Volume 13 Number 7 1985 Nucleic Acids Research

The nucleotide sequence of the *RAD3* gene of *Saccharomyces cerevisiae:* a potential adenine nucleotide binding amino acid sequence and a nonessential acidic carboxyl terminal region

Paul Reynolds, David R.Higgins*, Louise Prakash and Satya Prakash*

Department of Radiation Biology and Biophysics, University of Rochester School of Medicine, Rochester, NY 14642, and *Department of Biology, University of Rochester, Rochester, NY 14627, USA

Received 21 December 1984; Revised and Accepted 8 March 1985

ABSTRACT

The *RAD3* gene of *Saccharomyces cerevisiae* is required for excision of pyrimidine dimers and is essential for viability. We present the nucleotide sequence of the *RAD3* protein coding region and its flanking regions, and the deduced primary structure of the RAD3 protein. In addition, we have mapped the 5' end of *RAD3* mRNA. The predicted RAD3 protein contains 778 amino acids with a calculated molecular weight of 89,779. A segment of the RAD3 protein shares homology with several adenine nucleotide binding proteins, suggesting that RAD3 protein may react with ATP. The twenty carboxyl terminal amino acids of RAD3 protein are predominantly acidic; however, deletion of this acidic region has no obvious effect on viability or DNA repair.

INTRODUCTION

In the yeast *Saccharomyces cerevisiae*, excision of pyrimidine dimers or interstrand DNA crosslinks requires a large number of genes - *RAD1*, *RAD2*, *RAD3*, *RAD4*, *RAD10*, *MMS19*, *RAD7*, *RAD14*, *RAD16*, and *RAD23* (1-12). A mutation in any of the first six of the ten genes listed results in highly defective incision of DNA containing pyrimidine dimers (13,14) or interstrand DNA crosslinks (8,11,12), while a mutation in *RAD14*, produces reduced incision proficiency compared with the Rad$^+$ strain (9,13). The *RAD3* gene has been cloned and partially characterized (15,16). Previously, we had localized the *RAD3* gene to a DNA fragment of approximately 2.6 kb, and identified a 2.5 kb *RAD3* mRNA and determined its direction of transcription (15). By integrating plasmids containing different internal fragments of the *RAD3* gene in the yeast chromosomal *RAD3* site, we, and others, deleted part of the *RAD3* gene and found these deletions to be recessive lethal (15,17), indicating that *RAD3* plays an essential role in cellular processes in addition to incision of damaged DNA. This finding is in contrast to the effect of *rad1*, *rad2*, and *rad10* deletions or disruptions, which are viable (18-21) and suggests that the *RAD3* gene plays a more complex role *in vivo* than do these other genes involved in incision of pyrimidine dimer-containing DNA.

In this paper, we have mapped the 5' end of the *RAD3* mRNA and determined

the nucleotide sequence of 3383 bp of *RAD3* region DNA. The 5' mRNA end maps 117 nucleotides upstream of the first translation initiating ATG codon. The coding region of *RAD3* is 2334 nucleotides long, which encodes a protein of 778 amino acids, with a calculated molecular weight of 89,779. The carboxyl terminal region of the RAD3 protein is highly acidic: the last 20 amino acids contain 12 acidic and 1 basic residue, including 7 consecutive acidic residues. The *RAD3* gene deleted for the acidic carboxyl terminus in plasmid pSP6 restores wild type levels of viability and ultraviolet light resistance to a yeast strain lacking the entire chromosomal *RAD3* gene. The RAD3 protein contains a sequence similar to the consensus sequence that has been identified in the *Escherichia coli* UvrD, RecA, and DnaB proteins, and in several other adenine nucleotide binding proteins, and which may be involved in ATP binding.

METHODS

Mapping the 5' mRNA Terminus

The 5' terminus of the *RAD3* transcript was mapped by a modification (22) of the Berk and Sharp method (23), using S1 nuclease and a 5' end-labeled single-stranded DNA probe. A *Kpn*I-*Hin*dIII fragment from the 5' end of the *RAD3* gene (Fig. 1) was dephosphorylated by bacterial alkaline phosphatase, and labeled with [γ-^{32}P]ATP by T4 polynucleotide kinase. The 5' protruding *Hin*dIII end is labeled with much greater efficiency than the 5' recessed *Kpn*I end, under the reaction conditions used (24, p. 122).

Approximately 10 ng of 5' end-labeled DNA fragment was precipitated with 10 µg of poly(A)$^+$ RNA and 60 µg carrier yeast tRNA. The nucleic acids were resuspended in 80% formamide, 40 mM PIPES, pH 6.4, 0.4 M NaCl and 1 mM EDTA. After heating to 100°C for 2 minutes, the mixture was incubated at 45°C for 5 to 6 hours, then 190 µl of S1 buffer (0.28 M NaCl, 0.05 M NaOAc, pH 4.6, 0.0045 M $ZnSO_4$ and 2 µg/ml denatured salmon sperm DNA) containing 1 unit/µl S1 nuclease was added. The nucleic acids were digested for 1 hour at 15°C, then 37 µl of 0.4 M NH_4OAc, 0.1 M EDTA containing 40 µg tRNA was added to stop the reaction. The protected DNA-RNA fragment was ethanol precipitated, resuspended in sequencing stop buffer (98% formamide, 10 mM EDTA, and 0.3% each of xylene cyanol and bromophenol blue), denatured, and fractionated on 8 M urea/5.3% polyacrylamide DNA sequencing gels along with Sanger dideoxy sequencing reactions containing known DNA sequences as standards.

Sequencing Strategy

DNA fragments of various lengths were obtained by digestion with restriction enzymes having either a 4 base or 6 base recognition sequence (Fig. 1). These

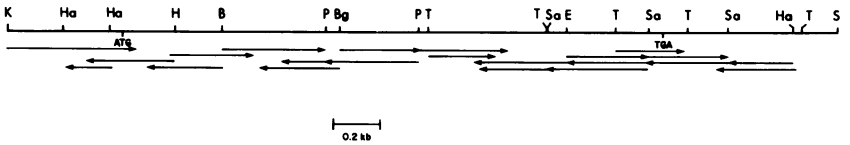

Figure 1. Restriction map of the *Kpn*I-*Sal*I *RAD3* DNA fragment and strategy for sequencing the *RAD3* gene. Symbols for restriction enzymes are as follows: B, *Bam*HI; Bg, *Bgl*II; E, *Eco*RI; H, *Hin*dIII; Ha, *Hae*III; K, *Kpn*I; P, *Pst*I; S, *Sal*I; Sa, *Sau*3A; and T, *Taq*I. The recognition sites for *Hae*III, *Sau*3A and *Taq*I are indicated only in the region where they were used for M13 cloning and DNA sequencing. The ATG initiating codon of the *RAD3* gene, to the left of the *Hin*dIII site, and the TGA termination codon between the second *Sau*3A and *Taq*I sites, are indicated. The *Sau*3A/*Bam*HI fusion junction of *RAD3* DNA with pBR322 in plasmid pSP6 (15) is at the *Sau*3A site to the left of the TGA codon. The horizontal lines with arrowheads indicate the extent and direction of sequencing.

fragments were then cloned into phages M13mp8, M13mp9, M13mp18, and M13mp19, for DNA sequencing by the dideoxy method (25) using deoxyadenosine 5'-(α-[^{35}S]thio)triphosphate as described (26).

