## Nucleotide sequence of the RAD10 gene of Saccharomyces cerevisiae

# Paul Reynolds, Louise Prakash, Dennis Dumais<sup>1</sup>, Giuditta Perozzi<sup>1</sup> and Satya Prakash<sup>1</sup>

Department of Radiation Biology and Biophysics, University of Rochester School of Medicine, Rochester, NY 14642, and <sup>1</sup>Department of Biology, University of Rochester, Rochester, NY 14627, USA

Communicated by F.Lacroute

The *RAD10* gene is one of several genes in *Saccharomyces* cerevisiae required for incision of u.v.-irradiated or crosslinked DNA. We have determined the nucleotide sequence of the *RAD10* gene and its flanking regions. The *RAD10* nucleotide sequence presented here differs significantly from that recently reported. The RAD10 protein predicted from the nucleotide sequence contains 210 amino acids with a calculated mol. wt. of 24 310. The middle portion of the RAD10 protein, which is highly basic and also contains eight of the total of 10 tyrosine residues present in the protein, may be involved in DNA binding by ionic interactions and tyrosine intercalation between the bases of DNA. A genomic deletion of the entire *RAD10* gene does not affect viability; however, the *rad10* deletion mutant is highly u.v. sensitive.

Key words: DNA repair/incision/RAD10 protein/DNA binding/Saccharomyces cerevisiae

#### Introduction

In Saccharomyces cerevisiae, the RAD1, RAD2, RAD3, RAD4, RAD10, MMS19, RAD7, RAD14, RAD16 and RAD23 genes are required for excision of pyrimidine dimers or inter-strand DNA cross-links. The rad1, rad2, rad3, rad4, rad10 and mms19 mutants are highly defective in incision of DNA containing pyrimidine dimers (Wilcox and Prakash, 1981; Reynolds and Friedberg, 1981) or inter-strand cross-links (Miller et al., 1982a; Jachymczyk et al., 1981; Magaña-Schwenke et al., 1982), while the other mutants show varying degrees of incision defect (Wilcox and Prakash, 1981; Miller et al., 1982a, 1982b). To elucidate the structure, function and regulation of genes involved in incision, we have cloned and characterized several of these genes (Higgins et al., 1983a, 1983b, 1984; Prakash et al., 1985; Reynolds et al., 1985; Nagpal et al., 1985). Previously, we had physically mapped the rad10 complementing function to a 1.0-kb DNA fragment, and had found that a genomic disruption of the RAD10 gene showed much greater u.v. sensitivity than the previously available rad10 mutants (Prakash et al., 1985). Here we present the nucleotide sequence of the RAD10 protein coding region and flanking regions and examine the effects of a genomic deletion of the entire RAD10 gene. The RAD10 gene encodes a protein of 210 amino acids with a calculated mol. wt. of 24 310. Even though the RAD10 protein shows no amino acid sequence homology with any of the single-stranded DNA-binding proteins of bacteriophages and Escherichia coli, some of its structural features resemble the fd gene 5 and T4 gene 32 single-strand DNA-binding proteins.

## Nucleotide sequence of the RAD10 gene

Previously, we had shown that the rad10 complementing activity is located within  $\sim 1.0$  kb of DNA between a PvuII site on the left and between XbaI and EcoRV sites on the right (Figure 1; Prakash et al., 1985). The nucleotide sequence of the RAD10 gene (Figure 2) shows a long open reading frame in this region starting with the ATG codon at +1 and ending with a TGA codon at +631. There is no other ATG codon until position +367. We had previously obtained RAD10 region DNA segments of varying lengths by Bal31 deletions generated from the EcoRV site (Figure 1 and Prakash et al., 1985). These fragments were cloned into yeast multicopy plasmids and tested for their ability to complement rad10 mutants. Nucleotide sequence determination of the RAD10 insert generating the rad10 complementing plasmid pDD16 revealed its end point at +667, which lies 34 nucleotides 3' to the termination codon TGA. On the other hand, the RAD10 insert in the rad10 non-complementing plasmid pDD17 has its end point at +557, which is 25 codons upstream of the RAD10 termination codon. These results are consistent with the location of the RAD10 open reading frame (Figure 2).

