see Figures 5 and 6). Second, the helix-turn-helix motif is present and located 92 amino acids from the carboxyl terminus (see Figures 5 and 6) so that the carboxyl terminus of MSH2 is extended by an additional 30–40 amino acids from this feature with respect to the other homologs (except MSH1). The length of the amino terminus of MSH2 appears to closely match the bacterial sequences (see Figure 5).

Overall, the predicted MSH2 protein contains less extensive amino terminal homology to the bacterial homologs than does MSH1 (for MSH1 there is 37% identity to the amino terminal 156 amino acids of the *E. coli* MutS protein, whereas MSH2 shows only 16% identity for the same region). The amino terminus of MSH2 contains two potential nuclear localization sites at amino acid positions 19 and 26 (MORELAND *et al.* 1987; RICHARDSON, ROBERTS and SMITH 1986) (Figure 6). The presence of putative nuclear localization sequences suggests that *MSH2* functions in the nucleus.

DISCUSSION

We have identified two yeast genes, *MSH1* and *MSH2*, whose putative gene products show a high degree of homology (>25%) to the predicted products of the *E. coli mutS* gene and the *S. pneumoniae hexA* gene which function in mismatch repair and mammalian *mutS* homologs for which no function has yet

been identified. The rationale for searching for such homologs was provided by the identification of a *mutL* homolog (*PMS1*) in yeast (KRAMER *et al.* 1989a; PRUDHOMME *et al.* 1989) and the existence of higher eucaryotic homologs of *S. pneumoniae hexA* and *E. coli mutS*. We were surprised to identify two *mutS* homologs in yeast.

The predicted molecular masses of the *MSH1* and *MSH2* gene products are nearly identical even though they share only 17% identity. The yeast proteins (109 kD) are larger than the procaryotic proteins (95 kD) and smaller than the mouse or human proteins (123 and 127 kD, respectively). Another feature of the primary structure of the two yeast proteins is their carboxyl termini. The MSH1 carboxyl terminus is located 3 amino acids after the putative helix-turn-helix, whereas the carboxyl terminus of the procaryotic and mammalian proteins is located approximately 60 amino acids after the helix-turn-helix. The carboxyl terminus of the MSH2 protein is 40 amino acids further from the helix-turn-helix than the procaryotic or mammalian proteins and almost 100 amino acids further from the helix-turn-helix than in the case of the MSH1 carboxyl terminus. The carboxyl termini regions of both MSH1 and MSH2 have in common the highly conserved nucleotide binding domain and the putative helix-turn-helix motifs. This strongly suggests that the carboxyl-terminal region of the MSH proteins possess the DNA binding region rather than a region for potential MSH/MutL homolog interaction. One might expect to find such a MutL interaction domain in a more divergent region such as in the amino-terminal half of the MSH proteins. This is because we imagine there are less evolutionary constraints on protein-protein interactions where compensating changes can occur in two proteins than on protein-DNA interactions where the structure of the DNA containing the mispair is fixed.

The amino-terminal regions of the MSH1 and MSH2 proteins provide a preliminary indication of the function and localization of these proteins. The amino terminus of the MSH1 protein possesses a potential mitochondrial targeting sequence that is absent in the other MutS homologs and suggests that MSH1 functions in the mitochondria. Recent analysis of the purified MSH1 protein indicates that its amino terminus is processed in a manner consistent with its being imported into mitochondria (N.-W. CHI, E. ALANI and R. KOLODNER, unpublished). One the other hand, the MSH2 amino terminus possesses two potential nuclear localization sites, suggesting that the MSH2 functions in the nucleus. The strong homology of MSH1 and MSH2 to the procaryotic MutS homologs known to function in mismatch repair suggests that the yeast homologs are involved in mismatch repair of DNA.