RESULTS

Isolation of the 3' Terminus of the *RAD3* Gene

We had previously localized the *rad3* complementing activity in plasmid pSP6 within a DNA fragment extending from the *Kpn*I site at the 5' end to the *Sau*3A site, at the fusion junction of the *RAD3* insert with the plasmid pBR322, at the 3' end (Fig. 1 and ref. 15). The nucleotide sequence of this entire DNA fragment did not reveal any termination codon in the putative *RAD3* open reading frame. A termination codon occurs downstream, following 17 in-frame codons in pBR322. Presumably, the pSP6 encoded RAD3 protein contains 17 amino acids at the carboxyl end, encoded by pBR322 DNA. To obtain the missing portion, we cloned the 3' end of the *RAD3* gene by integration-excision, as illustrated in Fig. 2. Plasmid pSP27, containing the 3.2 kb *Eco*RI *RAD3* region DNA fragment of pSP6 inserted into plasmid YIp5, which transforms yeast only by integration via homologous recombination (27), was used to transform the Rad$^+$ *ura3-52* strain, DBY747, to Ura$^+$. Physical evidence that integration had occurred at the *RAD3* locus was obtained from Southern analysis (28) of genomic DNA digested with appropriate restriction enzymes. Genomic DNA from such an integrant was restricted with *Sal*I, circularized with DNA ligase, and used to transform *E. coli* strain HB101 to ampicillin resistance. The resulting plasmid, pSP29 (Fig. 2) has an addition of a 1.1 kb *Eco*RI-*Sal*I fragment at the 3' end of the 3.2 kb *Eco*RI DNA fragment, and contains about 760 nucleotide pairs beyond the terminal 3' *Sau*3A site in the *RAD3* DNA fragment in plasmid pSP6. A restriction map of the complete *RAD3* gene is

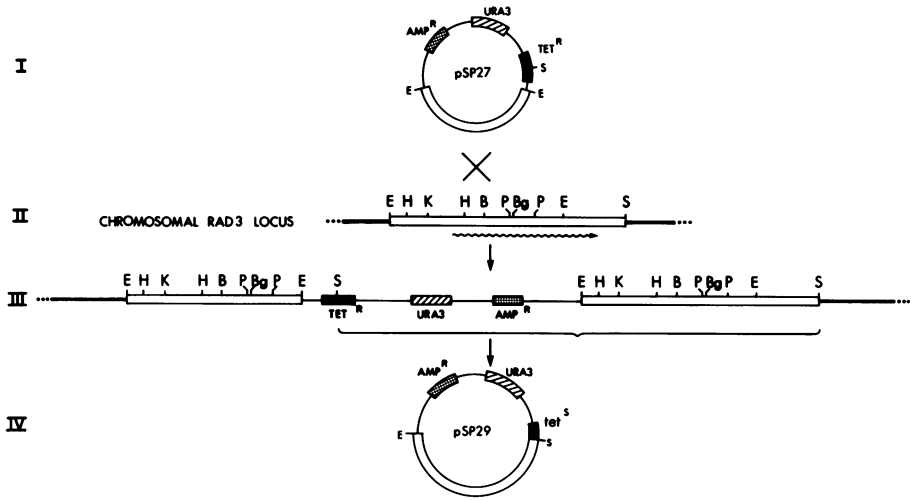

Figure 2. Cloning of the 3' end of the *RAD3* gene. (I) Plasmid pSP27, which is YIp5 containing the *RAD3* region 3.2 kb *Eco*RI fragment. See II for the other restriction sites in this fragment. (II) The *RAD3* region DNA in the yeast chromosome, including the 1.1 kb *Eco*RI-*Sal*I fragment at the 3' end of the *RAD3* gene (16). (III) The *RAD3* region of DNA in a Ura$^+$ yeast strain after integration of plasmid pSP27. (IV) Plasmid pSP29, obtained following transformation of *E. coli* to ampicillin resistance (AMP^R) with a *Sal*I digest of the DNA shown in (III). The thin lines indicate pBR322 sequences; the open bar, *RAD3* region DNA; the solid bar, *E. coli* tetracycline resistance (TET^R) gene; hatched bar, yeast *URA3* gene; dotted bar, *E. coli* AMP^R gene; thick line, yeast chromosomal DNA. Symbols for restriction enzymes are as in Fig. 1. The wavy line with an arrowhead denotes the direction of transcription of the *RAD3* gene.