We next determined if the *RAD10* open reading frame is translated in *S. cerevisiae* and produces a protein by fusing the *E. coli lacZ* gene with the *RAD10* gene. The 6.8-kb *Bam*HI fragment of pMC931 containing the *lacZ* gene but missing its promoter and the first seven amino acid codons (Casadaban *et al.*, 1980) was fused with the *Bam*HI-linkered *RAD10* gene at position +421 (Figure 3). This fusion connects the *RAD10* reading frame with the *lacZ* reading frame and expresses  $\beta$ -galactosidase in *S. cerevisiae*. However, a *lacZ* fusion at position +557, which is out-of-frame with the *RAD10* reading frame (Figure 3), does not show any  $\beta$ -galactosidase activity in *S. cerevisiae*. These results indicate that the *RAD10* open reading frame is translated in *S. cerevisiae*. The *RAD10*-encoded protein predicted by the nucleotide sequence shown in Figure 2 would contain 210 amino acids with a predicted mol. wt. of 24 310.

Recently, Weiss and Friedberg (1985) published a nucleotide sequence for the RAD10 gene containing an open reading frame of 195 codons. Our sequence differs from theirs primarily downstream of position +531. Our sequence, which contains an additional C at +532 (Figure 4A), continues for 33 codons until the termination codon TGA is reached at +631. Their sequence lacks the C at +532 and thus changes the reading frame downstream of this position. Our evidence from the productive in-frame and non-productive out-of-frame RAD10:lacZ fusions (Figure 3) confirms the identity of the RAD10 open reading frame presented in Figure 2. In addition, we observe GAC(Asp)GAT(Asp), at positions +508 to +513 (Figure 4B), whereas they report the codons GAG(Glu)CAT(His); and the nucleotide A present at +563 in our sequence (Figure 4C) is missing in their sequence. Other differences are located in the 3'-untranslated region: we observe the nucleotides C, A, C and C, at positions +795, +802,



Fig. 1. Strategy for sequencing the *RAD10* gene. The recognition sites for *Sau3A* are indicated only in the region where they were used for M13 cloning and DNA sequencing. The ATG initiation codon of the *RAD10* gene, to the left of the *Sau3A* site, and the TGA termination codon between the *Xba1* and *EcoRV* sites, are indicated. The *EcoRV* site was converted to a *BamHI* site by *BamHI* linker ligation to facilitate cloning in M13mp18 and M13mp19 phages. The horizontal lines with arrowheads indicate the extent and direction of sequencing. The asterisks indicate the site at which *BamHI* linkers were attached following *Bal31* digestion from the *EcoRV* site.