The data presented in this report suggest that yeast possesses a mismatch repair system similar to the *hexAB* system of *S. pneumoniae* and the *mutHLS* system of *E. coli* even though the yeast system is unlikely be methyl-directed like the *E. coli* system. A third yeast *mutS* homolog (now called *MSH3*), which could not have been identified using our primer set, has been identified and is more closely related to the mammalian *mutS* homologs than *MSH1* and *MSH2* are (GRAY CROUSE, personal communication). The observation of multiple homologs in yeast suggests that it may not be straightforward to define a function for the single presently identified mammalian gene encoding a *mutS* homolog. The existence of three yeast *mutS* homologs suggests the possible existence of multiple mismatch repair pathways and more *mutL* homologs than just *PMS1*. This is true since the *MSH1* and *MSH2* genes appear to have repair functions in different organelles as demonstrated in a companion study (REENAN and KOLODNER 1992). A remaining question is whether yeast contains homologs of *uvrD* and *mutH*. Yeast lacks detectable DNA methylation (PROFFITT *et al.* 1984) and therefore is unlikely to possess a methyl-directed mismatch repair system. This suggests that *mutH* homologs, which would be expected to recognize asymmetries in DNA methylation (LU, CLARK and MODRICH 1983; LAHUE, SU and MODRICH 1987; WELSH *et al.* 1987), may not be required in yeast. However, there may be as yet unidentified *mutH* homologs responsible for directing the strand specificity of the repair process in some way other than recognizing the state of methylation. The *RAD3* gene, which encodes a helicase, and the *HPR5/RADH/SRS2* genes are homologous to the *E. coli uvrD* gene, which is the DNA helicase involved in UV excision repair and methyl-directed mismatch repair excision (MODRICH 1987; ABOUSSEKHRA *et al.* 1989; RONG *et al.* 1991; SUNG *et al.* 1987). Although some alleles of *HPR5/RADH/SRS2* confer a hyper-gene conversion phenotype (RONG *et al.* 1991) and one allele of *RAD3* confers increased spontaneous mutability (GOLIN and ESPOSITO 1977; MONTELONE, HOEKSTRA and MALONE 1988), neither the *HPR5/RADH/SRS2* gene nor the *RAD3* gene have been definitively implicated as encoding the mismatch repair helicase of yeast. The multiple *mutS* homologs of yeast suggest that there may also be multiple as yet unidentified *uvrD* homologs.

The authors would like to thank NANCY KLECKNER for the strains she provided, RHONA BORTS, JIM HABER, DAN TISHKOFF and CHRIS WALSH for helpful discussions and ERIC ALANI, ARLEN JOHNSON and PAUL MODRICH for their comments on the manuscript. We are particularly grateful to GRAY CROUSE for sharing his unpublished results with us. This study was supported by National Institutes of Health grant HG00305 to R.D.K.

LITERATURE CITED

ABOUSSEKHRA, A., R. CHANET, Z. ZGAGA, C. CASSIER-CHAUVAT, M. HEUDE and F. FABRE, 1989 *RADH*, a gene of *Saccharomyces cerevisiae* encoding a putative DNA helicase involved in DNA repair. Nucleic Acids Res. **17:** 7211–7219.

ALTSCHUL, S. F., and B. W. ERICKSON, 1986 Optimal sequence alignment using affine gap costs. Bull. Math. Biol. **48:** 603–616.

BISHOP,D. K., J. ANDERSON and R. D. KOLODNER, 1989 Specificity of mismatch repair following transformation of *Saccharomyces cerevisiae* with heteroduplex plasmid DNA. Proc. Natl. Acad. Sci. USA **86:** 3713–3717.

BISHOP, D., and R. KOLODNER, 1986 Repair of heteroduplex plasmid DNA after transformation into *Saccharomyces cerevisiae*. Mol. Cell. Biol. **6:** 3401–3409.

BISHOP, D., M. S. WILLIAMSON, S. FOGEL and R. KOLODNER, 1987 The role of heteroduplex correction in gene conversion in *Saccharomyces cerevisiae*. Nature **328:** 362–364.

BORTS, R. H., W. LEUNG, W. KRAMER, B. KRAMER, M. WILLIAMSON, S. FOGEL and J. E. HABER, 1990 Mismatch repair-induced meiotic recombination requires the *PMS1* gene product. Genetics **124:** 573–584.

BOYER, H. W., and D. ROULLAND-DUSSOIX, 1969 A complementation analysis of the restriction and modification of DNA in *Escherichia coli*. J. Mol. Biol. **41:** 459–472.

BROWN, T. C., and J. JIRICNY, 1987 A specific mismatch repair event protects mammalian cells from loss of 5-methylcytosine. Cell **50:** 945–950.

BROWN, T. C., and J. JIRICNY, 1988 Different base/base mispairs are corrected with different efficiencies and specificities in monkey kidney cells. Cell **54:** 705–711.

CAO, L., ALANI, E. and N. KLECKNER, 1990 A pathway for generation and processing of double-strand breaks during meiotic recombination in *S. cerevisiae*. Cell **61:** 1089–1101.

CHEN, W., and K. STRUHL, 1988 Saturation mutagenesis of a yeast *his3* "TATA element": genetic evidence for a specific TATA-binding protein. Proc. Natl. Acad. Sci. USA **85:** 2691–2695.