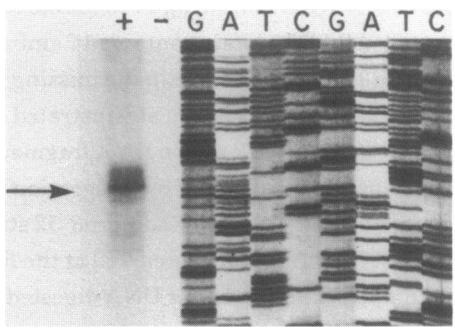

Figure 3. Mapping the 5' terminus of the *RAD3* transcript by S1 nuclease. The 5' terminus of the *RAD3* transcript was mapped by the S1 nuclease protection method using the 720 bp *Kpn*I-*Hind*III 5' end-labeled fragment (Fig. 1). + and - lanes indicate the addition of poly (A)$^+$ RNA or its omission, respectively, in the hybridization to the 5' end-labeled probe prior to treatment with S1 nuclease and fractionation on a DNA sequencing gel. The position of the protected DNA fragment is indicated by the arrow.

```
        -480       -470       -460       -450       -440       -430       -420       -410       -400
          TGGTACCGTCTCTGGGGAAACATCAGATATTATCTTCCGAAATTATTGAAACCTTATAACCAGCAAATTCGGCTAATAGGGTGATTTA

-390       -380       -370       -360       -350       -340       -330       -320       -310       -300
ACGCTACACTAACTGCAAGTATATGTTAACCTTCCCCGAGACTTTGAAAAACCGTGACTCTAGTTGGAAGTCAGCATCTCGTCAACCAAATGATTACA

  -290       -280       -270       -260       -250       -240       -230       -220       -210       -200
GTTCACTTTGAATATTCTGAATAGCCTTTATGATATGAGTTAATCCTATATTAATCATGGCCGACGGCATTTAAGCGATGTATATGAAATTTATGAAA

    -190       -180       -170       -160       -150       -140       -130       -120       -110       -100
AATTAAAGTAGCGTACTTTATGCTACATAGCTTTTGAAAAGAAATCTCTGGTCTGTAAATAATAAATAAATATATATATATGTATATCTTAAAGTTGA

       -90        -80        -70        -60        -50        -40        -30        -20        -10
CAAAGTACTGTTAGCCATTCATAGAAATACTATATTTCATCTTGGGTTGAAGGTGATAATCGGCCCGATTTGACTACACTTTAAGAAGATTGGAAACA

1          10         20         30         40         50         60         70
ATG AAG TTT TAT ATA GAT GAT TTA CCA GTG CTT TTT CCA TAC CCC AAG ATA TAT CCA GAG CAG TAT AAT TAT
Met Lys Phe Tyr Ile Asp Asp Leu Pro Val Leu Phe Pro Tyr Pro Lys Ile Tyr Pro Glu Gln Tyr Asn Tyr

         80         90        100        110        120        130        140
ATG TGC GAT ATT AAA AAG ACT CTG GAT GTA GGT GGA AAT AGT ATC TTG GAG ATG CCT TCA GGA ACA GGT AAA
Met Cys Asp Ile Lys Lys Thr Leu Asp Val Gly Gly Asn Ser Ile Leu Glu Met Pro Ser Gly Thr Gly Lys

     150        160        170        180        190        200        210
ACG GTC TCA TTA CTA TCC CTC ACA ATT GCC TAC CAG ATG CAC TAC CCA GAA CAT AGA AAG ATC ATA TAT TGT
Thr Val Ser Leu Leu Ser Leu Thr Ile Ala Tyr Gln Met His Tyr Pro Glu His Arg Lys Ile Ile Tyr Cys

   220        230        240        250        260        270        280
TCT CGT ACT ATG TCT GAA ATT GAA AAA GCT TTA GTA GAG TTA GAG AAC CTT ATG GAT TAC AGA ACT AAA GAA
Ser Arg Thr Met Ser Glu Ile Glu Lys Ala Leu Val Glu Leu Glu Asn Leu Met Asp Tyr Arg Thr Lys Glu

290        300        310        320        330        340        350        360
CTA GGC TAT CAA GAG GAT TTT CGA GGT CTT GGC TTG ACA TCA AGA AAA AAT TTG TGT TTG CAT CCC GAA GTG
Leu Gly Tyr Gln Glu Asp Phe Arg Gly Leu Gly Leu Thr Ser Arg Lys Asn Leu Cys Leu His Pro Glu Val

         370        380        390        400        410        420        430
AGT AAG GAA CGA AAA GGT ACA GTA GTC GAT GAA AAG TGC CGT AGA ATG ACA AAT GGG CAG GCG AAG AGA AAA
Ser Lys Glu Arg Lys Gly Thr Val Val Asp Glu Lys Cys Arg Arg Met Thr Asn Gly Gln Ala Lys Arg Lys

       440        450        460        470        480        490        500
TTA GAA GAG GAT CCA GAG GCA AAT GTA GAA TTG TGT GAA TAC CAT GAG AAT TTG TAC AAT ATT GAA GTA GAG
Leu Glu Glu Asp Pro Glu Ala Asn Val Glu Leu Cys Glu Tyr His Glu Asn Leu Tyr Asn Ile Glu Val Glu

     510        520        530        540        550        560        570
GAT TAT CTT CCA AAA GGC GTA TTT TCT TTT GAA AAA CTT TTG AAA TAC TGC GAA GAA AAA ACA CTT TGT CCA
Asp Tyr Leu Pro Lys Gly Val Phe Ser Phe Glu Lys Leu Leu Lys Tyr Cys Glu Glu Lys Thr Leu Cys Pro

   580        590        600        610        620        630        640
TAT TTT ATT GTT CGT CGT ATG ATT TCT CTT TGT AAC ATT ATT ATT TAT TCT TAC CAT TAT CTA TTA GAT CCT
Tyr Phe Ile Val Arg Arg Met Ile Ser Leu Cys Asn Ile Ile Ile Tyr Ser Tyr His Tyr Leu Leu Asp Pro

650        660        670        680        690        700        710        720
AAA ATT GCT GAA AGA GTT TCC AAC GAG GTT TCT AAA GAT AGC ATT GTC ATT TTT GAT GAA GCG CAC AAT ATT
Lys Ile Ala Glu Arg Val Ser Asn Glu Val Ser Lys Asp Ser Ile Val Ile Phe Asp Glu Ala His Asn Ile

         730        740        750        760        770        780        790
GAT AAT GTG TGT ATC GAA TCG CTG TCA TTA GAC TTG ACA ACG GAT GCA TTG AGA AGA GCC ACA CGA GGT GCT
Asp Asn Val Cys Ile Glu Ser Leu Ser Leu Asp Leu Thr Thr Asp Ala Leu Arg Arg Ala Thr Arg Gly Ala

       800        810        820        830        840        850        860
AAT GCG TTA GAT GAA CGT ATT TCT GAG GTC AGA AAG GTT GAC TCA CAG AAA TTA CAG GAT GAA TAC GAA AAA
Asn Ala Leu Asp Glu Arg Ile Ser Glu Val Arg Lys Val Asp Ser Gln Lys Leu Gln Asp Glu Tyr Glu Lys

     870        880        890        900        910        920        930
CTA GTT CAA GGT CTC CAT TCT GCA GAT ATT CTT ACC GAC CAG GAA GAG CCA TTT GTG GAA ACA CCT GTA TTG
Leu Val Gln Gly Leu His Ser Ala Asp Ile Leu Thr Asp Gln Glu Glu Pro Phe Val Glu Thr Pro Val Leu

   940        950        960        970        980        990       1000
CCC CAA GAT CTT CTA ACA GAA GCA ATC CCG GGA AAT ATA CGA AGA GCC GAG CAT TTT GTT TCA TTT TTG AAA
Pro Gln Asp Leu Leu Thr Glu Ala Ile Pro Gly Asn Ile Arg Arg Ala Glu His Phe Val Ser Phe Leu Lys
```

```
1010        1020        1030        1040        1050        1060        1070        1080
AGA TTG ATA GAA TAT CTG AAG ACC AGA ATG AAA GTT CTT CAC GTT ATT TCA GAA ACG CCA AAA TCA TTT CTA
Arg Leu Ile Glu Tyr Leu Lys Thr Arg Met Lys Val Leu His Val Ile Ser Glu Thr Pro Lys Ser Phe Leu

        1090        1100        1110        1120        1130        1140        1150
CAG CAT TTA AAA CAG TTA ACT TTC ATA GAA AGG AAA CCT CTT CGG TTT TGC TCA GAA AGG CTA TCA TTA CTT