															Salī	-50	0		
															GTC	GACC	ATGA	ATT	
	-49	0	-	480		-47	0	-	460		-45	0	-	440		-43	0		
	CTTATTGTCTACAATCTTACCTGG							TGTA	* T <b>AA</b> T	GATG	CAGA	TGCT	GGAA	* CCAG	* IGTAGTTTTTGCCCTTCA				
	-42	0	-	410		-40	0	-	390		- 38	0	-	370		- 36	0		
	* тт	тсст	ттла	*	GACT	* 6TTT			* TC & &	TACA	*	GTGA	ACTT	*	TTAG	* •	CGTT		
	25	n		240			•		220	1.101		0. 0		3000	1170	20	0		
	*	-350 -340 * *			- 330			*		*	•	-	*	-290 			Л		
	CTTGCAAGGACTGTGTGAAATGCTGCAA							AATT	ATTCTTGGTCGAATTGCTGTTTT					ACAG	CATT	CAGC	TGCT	CGGG	
	-28	0	-	270 *		-26	0	-	250 *		-24	0	-	230 *		-22	0		
	AT	TTAG	TTTA	GTTA	MTC	GACT	AAGT	TTCG	AACA	TGCA	GTAA	CTTA	AAAG	TTAG	TAAA	GTGT	CTGT	ACGG	
	-21	0	-	200		-19	0	-	180		-17	0	-	160		-15	0		
	ŤG	GCAA	CAAC	* GAAG	GAGC	TATT	CATT	TACA	* Tact	ттт	* TGAG	AGGA	CATG	* GCTT	GATT	TTTA	CAGT	GCTC	
	-14	0	_	130		-12	0	-	110		-10	0	-	90		-80			
	*	-		*		*			*		*	-		*		*			
TTTCTTGTTTGGGTCACAGCAAGATTTTCATCTAAGACACTTTTCTATTTAAAATCGTTAGAACAAAAAA																			
	-70 *		-	60 *		-50 *		-	40 *		-30		-	20 *		-10			
	GA	AAAT	TGTA	ACTT	ATGA	GACA	GCCA	CGTA	ACAC	****	AAGG	GCAT	AAAC.	AAAG	TTGG	TTAT	CCTA	GAAG	
1			10			20			3	0			40 *			50 *			60 *
ATG	AAC	MT	ACT	GAT	сст	ACT	TCA	TTT	GAA	AGT	ATA	TTG	GCT	GGT	GTG	GCC		TTG	AGA
Met	Asn	Asn	Ihr	Asp	Pro	Ihr	Ser	Phe	Glu	Ser	lle	Leu	Ala	Gly	Val	Ala	Lys	Leu	Arg
			70 *			80 *			9	0			100			110			120 *
AAG	GAA	AAG	AGT	GGT	GCA	GAT	ACT	ACC	GGA	TCT	CAA	TCG	TTA	GAA	ATA	GAT	GCA	TCA	<b>AAA</b>
Lys	910	Lys	ser	GIY	714	ASP	Inr	Inr	GIY	ser	GIN	ser	Leu	GIU	Tie	ASP	~14	ser	Lys
			130			140			15	0			160 *			170			180 *
CTT	CAG G1n	CAG G1n	CAG G1n	GAA G1u	CCG	CAA Gln	ACA Thr	TCA	AGA Ara	CGC Ara	ATA	AAT Asn	TCA	AAC Asn	CAG Gln	GTC Val	ATA	AAC Asn	GCC Ala
	<b>u</b>		100			200		501	~ ~ ~	~~ y			200			220	•••	~	240
_			*			*			*			4	*			*			*
TTC Phe	AAT Asn	CAG G1n	CAA G1 n	AAA Lys	CCG Pro	GAA G1u	GAA G1u	TGG Trp	ACC Thr	GAC Asp	TCA Ser	AAG Lys	GCT Ala	ACA Thr	GAT Asp	GAC Asp	TAT Tyr	AAT Asn	CGA Arg
		:	250			260			27	D		;	280			290			300
AAG	AGA	222	* 111	AGG	AGT	*	664	977	*		ACT	GTA	* CTA	GTG		* ACT	104		*
Lys	Arg	Pro	Phe	Arg	Ser	Thr	Arg	Pro	Gly	Lys	Thr	Val	Leu	Val	Asn	Thr	Thr	Gln	Lys