CHURCH, G. M., and W. GILBERT, 1984 Genomic sequencing. Proc. Natl. Acad. Sci. USA **81:** 1991–1995.

CLAVERYS, J. P., and S. A. LACKS, 1986 Heteroduplex deoxyribonucleic acid base mismatch repair in bacteria. Microbiol. Rev. **50:** 133–165.

COX, E. C., 1976 Bacterial mutator genes and the control of spontaneous mutation. Annu. Rev. Genet. **10:** 135–156.

EISENBERG,D., R. M. WEISS and T. C. TERWILLIGER, 1984 The hydrophobic moment detects periodicity in protein hydrophobicity. Proc. Natl. Acad. Sci. USA **81:** 140–144.

EISENBERG, D., E. SCHWARZ, M. KOMAROMY and R. WALL, 1984 Analysis of membrane and surface protein sequences with the hydrophobic moment plot. J. Mol. Biol. **179:** 125–142.

FEINBERG, A. P., and B. VOGELSTEIN, 1983 A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. **132:** 6–13.

FITZGERALD, M., and T. SHENK, 1981 The sequence 5′-AAUAAA-3′ forms part of the recognition site for polyadenylation of late SV40 mRNAs. Cell **24:** 251–260.

FUJII, H., and T. SHIMADA, 1989 Isolation and characterization of cDNA clones derived from the divergently transcribed gene in the region upstream from the human dihydrofolate reductase gene. J. Biol. Chem. **264:** 10057–10064.

GLICKMAN, B., P. VAN DEN ELSEN and M. RADMAN, 1978 Induced mutagenesis in *dam*-mutants of *Escherichia coli*: A role for 6-methyladenine residues in mutation avoidance. Mol. Gen. Genet. **163:** 307–312.

GLICKMAN, B. W., and M. RADMAN, 1980 *Escherichia coli* mutator mutants deficient in methylation-instructed DNA mismatch correction. Proc. Natl. Acad. Sci. USA **77:** 1063–1067.

GOLIN, J., and M. S. ESPOSITO, 1977 Evidence for joint genetic control of spontaneous mutation and genetic recombination during mitosis in *Saccharomyces*. Mol. Gen. Genet. **150:** 127–135.

GORBALENYA, A. E., and E. V. KOONIN, 1990 Superfamily of uvrA-related NTP-binding proteins. Implications for rational classification of recombination repair systems. J. Mol. Biol. **213:** 583–591.

GRILLEY, M., K. M. WELSH, S. SU and P. MODRICH, 1989 Isolation and characterization of the *Escherichia coli mutL* gene product. J. Biol. Chem. **264:** 1000–1004.

HABER, L. T., and G. C. WALKER, 1991 Altering the conserved nucleotide binding motif in the *Salmonella typhimurium* MutS mismatch repair protein affects its ATPase and mismatch binding activities. EMBO J. **10:** 2707–2715.

HABER, L. T., P. P. PANG, D. I. SOBELL, J. A. MANKOVITCH and G. C. WALKER, 1988 Nucleotide sequence of the *Salmonella typhimurium mutS* gene required for mismatch repair: homology of *mutS* and *hexA* of *Streptococcus pneumoniae*. J. Bacteriol. **170:** 197–202.

HAMILTON, R., C. K. WATANABE and H. A. DE BOER, 1987 Compilation and comparison of the sequence context around the AUG start codons in *Saccharomyces cerevisiae* mRNAs. Nucleic Acids Res. **15:** 3581–3593.

HOLMES, D. S., and M. QUIGLEY, 1981 A rapid boiling method for the preparation of bacterial plasmids. Anal. Biochem. **114:** 193–197.

HOLMES, J., S. CLARK and P. MODRICH, 1990 Strand-specific mismatch correction in nuclear extracts of human and *Drosophila melanogaster* cell lines. Proc. Natl. Acad. Sci. USA **87:** 5837–5841

JIRICNY, J., M. HUGHES, N. CORMAN and B. B. RUDKIN, 1988 A human 200-kDa protein binds selectively to DNA fragments containing G/T mismatches. Proc. Natl. Acad. Sci. USA **85:** 8860–8864.

KRAMER, B., W. KRAMER and H. J. FRITZ, 1984 Different base/base mismatches are corrected with different efficiencies by the methyl-directed DNA mismatch-repair system of *E. coli*. Cell **38:** 879–887.