Gln His Leu Lys Gln Leu Thr Phe Ile Glu Arg Lys Pro Leu Arg Phe Cys Ser Glu Arg Leu Ser Leu Leu

      1160        1170        1180        1190        1200        1210        1220
GTA AGA ACT TTA GAA GTT ACA GAG GTA GAA GAT TTT ACT GCA TTG AAA GAC ATA GCA ACT TTT GCT ACT CTT
Val Arg Thr Leu Glu Val Thr Glu Val Glu Asp Phe Thr Ala Leu Lys Asp Ile Ala Thr Phe Ala Thr Leu

    1230        1240        1250        1260        1270        1280        1290
ATA TCA ACA TAT GAG GAA GGG TTT TTA CTA ATT ATT GAA CCG TAT GAA ATC GAA AAT GCT GCA GTT CCG AAT
Ile Ser Thr Tyr Glu Glu Gly Phe Leu Leu Ile Ile Glu Pro Tyr Glu Ile Glu Asn Ala Ala Val Pro Asn

  1300        1310        1320        1330        1340        1350        1360
CCG ATT ATG AGA TTT ACT TGC TTA GAT GCA TCG ATT GCC ATC AAA CCA GTC TTT GAG AGA TTT TCT TCC GTT
Pro Ile Met Arg Phe Thr Cys Leu Asp Ala Ser Ile Ala Ile Lys Pro Val Phe Glu Arg Phe Ser Ser Val

1370        1380        1390        1400        1410        1420        1430        1440
ATT ATC ACT TCA GGG ACC ATA TCA CCA TTA GAC ATG TAT CCA AGA ATG TTA AAC TTT AAA ACT GTT TTA CAA
Ile Ile Thr Ser Gly Thr Ile Ser Pro Leu Asp Met Tyr Pro Arg Met Leu Asn Phe Lys Thr Val Leu Gln

        1450        1460        1470        1480        1490        1500        1510
AAA TCA TAT GCC ATG ACC TTA GCC AAA AAA TCA TTT CTA CCA ATG ATT ATT ACC AAG GGT TCT GAT CAA GTT
Lys Ser Tyr Ala Met Thr Leu Ala Lys Lys Ser Phe Leu Pro Met Ile Ile Thr Lys Gly Ser Asp Gln Val

      1520        1530        1540        1550        1560        1570        1580
GCA ATT TCT TCA AGA TTT GAA ATC AGA AAC GAT CCT AGT ATT GTT CGT AAT TAC GGT TCT ATG CTA GTA GAG
Ala Ile Ser Ser Arg Phe Glu Ile Arg Asn Asp Pro Ser Ile Val Arg Asn Tyr Gly Ser Met Leu Val Glu

    1590        1600        1610        1620        1630        1640        1650
TTT GCC AAG ATC ACA CCT GAT GGA ATG GTT GTT TTT TTC CCC TCA TAT CTA TAT ATG GAA AGT ATT GTT TCA
Phe Ala Lys Ile Thr Pro Asp Gly Met Val Val Phe Phe Pro Ser Tyr Leu Tyr Met Glu Ser Ile Val Ser

  1660        1670        1680        1690        1700        1710        1720
ATG TGG CAA ACA ATG GGT ATT CTT GAC GAA GTT TGG AAA CAT AAA TTA ATT TTA GTT GAG ACT CCT GAT GCT
Met Trp Gln Thr Met Gly Ile Leu Asp Glu Val Trp Lys His Lys Leu Ile Leu Val Glu Thr Pro Asp Ala

1730        1740        1750        1760        1770        1780        1790        1800
CAA GAA ACT TCT TTA GCC TTA GAA ACC TAT AGA AAG GCT TGC TCA AAT GGG CGT GGG GCA ATT TTG CTT TCT
Gln Glu Thr Ser Leu Ala Leu Glu Thr Tyr Arg Lys Ala Cys Ser Asn Gly Arg Gly Ala Ile Leu Leu Ser

        1810        1820        1830        1840        1850        1860        1870
GTT GCT AGA GGA AAG GTA TCT GAA GGT ATC GAT TTT GAT CAT CAA TAT GGC AGA ACT GTG CTG ATG ATA GGT
Val Ala Arg Gly Lys Val Ser Glu Gly Ile Asp Phe Asp His Gln Tyr Gly Arg Thr Val Leu Met Ile Gly

      1880        1890        1900        1910        1920        1930        1940
ATC CCG TTT CAA TAC ACA GAA TCG CGT ATT TTG AAA GCT CGC CTA GAA TTC ATG AGG GAG AAC TAT CGC ATC
Ile Pro Phe Gln Tyr Thr Glu Ser Arg Ile Leu Lys Ala Arg Leu Glu Phe Met Arg Glu Asn Tyr Arg Ile

    1950        1960        1970        1980        1990        2000        2010
AGA GAA AAC GAC TTC TTA TCT TTC GAT GCG ATG AGA CAT GCA GCT CAA TGT CTG GGG AGA GTA CTG AGA GGG
Arg Glu Asn Asp Phe Leu Ser Phe Asp Ala Met Arg His Ala Ala Gln Cys Leu Gly Arg Val Leu Arg Gly

  2020        2030        2040        2050        2060        2070        2080
AAG GAC GAC TAT GGT GTA ATG GTA CTA GCA GAC CGT AGG TTT TCA AGA AAA AGA AGC CAG TTA CCA AAA TGG
Lys Asp Asp Tyr Gly Val Met Val Leu Ala Asp Arg Arg Phe Ser Arg Lys Arg Ser Gln Leu Pro Lys Trp

2090        2100        2110        2120        2130        2140        2150        2160
ATT GCT CAA GGT TTG TCT GAC GCC GAT TTG AAC CTT TCG ACT GAC ATG GCC ATA TCC AAT ACC AAA CAA TTC
Ile Ala Gln Gly Leu Ser Asp Ala Asp Leu Asn Leu Ser Thr Asp Met Ala Ile Ser Asn Thr Lys Gln Phe

        2170        2180        2190        2200        2210        2220        2230
TTG AGA ACA ATG GCA CAA CCC ACA GAC CCT AAA GAC CAA GAG GGT GTA TCT GTT TGG AGT TAT GAA GAT TTA
Leu Arg Thr Met Ala Gln Pro Thr Asp Pro Lys Asp Gln Glu Gly Val Ser Val Trp Ser Tyr Glu Asp Leu

      2240        2250        2260        2270        2280        2290        2300
ATA AAG CAC CAG AAT AGC AGA AAA GAT CAA GGT GGA TTT ATT GAA AAC GAA AAC AAA GAA GGA GAA CAG GAT
Ile Lys His Gln Asn Ser Arg Lys Asp Gln Gly Gly Phe Ile Glu Asn Glu Asn Lys Glu Gly Glu Gln Asp
```

```
    2310          2320          2330          2340      2350      2360      2370      2380
GAA GAT GAA GAT GAA GAT ATA GAA ATG CAG TGA TGCAATGATACGCTTTTGCTATAAACTGTATATATCACAATTAGATTAAAT
Glu Asp Glu Asp Glu Asp Ile Glu Met Gln ---

2390      2400      2410      2420      2430      2440      2450      2460      2470      2480
AGCCGCAAGAGAATGTTATATATTGAAATCCATTCGATTATCCAGGACTAAACAATGATTTTATTTCACATTTATTTCAAAGGACAACTCTTT

      2490      2500      2510      2520      2530      2540      2550      2560      2570
ATCTGCGTCAAGATATGAATCACAGACACACCAAATTGTTAAGTTATGTTTACCAGATGTCGGAGTGTCAAATTCCAATTCATATTGCTGAGT

   2580      2590      2600      2610      2620      2630      2640      2650      2660
CTCTTTATTTAAAGTAACTTTCTTGATCGCATAAAGCTCTTTTTTAGATACTTCACCTAAAACCAACCACCAACTTTCTAGCTTATCAAACGG

2670      2680      2690      2700      2710      2720      2730      2740      2750      2760
ATACTTTTCAGATGTTACTTGTAAGTTCTCAGGTTCAACATCTCTTGTCAATTGAATGGTAATTTTCTGTTTTACACCTGAAATTAACGAATC

      2770      2780      2790      2800      2810      2820      2830      2840      2850
AGAATTATTCAACGAATAAGTAAGTTCAACGTTGGGGTAATTGTTAACAAACGCGGCGACCTGCGCAAGCTGTGAATCAGTAAGCGTTAAAAT

   2860      2870      2880      2890      2900
TTCATCCCTCTCTTCATCCTCAAGGGCCATTATATCATAGACCGTCTCGA
```