310				320	330					340				350 3			360		
GAA Glu	AAT Asn	CCA Pro	× CTC Leu	CTG Leu	AAC Asn	CAC His	TTA Leu	AAG Lys	AGC Ser	ACC Thr	AAT Asn	TGG Trp	ÅGA Årg	TAT Tyr	GTA Val	TCT Ser	TCA Ser	ACA Thr	GGA G1y
			370			380			390	)		4	4 <b>0</b> 0			410			420 *
ATT Ile	AAT Asn	ATG Met	ATA Ile	TAC Tyr	TAC Tyr	GAC Asp	TAT Tyr	CTA Leu	GTT Val	CGT Arg	GGA G1y	AGA Arg	AGT Ser	GTA Val	CTG Leu	TTC Phe	TTA Leu	ACT Thr	TTG Leu
			430			440			450	)			460			470	Xb	aI	480 *
ACT Thr	TAT Tyr	CAC His	ÂAA Lys	TTA Leu	TAT Tyr	GTC Val	GAT Asp	TAT Tyr	ATC 11e	TCT Ser	AGG Arg	AGA Arg	ATG Met	CAG G1n	CCC Pro	TTA Leu	TCT Ser	AGA Arg	AAC Asn
	490				500 *		510				520 *				530 *			540 *	
GAA Glu	AAT Asn	AAT Asn	ATA Ile	CTG Leu	ATA Ile	TTC Phe	ATA Ile	GTA Val	GAC Asp	GAT Asp	AAC Asn	AAC Asn	TCT Ser	GAA G1u	GAT Asp	ACC Thr	CTT Leu	AAT Asn	GAC Asp
			550			560			570	)			580			590 *			600 *
ATT Ile	ACA Thr	AAA Lys	CTA Leu	TGT Cys	ATG Met	TTC Phe	AAC Asn	GGA G1y	TTT Phe	ACT Thr	CTA Leu	CTT Leu	TTA Leu	GCA Ala	TTT Phe	AAT Asn	TTT Phe	GAA G1u	CAA G1n
			610			620			63	)			6	40		650 *			
						_			-										
GCT Ala	GCA A1a	AAA Lys	* TAT Tyr	ATT Ile	GAA Glu	TAT Tyr	TTG Leu	AAT Asn	TTA Leu	TGA		AC.	AATTO	CCTA	ITTI	CTTT	TAAT	MAT	
GCT Ala	GCA A1a 660	AAA Lys	* TAT Tyr 6	ATT Ile 70	GAA Glu	TAT Tyr 680	TTG Leu	AAT Asn 6	TTA Leu 90	TGA	7 <b>0</b> 0	AC.	<b>AATT</b>	сста <sup>.</sup> 10 *	TTT	720	TAAT	MAT	
GCT Ala	GCA Ala 660 *	AAA Lys GTTC	TAT Tyr 6 CATG	ATT Ile 70 *	GAA Glu TTAC	TAT Tyr 680 *	TTG Leu	AAT Asn 6 GTCA	TTA Leu 90 * TTTT	TGA 	700 *	AC. TAGC	AATTO 7 ACCG	CCTA 10 * •	GAAT	CTTT 720 GGCC	CTTT	NAAT TTGG	
GCT Ala	GCA A1a 660 * CT/ 730	AAA Lys GTTC	TAT Tyr 6 CATG	ATT 11e 70 * CTTA 40	GAA Glu TTAC	TAT Tyr 680 * CATCI 750	TTG Leu	AAT Asn 6 GTCA 7	TTA Leu 90 * TTTT/ 60	TGA 	700 cttt 770	AC. TAGC	AATT 7 ACCG	CCTA 10 * MATTO 80 *	GAAT	720 720 GGCCI 790	TAAT, CTTT Hinc]	MAAT TTGG	
GCT Ala	GCA A1a 660 * CT 730 * TT	AAA Lys GTTC CTGA	TAT Tyr 6 CATG 7 .TGAT	ATT 11e 70 CTTA 40 40	GAA Glu TTAC	TAT Tyr 680 CATCO 750 CAAC	TTG Leu CTTTC	AAT Asn 6 GTCA 7 AATA	TTA Leu 90 * TTTT 60 * CTGC	TGA  AGCG	700 CTTT 770 CAAC	AC. TAGC	AATT 7 ACCG 7 CGAA	CCTA 10 * AATT( 80 * GGAT	GAAT	720 66CCC 790 71	TAAT, CTTT Hinc] TAAC	TTGG I CTTG	
GCT Ala	GCA A1a 660 * CT/ 730 * TT 800	AAA Lys GTTC CTGA	* Tyr 6 CATG 7 .TGAT. 8	ATT 11e 70 CTTA 40 ATTC 10	GAA Glu TTAC	TAT Tyr 680 CATCO 750 CAAC 820	TTG Leu CTTTC	AAT Asn 6 GTCA 7 AATA 8	TTA Leu 90 * TTTT 60 * CTGC 30	TGA  NGCG	700 CTTT 770 CAAC 840	AC. TAGC	AATT 7 ACCG 7 CGAA	CCTA 10 * AATTO 80 * GGAT 50	GAAT	720 GGCC 790 TTGT 860	TAAT, CTTT Hinc] TAAC	TTGG I CTTG	