KRAMER, B., W. KRAMER, M. S. WILLIAMSON and S. FOGEL, 1989a Heteroduplex DNA correction in *Saccharomyces cerevisiae* is mismatch specific and requires functional *PMS* genes. Mol. Cell. Biol. **9:** 4432–4440.

KRAMER, W., B. KRAMER, M. S. WILLIAMSON and S. FOGEL, 1989b Cloning and nucleotide sequence of DNA mismatch repair gene *PMS1* from *Saccharomyces cerevisiae*: homology of *PMS1* to procaryotic *mutL* and *hexB*. J. Bacteriol. **171:** 5339–5346.

LAHUE, R. S., S. S. SU and P. MODRICH, 1987 Requirement for d(GATC) sequences in *Escherichia coli mutHLS* mismatch correction. Proc. Natl. Acad. Sci. USA **84:** 1482–1486.

LAWRENCE, C. B., and D. A. GOLDMAN, 1988 Definition and identification of homology domains. Comput. Appl. Biosci. **4:** 25–31.

LICHTEN, M., C. GOYON, N. P. SCHULTES, D. TRECO, J. W. SZOSTAK, J. E. HABER and A. NICOLAS, 1990 Detection of heteroduplex DNA molecules among the products of *Saccharomyces cerevisiae* meiosis. Proc. Natl. Acad. Sci. USA **87:** 7653–7657.

LINTON, J. P., J. J. YEN, E. SELBY, Z. CHEN, J. M. CHINSKY, K. LIU, R. E. KELLEMS and G. F. CROUSE, 1989 Dual bidirectional promoters at the mouse dhfr locus: cloning and characterization of two mRNA classes of the divergently transcribed *Rep-1* gene. Mol. Cell. Biol. **9:** 3058–3072.

LU, A. L., S. CLARK and P. MODRICH, 1983 Methyl-directed repair of DNA base-pair mismatches *in vitro*. Proc. Natl. Acad. Sci. USA **80:** 4639–4643.

MATSON, S. W., and J. W. GEORGE, 1897 DNA helicase II of

Escherichia coli. Characterization of the single-stranded DNA-dependent NTPase and helicase activities. J. Biol. Chem. **262:** 2066–2076.

MESSING, J., 1983 New M13 vectors for cloning. Methods Enzymol. **101:** 10–77.

MODRICH, P., 1987 DNA mismatch correction. Annu. Rev. Biochem. **56:** 436–466.

MORELAND, R. B., G. L. LANGEVIN, R. H. SINGER, R. L. GARCEA and L. M. HEREFORD, 1987 Amino acid sequences that determine the nuclear localization of yeast histone 2B. Mol. Cell. Biol. **7:** 4048–4057.

MONTELONE, B. A., M. F. HOEKSTRA and R. E. MALONE, 1988 Spontaneous mitotic recombination in yeast: the hyper-recombinational *rem1* mutations are alleles of the *RAD3* gene. Genetics **119:** 289–301.

MUSTER-NASSAL, C., and R. KOLODNER, 1986 Mismatch correction catalyzed by cell-free extracts of *Saccharomyces cerevisiae*. Proc. Natl. Acad. Sci. USA **83:** 7618–7622.

PABO, C. O., and R. T. SAUER, 1984 Protein-DNA recognition. Annu. Rev. Biochem. **53:** 293–321.

PANG, P. P., S.-D. TSEN, A. S. LUNDBERG and G. C. WALKER, 1984 The *mutH*, *mutL*, *mutS*, and *uvr*D genes of *Salmonella typhimurium* LT2. Cold Spring Harbor Symp. Quant. Biol **49:** 597–602.

PRIEBE, S. D., S. H. HADI, B. GREENBERG and S. A. LACKS, 1988 Nucleotide sequence of the *hexA* gene for DNA mismatch repair in *Streptococcus pneumoniae* and homology of *hexA* to *mutS* of *Escherichia coli* and *Salmonella typhimurium*. J. Bacteriol. **170:** 190–196.

PROFFITT, J. H., J. R. DAVIE, D. SWINTON and S. HATTMAN, 1984 5-Methylcytosine is not detectable in *Saccharomyces cerevisiae* DNA. Mol. Cell. Biol. **4:** 985–988.

PRUDHOMME, M., B. MARTIN, V. MEJEAN and J. CLAVERYS, 1989 Nucleotide sequence of the *Streptococcus pneumoniae hexB* mismatch repair gene: homology of *HexB* to *MutL* of *Salmonella typhimurium* and to *PMS1* of *Saccharomyces cerevisiae*. J. Bacteriol. **171:** 5332–5338.