Figure 4. Nucleotide sequence of the *RAD3* gene. The sequence of 3383 nucleotides in the DNA strand identical to the mRNA strand is presented. The numbering is in relation to the first base of the ATG translation initiation codon, indicated as +1. The second digit from the right denotes the numbered base. The predicted amino acid sequence encoded by the *RAD3* open reading frame is given below the nucleotide sequence. The location of the 5' mRNA terminus is denoted by an arrow. Possible "TATA" sequences are underlined in the 5' upstream nontranslated region. Restriction sites for some enzymes noted in Fig. 1 are as follows: *Kpn*I, -479 to -474; *Hin*dIII, +242 to +247; *Bam*HI, +441 to +446; *Pst*I at +884 to +889 and at +1283 to +1288; *Bgl*II, +942 to +947; *Eco*RI, +1918 to +1923; and the *Sau*3A site at +2257 to +2260, the site of the fusion of *RAD3* region DNA with the pBR322 sequence in plasmid pSP6.

given in Fig. 1.

5' End Mapping of *RAD3* mRNA

The 5' end of the *RAD3* mRNA was mapped by S1 nuclease digestion of hybrids between mRNA and the 5' end-labeled 720 bp *Kpn*I-*Hin*dIII DNA fragment from the 5' end of the gene (Fig. 1), and sizing of the protected DNA fragment on an 8 M urea/5.3% polyacrylamide DNA sequencing gel. A protected fragment of 364 nucleotides is observed (Fig. 3), indicating a transcriptional start located at 117 nucleotides upstream of the first ATG codon in the RAD3 open reading frame (Fig. 4).

Sequence of the *RAD3* Gene and Protein

The strategy employed for determination of the nucleotide sequence is given in Fig. 1. The nucleotide sequence of the DNA strand of the *RAD3* gene identical to the mRNA strand is shown in Fig. 4. There is only one long open reading frame within the *rad3* complementing DNA fragment, which starts with the ATG at +1 and continues for 2334 nucleotides until the termination codon TGA. Experiments with initiation mutants of the yeast *CYC1* gene (29) and mutational studies in the yeast *HIS4* gene (30) indicate no particular requirement for a specific sequence 5' to the initiating ATG codon and that initiation of translation occurs at the first ATG codon

Table 1. Codon usage and amino acid composition in RAD3
(Percent occurrence of codons and amino acids is given in parentheses)

Phe	UUU	27	(3.5)	Ser	UCU	18	(2.3)	Tyr	UAU	22	(2.8)	Cys	UGU	7	(0.9)
Phe	UUC	6	(0.8)	Ser	UCC	4	(0.5)	Tyr	UAC	11	(1.4)	Cys	UGC	6	(0.8)
Leu	UUA	26	(3.3)	Ser	UCA	20	(2.6)		UAA	-			UGA	1	
Leu	UUG	18	(2.3)	Ser	UCG	4	(0.5)		UAG	-		Trp	UGG	4	(0.5)
Leu	CUU	16	(2.1)	Pro	CCU	8	(1.0)	His	CAU	10	(1.3)	Arg	CGU	9	(1.2)
Leu	CUC	2	(0.3)	Pro	CCC	5	(0.6)	His	CAC	4	(0.5)	Arg	CGC	2	(0.3)
Leu	CUA	13	(1.7)	Pro	CCA	14	(1.8)	Gln	CAA	15	(1.9)	Arg	CGA	4	(0.5)
Leu	CUG	6	(0.8)	Pro	CCG	5	(0.6)	Gln	CAG	12	(1.5)	Arg	CGG	1	(0.1)
Ile	AUU	32	(4.1)	Thr	ACU	15	(1.9)	Asn	AAU	17	(2.2)	Ser	AGU	5	(0.6)
Ile	AUC	12	(1.5)	Thr	ACC	7	(0.9)	Asn	AAC	10	(1.3)	Ser	AGC	3	(0.4)
Ile	AUA	13	(1.7)	Thr	ACA	17	(2.2)	Lys	AAA	35	(4.5)	Arg	AGA	29	(3.7)
Met	AUG	26	(3.3)	Thr	ACG	3	(0.4)	Lys	AAG	15	(1.9)	Arg	AGG	4	(0.5)
Val	GUU	21	(2.7)	Ala	GCU	11	(1.4)	Asp	GAU	34	(4.4)	Gly	GGU	15	(1.9)
Val	GUC	5	(0.6)	Ala	GCC	10	(1.3)	Asp	GAC	14	(1.8)	Gly	GGC	4	(0.5)
Val	GUA	15	(1.9)	Ala	GCA	13	(1.7)	Glu	GAA	51	(6.6)	Gly	GGA	7	(0.9)
Val	GUG	5	(0.6)	Ala	GCG	4	(0.5)	Glu	GAG	20	(2.6)	Gly	GGG	7	(0.9)

Ala	38	(4.9)	Leu	81	(10.4)
Arg	49	(6.3)	Lys	50	(6.4)
Asn	27	(3.5)	Met	26	(3.3)
Asp	48	(6.2)	Phe	33	(4.2)
Cys	13	(1.7)	Pro	32	(4.1)
Gln	27	(3.5)	Ser	54	(6.9)
Glu	71	(9.1)	Thr	42	(5.4)
Gly	33	(4.2)	Trp	4	(0.5)
His	14	(1.8)	Tyr	33	(4.2)
Ile	57	(7.3)	Val	46	(5.9)

Mr = 89,779

closest to the 5' terminus of the mRNA. Therefore, the initiating ATG codon of the *RAD3* gene is most likely to be the one as indicated in Fig. 4. The size of this open reading frame is consistent with the size of the 2.5 kb *RAD3* mRNA (15).

Evidence for the correct identification of the *RAD3* open reading frame comes from *RAD3-lacZ* fusions. The *E. coli lacZ* gene, missing its promoter and the first 7 amino acids, was fused with the *RAD3* gene in the *Bam*HI site at +441 to +446 and also at the *Bgl*II site at +942 to +947. These fusions connect the *E. coli lacZ* gene in the same reading frame as the *RAD3* open reading frame and produce β-galactosidase in yeast (31), indicating that the *RAD3* open reading frame is translated. The predicted RAD3 protein contains 778 amino acids with a molecular weight of 89,779.

Codon Usage in *RAD3*

The amino acid sequence encoded by the *RAD3* open reading frame contains 40.7% nonpolar, 29.4% polar, 15.3% acidic and 14.5% basic amino acids, which

represents a random distribution of the four classes of amino acids. The base composition of the open reading frame is 33.6% A, 29.3% T, 20.4% G, and 16.7% C, and that of the wobble position is 35.0% A, 34.3% T, 17.2% G, and 13.5% C. Codon usage, and the amino acid composition are given in Table 1. In *RAD3* all of the 61 codons are used. This is in striking contrast to the highly expressed genes of *S. cerevisiae* such as alcohol dehydrogenase I and glyceraldehyde 3-phosphate dehydrogenase, which show extreme codon bias and in which 96% of the amino acid residues are encoded by only a select 25 of the 61 possible coding triplets, and these preferred codons correspond to the anticodons of the major isoacceptor tRNA species of yeast (32). In *RAD3*, the UUC phenylalanine, UUG leucine, UAC tyrosine, AAG lysine, and GAC aspartic acid codons, which correspond to the major tRNA isoacceptor species in yeast, are not used as often as the other codons for these amino acids; in contrast, the GAA glutamic acid and AGA arginine codons, which correspond to the major isoacceptor tRNA species of yeast, are used quite frequently. Overall, codon usage in the *RAD3* gene would indicate this gene not to be a highly expressed one.