GCT A1a	GCA A1a 660 * CT 730 * TT 800 * AA	AAA Lys GTTC CTGA	TAT Tyr 6 CATG 7 .TGAT. 8 ATGG	ATT 11e 70 CTTA 40 ATTC 10 SAAT	GAA Glu TTAC CATC	TAT Tyr 680 CATCO 750 CAAC 820 GCTG	TTG Leu CTTTC TTCA	AAT Asn 6 GTCA 7 AATA 8 TAAC	TTA Leu 90 * TTTT/ 60 * CTGC 30 * TTTA	TGA  AGCG GAGG	700 CTTT 770 CAAC 840 ATAT	AC. TAGC TTTT TCAT	AATT 7 ACCG 7 CGAA 8 CTAA	CCTA 10 * AATTO 80 * GGAT 50 * AAATO	GAAT TAAT GGGT/	720 GGCC0 790 TTGT 860	CACC	TTGG I CTTG TCCA	
GCT Ala	GCA A1a 6600 * 730 * TT 8000 * 800 * 800 *	AAA Lys GTTC CTGA	TAT Tyr 6 CATG 7 TGAT 8 ATGG 8	ATT 11e 70 * CTTA 40 * ATTC 10 * GAAT 80 *	GAA Glu TTAC	TAT Tyr 680 CATCI 750 CAAC 820 6CTG 890	TTG Leu CTTTC TTCA	AAT Asn GTCA GTCA 7 AATA 8 TAAC 9	TTA Leu 90 * TTTT/ 60 * CTGC 30 * TTTA 00	TGA  AGCG GAGG	700 CTTT 770 CAAC 840 ATAT 910	AC. TAGC TTTT TCAT	AATT 7 ACCG 7 CGAA 8 CTAA 9	CCTA 10 * AATTO 80 * GGAT 50 * AAATO 20 *	GAAT TAAT GGGT/	720 GGCC0 790 TTGT 860 ATCT0 930	TAAT	TTGG I CTTG TCCA	
GCT Ala	GCA A1a 660 * 730 * TT 800 * AA 870 * CA	AAA Lys GTTC CTGA	* TAT Tyr 6 CATGO 7 TGAT. 8 ATGG 8 ATGG	ATT Ile 70 * 40 * ATTC 10 * 50 * AGCT	GAA Glu TTAC CATC TTCT	TAT Tyr 680 * CATCO 750 * CAAC 820 * 820 * 820 * * 820 * * * * * * * * * * * * * * * * * * *	TTG Leu CTTTC TTCA TAGA	AAT Asn 6 GTCA 7 AATA 8 TAAC 9 TCAC	TTA Leu 90 * TTTT/ 60 * CTGC 30 * TTTA * 00 * GATT	TGA  GAGCG GAGG TAAC.	700 cttt 770 cAAC 840 ATAT 910 TTGC	AC. TAGC TTTT TCAT GATC	AATT 7 ACCG 7 CGAA 8 CTAA 9 ATAA	CCTA 10 * AATTO 80 * 60 * 50 * AAATO 20 * ATGT	GAAT TAAT GGGT	720 720 GGCCC 790 * TTGT 860 * 930 * TATT	TAAT. CTTT Hincl TAAC CACC	MAT TTGG I CTTG TCCA	
GCT Ala	GCA A1a 660 * 730 * TT 800 * AA 870 * CA 940 *	AAA Lys GTTC CTGA	TAT Tyr 6 CATG 7 TGAT 8 ATGG 8 ATGG 9	ATT 11e 70 * CTTA 40 * ATTC 10 * GAAT 80 * AGCT. 50	GAA Glu TTAC CATC TTCT	TAT Tyr 680 CATCI 750 CAAC 820 GCTG 890 CTTT, 960	TTG Leu CTTTC TTCA TAGA	AAT Asn 6 GTCA 7 AATA 8 TAAC 9 TCAC 9	TTA Leu 90 * TTTT/ 60 * CTGC: 30 * TTTA * GATT: 70 *	TGA  GAGG TAAC	700 cttt 770 cAAC 840 ATAT 910 TTGC 980 *	AC. TAGC TTTT TCAT GATC	7 ACCG 7 CGAA 8 CTAA 9 ATAA	CCTA 10 * 80 * 60 * 60 * 60 * 80 * 60 * 60 * 60 * 60 * 60 * 60 * 60 * 60 * 60 * * 80 * * 80 * * 80 * * 80 * * 80 * * 80 * * 80 * * 80 80 * * * 80 * * 80 * * 80 * * 80 * * * * * * * * * * * * *	GAATI GAATI GGGT/	720 566CCI 790 TTGT 860 4 4 7 7 930 7 4 7 4 7	TAAT. CTTT Hinc] TAAC CACC	NAAT TTGG I CTTG TCCA	