REENAN, R. A. G., and R. D. KOLODNER, 1992 Characterization of insertion mutations in the *Saccharomyces cerevisiae MSH1* and *MSH2* genes: evidence for separate mitochondrial and nuclear function. Genetics **132:** xxx–xxx.

RICHARDSON, W. D., B. L. ROBERTS and A. E. SMITH, 1986 Nuclear localization signals in polyoma virus large-T. Cell **44:** 77–85.

RONG, L., F. PALLADINO, A. AGUILERA and H. L. KLEIN, 1991 The hyper-gene conversion *hpr5–1* mutation of *Saccharomyces cerevisiae* is an allele of the *SRS2/RADH* gene. Genetics **127:** 75–85.

ROSE, M. D., P. NOVICK, J. H. THOMAS, D. BOTSTEIN and G. R. FINK, 1987 A *Saccharomyces cerevisiae* genomic plasmid bank based on a centromere-containing shuttle vector. Gene **60:** 237–243.

SAMBROOK, J., E. F. FRITSCH and T. MANIATIS, 1989 *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

SHARP, P. M., E. COWE, D. G. HIGGINS, D. C. SHIELDS, K. H. WOLFE and F. WRIGHT, 1988 Codon usage patterns in *Escherichia coli, Bacillus subtilis, Saccharomyces cerevisiae, Schizosaccharomyces pombe, Drosophila melanogaster* and *Homo sapiens*; a review of the considerable within-species diversity. Nucleic Acids Res. **16:** 8207–8211.

STEPHENSON, C., and P. KARRAN, 1989 Selective binding to DNA base-pair mismatches by proteins from human cells. J. Biol. Chem. **264:** 21177–21182

SU, S., and P. MODRICH, 1986 *Escherichia coli mutS* -encodedprotein binds to mismatched DNA base pairs. Proc. Natl. Acad. Sci. USA **83:** 5057–5061.

SUNG, P., L. PRAKASH, S. W. MATSON and S. PRAKASH, 1987 RAD3 protein of *Saccharomyces cerevisiae* is a DNA helicase. Proc. Natl. Acad. Sci. USA **84:** 8951–8955

SYMINGTON, L. S., L. M. FOGARTY and R. KOLODNER, 1983 Genetic recombination of homologous plasmids catalyzed by cell-free extracts of *Saccharomyces cerevisiae*. Cell **35:** 805–813.

TAIT, R. C., R. L. RODRIGUES and R. W. WEST, 1980 The rapid purification of T4 DNA ligase from a *T4 lig* lysogen. J. Biol. Chem. **255:** 813–816.

THOMAS, D. C., J. D. ROBERTS and T. A. KUNKEL, 1991 Heteroduplex repair in extracts of human HeLa cells. J. Biol. Chem. **266:** 3744–3751

TISHKOFF, D., A. W. JOHNSON and R. KOLODNER, 1991 Molecular and genetic analysis of the gene encoding the *Saccharomyces cerevisiae* strand exchange protein SEP1. Mol. Cell. Biol. **11:** 2593–2608.

VARLET, I., M. RADMAN and B. BROOKS, 1990 DNA mismatch repair in *Xenopus* egg extracts: repair efficiency and DNA repair synthesis for all single base-pair mismatches. Proc. Natl. Acad. Sci. USA **87:** 7883–7887.

VON HEIJNE, G., 1986 Mitochondrial targeting sequences may form amphiphilic helices. EMBO J. **5:** 1335–1342.

WELSH, K. M., A. L. LU, S. CLARK and P. MODRICH, 1987 Isolation and characterization of the *Escherichia coli mutH* gene product. J. Biol. Chem. **262:** 15624–15621.

WENSINK, P. C., D. J. FINNEGAN, J. E. DONELSON and D. S. HOGNESS, 1974 A system for mapping DNA sequences in the chromosomes of *Drosophila melanogaster*. Cell **3:** 315–325.

WILLIAMSON, M. S., J. C. GAME and S. FOGEL, 1985 Meiotic gene conversion mutants in *Saccharomyces cerevisiae*. I. Isolation and characterization of *pms1-1* and *pms1-2*. Genetics **110:** 609–646.

WOBBE, C. R., and K. STRUHL, 1990 Yeast and human TATA-binding proteins have nearly identical DNA sequence requirements for transcription *in vitro*. Mol. Cell Biol. **10:** 3859–3867.

ZARET, K. S., and F. SHERMAN, 1982 DNA sequence required for efficient transcription termination in yeast. Cell **28:** 563–573.

Communicating editor: G. R. SMITH