5' and 3' Flanking Sequences of the *RAD3* Gene

The 480 nucleotides 5' to the *RAD3* initiating codon show an overall base composition of 34.4% A, 32.1% T, 17.1% G, and 16.5% C, similar to that in the open reading frame. In higher eukaryotes, the "TATA" box with the consensus sequence

$$5'\text{-TATA}\begin{matrix}A\ A\\ T\ T\end{matrix}\text{-}3'$$

is usually found 26 to 34 bp upstream of the site of initiation of transcription (33,34) and is apparently required for proper positioning of the mRNA start at a specific site by RNA polymerase II (35-38). In many yeast genes examined thus far, the distance between the mRNA start site and the TATA-like sequence is not as rigid as it is in higher eukaryotes, but varies considerably, being about 100 bp and 150 bp in the *PYK1* gene (39), 100 bp in the *ADH1* gene (32), and 39 bp in the *HIS4* gene (30). The *RAD3* gene has a TATATTA sequence at -248, and a TATAT sequence at -214, placing their distance from the mRNA start site at 125 and 93 bases, respectively. The sequence TAAATAA is found in tandem repeat at locations -141 and -134, and is followed by the sequence AT repeated six times. A number of yeast genes that encode abundant mRNAs seem to have their transcription start sites at or very near the sequence CAAG and a CT rich block is usually found upstream of the CAAG sequence (32,39,40). In the *RAD3* gene, there is no CT rich block nor a CAAG sequence in the vicinity of the mRNA start site.

It has been suggested that efficient initiation of translation in eukaryotes depends on a purine residue, usually an A, at position -3 relative to the initiating ATG codon (40,41). The -3 position of the *RAD3* gene is occupied by an A residue.

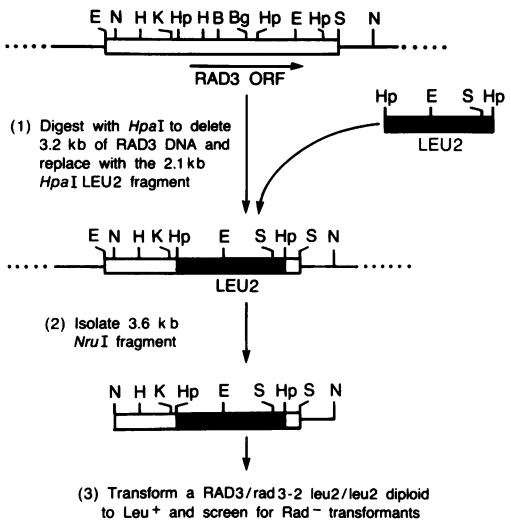

Figure 5. Construction of a *rad3* deletion strain. (1) The 3.2 kb *Hpa*I fragment of *RAD3* (from nucleotide -365 to +2805) which contains the entire *RAD3* open reading frame (ORF) was deleted from plasmid pSP29. The deletion extends 365 nucleotides 5' to the first ATG codon and 468 nucleotides 3' to the TGA (stop) codon of *RAD3*. The 2.1 kb *Hpa*I fragment containing the *LEU2* gene from plasmid YEp13 was ligated into the *RAD3* deleted plasmid pSP29, resulting in plasmid pSP36. (2) The 3.6 kb *Nru*I fragment deleted for the *RAD3* gene and containing the *LEU2* gene insert was purified from plasmid pSP36 by electroelution. (3) The 3.6 kb *Nru*I fragment was used to transform the diploid strain DH-225 (*MAT*a/*MAT*α *ARO7*/*aro7 CAN1*/*can1 his3*-Δ1/*HIS3 leu2*-3 *leu2*-112/*leu2*-3 *leu2*-112 *trp1*-289/*TRP1 ura3*-52/*ura3*-52 *RAD3*/*rad3*-2) to Leu$^+$ and transformants screened for Rad$^-$. The Rad$^-$ transformants arise from replacement of the chromosomal *RAD3* gene by the *LEU2* gene. Symbols for restriction enzymes are as follows: B, *BamH*I; Bg, *Bgl*II; E, *Eco*RI; H, *Hind*III; Hp, *Hpa*I; K, *Kpn*I; N, *Nru*I; S, *Sal*I.

Downstream of the ATG, a purine at position +4 (41) has been implicated as playing a role in the efficiency of initiation of translation, while Dobson et al. (40) suggest that efficient initiation in yeast depends on a pyrimidine, usually T, at position +6. The *RAD3* gene has an A at +4, but a G at +6.

The 566 nucleotides downstream of the TGA termination codon contain 32.0% A, 33.6% T, 14.8% G, and 19.6% C. In the 3' noncoding region of the mRNAs of higher eukaryotic genes, the sequence AATAAA is present about 20 nucleotides upstream from the 3' end of the mRNA (37,42), and has been postulated to be necessary for polyadenylation (43). Many of the sequenced yeast genes contain a sequence related to the model sequence TAAATAA$\begin{matrix}A\\G\end{matrix}$ approximately 28 to 33 nucleotides upstream from the 3' mRNA terminus (32). Zaret and Sherman (44)

have identified another consensus sequence, TAG.....TAGT or TATGT .. (A-T rich) ..TTT... in the 3' mRNA region of various yeast genes, whereas Henikoff et al. (45) have suggested that the sequence TTTTTATA is required for transcription termination in yeast. No sequences similar to the above mentioned ones exist in the 3' region of the *RAD3* gene. However, various AT-rich stretches exist in the 3' noncoding region.

In vivo function of the acidic carboxyl terminus of RAD3 protein

The carboxyl terminal region of the RAD3 protein is highly acidic; the sequence of the last 20 amino acids has 12 acidic and only 1 basic residue, and 7 of the acidic residues, from residue 768 through 774, are present in tandem (Fig. 4). The *RAD3* DNA fragment in plasmid pSP6, which fully complements the UV sensitive *rad3-2* mutation (15), is missing the last 25 amino acid codons of the *RAD3* gene, and in the RAD3 protein coded by the plasmid pSP6, these residues are presumably replaced by the pBR322 encoded sequence of 17 amino acids: Pro-Gln-Asp-Gly-Cys-Gly-Arg-His-Asp-Arg-Val-Val-Asp-Ser-Gly-Ser-Lys. The last 25 RAD3 encoded amino acids contain 12 acidic and 1 basic residue including 7 consecutive acidic residues (Fig. 4), while the last 17 amino acid residues encoded by the pBR322 sequence in plasmid pSP6 have only 3 acidic and 4 basic residues. However, these observations give no indication about the function of the acidic carboxyl terminal region in the RAD3 protein. *RAD3*, in addition to its function in DNA repair, is also required for viability (15,17). Since *rad3*-2 mutants are viable but defective in excision repair, the rad3-2 mutant protein is defective in the DNA repair but not in the viability function. Complementation of the *rad3-2* mutation by the *RAD3* DNA fragment in plasmid pSP6 only indicates that the RAD3 protein missing the last 25 amino acids can supply the DNA repair activity that is absent in the rad3-2 protein.

For determining the function of the acidic carboxyl terminus of RAD3, we used a yeast strain in which the entire genomic *RAD3* gene was deleted by the gene replacement method (46), as shown in Fig. 5. The 3.2 kb *RAD3* DNA that contains the entire amino acid coding region, plus 365 nucleotides upstream and 468 nucleotides downstream of the coding region, was replaced with the *LEU2* gene. The DNA fragment containing the *LEU2* gene and the flanking *RAD3* region sequences was used for transformation of a *RAD3/rad3-2 leu2/leu2* diploid strain to Leu$^+$ and transformants screened for UV sensitivity (Rad$^-$). The Leu$^+$ Rad$^-$ transformants arise from replacement of the *RAD3* gene by the *LEU2* DNA fragment by gene conversion, and thus are deleted for the entire chromosomal *RAD3* gene. The *LEU2 rad3-Δ/rad3-2 leu2/leu2* diploids were sporulated and, as expected for the recessive lethal *rad3-Δ* (15), genetic analysis of 200 random spores and 15 tetrads gave only

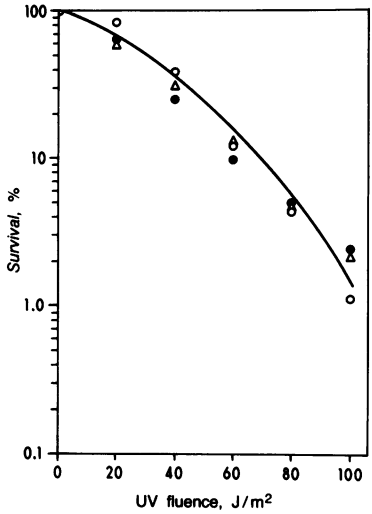

Figure 6. Survival after UV irradiation of a *rad3*-Δ strain harboring various *RAD3* insert-containing plasmids. Cells were grown in minimal medium supplemented with the necessary nutrients and lacking uracil. ○, Strain DH225-2A, *rad3*-Δ + pSP6; △, *rad3*-Δ + pSP32; ●, LP2649-1B, Rad$^+$ + YEp24.

Leu$^-$ spores. To examine if the RAD3 acidic carboxyl terminus affects cell viability, we transformed the *LEU2 rad3*-Δ/*rad3-2 leu2/leu2* diploid with either the multicopy autonomously replicating plasmid pSP6, in which the acidic carboxyl terminus is deleted, or the plasmid pSP32, which contains the entire *RAD3* gene. Plasmid pSP32 was constructed by cloning the 4.0 kb *Sal*I fragment containing the 2 μ circle sequences from plasmid YEp13 into plasmid pSP29 (Fig. 2). The transformed diploids were sporulated. In both cases, genetic analysis of 200 random spores and 20 tetrads yielded both Leu$^+$ and Leu$^-$ spores, indicating that, like plasmid pSP32, plasmid pSP6 can rescue *rad3*-Δ spores. The growth rates of the haploid *rad3*-Δ strain bearing either pSP6 or pSP32 were similar to that of the wild type (Rad$^+$) haploid strain. The UV survival of the haploid *rad3*-Δ strain containing either plasmid pSP6 or plasmid pSP32 was similar to the Rad$^+$ haploid strain (Fig. 6).

DISCUSSION

The *RAD3* 5' mRNA terminus maps at position -117, with putative "TATA" sequences present 93 and 125 nucleotides upstream of the mRNA start site. The *RAD3* gene contains 778 codons in the open reading frame. The 89,779 dalton protein encoded by the *RAD3* gene contains 15.3% acidic and 14.5% basic amino acids. The carboxyl terminal region of the RAD3 protein is highly acidic, containing

Protein	Residues	Sequence
RAD3	40-61	L E M P S G T G K T V S L L S L T I A Y Q M
UvrD	27-47	V L A G A G S G K T R V L V H R I A W L M

Figure 7. Sequence of the yeast RAD3 protein in the region showing homology with the *E. coli* UvrD protein. The underlined amino acids in the UvrD protein correspond to the highly conserved consensus sequence found in various adenine nucleotide binding proteins. The boxed amino acids are identical in the RAD3 and UvrD proteins. The two proteins have been aligned for optimal homology.

12 acidic and only 1 basic residue in the last 20 amino acids. The Kyte and Doolittle plot (47) of hydrophilic and hydrophobic regions in the RAD3 protein indicates a highly hydrophilic carboxyl terminal end. The amino terminal region contains both hydrophilic and hydrophobic sections, while the middle region is hydrophobic.

We have compared the amino acid sequence of the RAD3 protein with the *E. coli* UvrC and UvrD protein sequences (48,49). The UvrC protein, along with the UvrA and UvrB proteins, functions in the incision step of excision repair. No significant homology was evident between the RAD3 and the UvrC proteins. The UvrD protein is involved in excision repair (50-52), recombination (53), methyl directed DNA mismatch repair (54), and possibly DNA replication (55), and it possesses a single stranded DNA dependent ATPase and ATP dependent DNA unwinding activity (56). The amino acid sequence of the RAD3 protein from residue 40 to 61 and of the UvrD protein from residue 27 to 47 resemble each other in 8 residues (Fig. 7). This region of the UvrD protein contains the conserved consensus sequence I/V/L-X-A/G-X-X-X-X-G-K-T-X-X-X-X-X-X-I/V, that has been identified in a number of adenine nucleotide binding proteins - RecA, DnaB, Rho, thymidine kinase, ATPase β, ATPase α, AMP kinase, and myosin (49,57). The RAD3 protein has a similar sequence in this region which differs from the consensus sequence in having methionine in the third position rather than alanine or glycine, and the last conserved amino acid isoleucine is at a position following 7 rather than 6 residues after the conserved threonine residue at position 10. The RAD3 and UvrD proteins also contain 4 identical amino acid residues not included in the conserved consensus sequence. Fusions of the *RAD3* gene with a strong yeast promoter should facilitate the purification of the RAD3 protein which could then be characterized for various activities including DNA dependent ATPase and ATP dependent helicase.

To examine the *in vivo* function of the acidic carboxyl terminus of RAD3, we constructed a diploid strain in which the entire RAD3 amino acid coding region and the 5' and 3' noncoding sequences were deleted from one of the chromosomes, while the other homologue carried the *rad3-2* mutation. Sporulation of this diploid gave only *rad3-2* spores, since *rad3-Δ* spores are inviable. In plasmid pSP6, the last 25

amino acid codons from the RAD3 carboxyl terminus are replaced by 17 amino acid codons of pBR322, thereby changing this from a highly acidic to a slightly basic region. We observed that plasmid pSP6 could rescue the *rad3*-Δ spores, indicating that the highly acidic carboxyl terminus is not essential for viability. Moreover, the growth rate and UV survival of *rad3*-Δ strains harboring plasmid pSP6 were similar to the Rad$^+$ strains. Even though our observations do not demonstrate an essential function for the acidic carboxyl terminus, we cannot conclude that this region has no role. It could be that the function of the missing RAD3 acidic region is carried out by other protein(s) which also contain similar acidic regions. In that case, deletion of the acidic region from any one protein may have no effect, but deletions from more than one protein could produce adverse effects.

The *RAD6* gene of *S. cerevisiae*, which is required for DNA repair, induced mutagenesis, and sporulation, and which belongs to a different epistasis group than the *RAD3* gene, also encodes a protein with a very highly acidic carboxyl terminal region. The last 23 amino acids of the 172 residue long RAD6 protein contain 20 acidic amino acids, including 13 consecutive aspartate residues (58). The chromatin-associated high mobility group proteins HMG-1 and HMG-2 from calf thymus also contain a long block of acidic residues in the carboxyl region (59). The acidic carboxyl terminal regions of the RAD3 and RAD6 proteins may be involved in binding to histones or to basic regions of other interacting proteins.

ACKNOWLEDGEMENTS

We thank Alan Morrison for helpful discussions and Eugene LeClerc, Nancy Istock, and David Barker for advice on DNA sequencing. This work was supported by grants CA35035 to S.P. and GM19261 to L.P. from NIH, and contract No. DE-AC02-76EV03490 with the U.S. Department of Energy at the University of Rochester Department of Radiation Biology and Biophysics and has been assigned number DOE/EV/03490-2472.

REFERENCES

1. Unrau, P., Wheatcroft, R. and Cox, B.S. (1971) Mol. Gen. Genet. 113, 359-362.
2. Resnick, M.A. and Setlow, J.K. (1972) J. Bacteriol. 109, 979-986.
3. Waters, R. and Moustacchi, E. (1974) Biochim. Biophys. Acta. 353, 407-419.
4. Prakash, L. (1975) J. Mol. Biol. 98, 781-795.
5. Prakash, L. (1977) Mutat. Res. 45, 13-20.
6. Prakash, L. (1977) Mol. Gen. Genet. 152, 125-128.
7. Prakash, L. and Prakash, S. (1979) Mol. Gen. Genet. 176, 351-359.
8. Miller, R.D., Prakash, L. and Prakash, S. (1982) Mol. Cell Biol. 2, 939-948.
9. Miller, R.D., Prakash, L. and Prakash, S. (1982) Mol. Gen. Genet. 188, 235-239.
10. Reynolds, R.J. (1978) Mutat. Res. 50, 43-56.
11. Jachymczyk, W.J., von Borstel, R.C., Mowat, M.R.A. and Hastings, P.J. (1981) Mol. Gen. Genet. 182, 196-205.

12. Magaña-Schwencke, N., Henriques, J.A.P., Chanet, R. and Moustacchi, E. (1982) Proc. Natl. Acad. Sci. USA 79, 1722-1726.
13. Wilcox, D.R. and Prakash, L. (1981) J. Bacteriol. 148, 618-623.
14. Reynolds, R.J. and Friedberg, E.C. (1981) J. Bacteriol. 146, 692-704.
15. Higgins, D.R., Prakash, S., Reynolds, P., Polakowska, R., Weber, S. and Prakash, L. (1983) Proc. Natl. Acad. Sci. USA 80, 5680-5684.
16. Naumovski, L. and Friedberg, E.C. (1982) J. Bacteriol. 152, 323-331.
17. Naumovski, L. and Friedberg, E.C. (1983) Proc. Natl. Acad. Sci. USA 80, 4818-4821.
18. Higgins, D.R., Prakash, S., Reynolds, P. and Prakash, L. (1983) Gene 26, 119-126.
19. Higgins, D.R., Prakash, L., Reynolds, P. and Prakash, S. (1984) Gene 30, 121-128.
20. Prakash, L., Dumais, D., Polakowska R., Perozzi, G. and Prakash, S. (1985) Gene 34, 55-61.
21. Naumovski, L. and Friedberg E.C. (1984) Mol. Cell Biol. 4, 290-295.
22. Weaver R.F. and Weissman, C. (1979) Nucl. Acids Res. 7, 1175-1193.
23. Berk, A.J. and Sharp, P.A .(1978) Proc. Natl. Acad. Sci. USA 75, 1274-1278.
24. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) *Molecular Cloning. A laboratory manual.*, Cold Spring Harbor Laboratory, New York.
25. Sanger, F., Nicklen, S. and Coulson A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.
26. Biggin, M.D., Gibson , T.J. and Hong, G.F. (1983) Proc. Natl. Acad. Sci. USA 80, 3963-3965.
27. Scherer, S. and Davis, R.W. (1979) Proc. Natl. Acad. Sci. USA 76, 4951-4955.
28. Southern E.M. (1975) J. Mol. Biol. 98, 503-517.
29. Sherman F., Stewart, J.W. and Schweingruber, A.M. (1980) Cell 20, 215-222.
30. Donahue, T.F., Farabaugh, P.J. and Fink, G.R. (1982) Gene 18, 47-59.
31. Nagpal, M.L., Higgins, D.R. and Prakash, S. (1985) Mol. Gen. Genet., in press.
32. Bennetzen, J.L. and Hall, B.D. (1982) J. Biol. Chem. 257, 3018-3025.
33. Gannon, F., O'Hare, K., Perrin, F., LePennec, J.P., Benoist, C., Cochet, M., Breathnach, R., Royal, A., Garapin, A., Cami, B. and Chambon, P. (1979) Nature 278, 428-434.
34. Breathnach, R. and Chambon, P. (1981) Ann. Rev. Biochem. 50, 349-383.
35. Grosschedl, R. and Birnstiel, M.L. (1980) Proc. Natl. Acad. Sci. USA 77, 1432-1436.
36. Mathis, D.J. and Chambon, P. (1981) Nature 290, 310-315.
37. Benoist, C., O'Hare, K. Breathnach, R. and Chambon, P. (1980) Nucl. Acids Res. 8, 127-142.
38. Faye, G., Leung, D.W., Tatchell, K., Hall, B.D. and Smith, M. (1981) Proc. Natl. Acad. Sci. USA 78, 2258-2262.
39. Burke, R.L., Tekamp-Olson, P. and Najarian, R. (1983) J. Biol. Chem. 258, 2193-2201.
40. Dobson, M.J., Tuite, M.F., Roberts, N.A., Kingsman, A.J., Kingsman, S.M., Perkins, R.E., Conroy, S.C., Dunbar, B. and Fothergill, L.A. (1982) Nucl. Acids Res. 10, 2625-2637.
41. Kozak, M. (1981) Nucl. Acids Res. 9, 5233-5252.
42. Proudfoot, N.J. and Brownlee, G.G. (1976) Nature 263, 211-214.
43. Fitzgerald, M. and Shenk, T. (1981). Cell 24, 251-260.
44. Zaret, K.S. and Sherman, F. (1982) Cell 28, 563-573.
45. Henikoff, S., Kelly, J.D. and Cohen E.H. (1983) Cell 33, 607-614.
46. Rothstein, R.J. (1983) in *Methods in Enzymology* (ed. Wu, R., Grossman, L. and Moldave, K.) Academic Press, New York, pp. 202-211.
47. Kyte, J. and Doolittle, R.F. (1982) J. Mol. Biol. 157, 105-132.
48. Sancar, G.B., Sancar, A. and Rupp, W.D. (1984) Nucleic Acids Res. 12, 4593-4608.
49. Finch, P.W. and Emmerson, P.T. (1984) Nucleic Acids Res. 12, 5789-5799.

50. Rothman, R.H. and Clark, A.J. (1977) Mol. Gen. Genet. 155, 267-277.
51. Kuemmerle N.B. and Masker, W.E. (1980) J. Bacteriol. 142, 535-546.
52. Kuemmerle, N.B. and Masker, W.E. (1983) Nucleic Acids Res. 11, 2193-2204.
53. Horii, Z.-I., and Clark, A.J. (1973) J. Mol. Biol. 80, 327-344.
54. Lu, A.L., Clark, S. and Modrich, P. (1983) Proc. Natl. Acad. Sci. USA 80, 4639-4643.
55. Klinkert, M.Q., Klein, A. and Abdel-Monem, M. (1980) J. Biol. Chem. 255, 9746-9752.
56. Kumura, K. and Sekiguchi, M. (1984) J. Biol. Chem. 259, 1560-1565.
57. Walker, J.E., Saraste, M., Runswick, M.J. and Gay, N.J. (1982) EMBO J. 1, 945-951.
58. Reynolds P., Weber, S. and Prakash, L. (1985) Proc. Natl. Acad. Sci USA, 82, 168-172.
59. Walker, J.M. (1982) in *The HMG Chromosomal Proteins* (ed. Johns, E.W.) Academic Press, New York, pp. 69-87.