Fig. 2. Nucleotide sequence of the RAD10 gene. The sequence of 1491 nucleotides in the DNA strand identical to the mRNA is presented. The numbering is in relation to the first base of the ATG translation initiation codon, indicated as +1. The predicted amino acid sequence encoded by the RAD10 open reading frame is shown below the nucleotide sequence. Pertinent restriction sites are indicated. The last three nucleotides indicated in the sequences are the first three nucleotides of the *Eco*RV restriction site 5' GATATC 3'.

+868 and +872, respectively, whereas these four nucleotides are missing in their sequence.

## Codon usage in RAD10

The 630 nucleotides of the *RAD10* coding region possess 37.8% A, 26.2% T, 17.1% G and 18.9% C. The *RAD10*-encoded protein contains 35.2% non-polar, 40.5% polar, 11.0% acidic and 13.3% basic amino acids. In the *RAD10* gene, 52 of the possible 61 codons are used. In the majority of cases, the *RAD10* gene uses codons not used in the highly expressed genes of *S. cerevisiae* such as alcohol dehydrogenase I and glyceraldehyde-3-phosphate dehydrogenase (Bennetzen and Hall, 1982). For example, the codons GAU(Asp), AAA(Lys), GCA(Ala), UCA, AGU(Ser), ACA(Thr), GUA, GUG(Val), AUA(Ile), UUU(Phe), UAU(Tyr), AAU(Asn), UUA, CUA(Leu), GGA(Gly), CAG(Gln) and CCG, CCC(Pro) are used frequently in the *RAD10* gene whereas these codons are used rarely or are absent from the highly expressed yeast genes (Bennetzen and Hall, 1982).

## 5'- and 3'-flanking sequences of the RAD10 gene

In the region from -292 at the *Pvu*II site to +1 which is likely to include the entire *RAD10* upstream region, the base composition is 33.9% A, 30.8% T, 19.5% G, and 15.8% C. Our S1 mapping results indicate two *RAD10* mRNA start sites, approximately at positions -17 and -32 in Figure 2 (results not shown).

Г <sup></sup>	-RADIO se	equence -		('	Bam HI _ linker	I	— Laca	z in-frame	e sequence
410 5' TTC Phe	TTA Leu	ACT Thr	420 TTG Leu	A <b>GG</b> Arg	GAT Asp	<b>C</b> CC Pro	GTC Val	GTT Val	TTA 3' Leu

RADIO sequence ——— BamHI — LacZ out-of-frame ——													
	550 ČTA	тот	ATC	CCA	-	0.07	seque	ence					
5' Lys	Leu	Cys	Alg	GGA	ICC.	CGI	CGI	3					

Fig. 3. In-frame and out-of-frame *RAD10:LacZ* fusions. Nucleotide sequence of the *RAD10* gene in the vicinity of the *lacZ* fusion is shown followed by the nucleotide sequence of the *Bam*HI linker and the *lacZ* gene. The upper fusion represents an in-frame fusion which produces  $\beta$ -galactosidase and therefore the amino acids encoded by the *lacZ* gene are indicated below the nucleotide sequence. The lower fusion, on the other hand, represents an out-of-frame fusion and therefore, only the amino acids encoded by the *RAD10* gene are indicated. Numbers above the *RAD10* nucleotides indicate their position as in Figure 2.  $\beta$ -galactosidase activities were determined as in Nagpal et al. (1985).

Upstream of these mRNA initiation sites, there is no sequence identical to the 5'-TATAA/TAA/T-3' sequence that has been proposed to be required for proper transcription initiation in higher eukaryotes (Benoist et al., 1980; Grosschedl and Birnstiel, 1980; Breathnach and Chambon, 1981; Mathis and Chambon, 1981). However, an AT-rich sequence, TATTTAAAAT, occurs in the RAD10 gene at position -95 to -86. The sequences GGACATGGCTTGATTT at position -167 to -152 and GGTCACAGCAAGATTT at position -129 to -114 are direct repeats in which 11 of the 16 nucleotides are identical. In eukaryotes, translation usually begins with the first ATG codon in the mRNA (Kozak, 1984; Baim et al., 1985); however, it has been proposed that the efficiency of translation is influenced by the nucleotides at positions -3 and +4. A purine, an A, frequently occurs at position -3 and a purine, usually a G, is found at position +4 in eukaryotic mRNAs (Kozak, 1981, 1984). The *RAD10* gene has an A at positions -3 and at +4. The sequence 5'-TTATCCT-3' from positions -12 to -6 shows complementarity with the sequence 3'-ACUAGGA-5' present at the 3' ends of 18S rRNAs of S. cerevisiae (Rubtsov et al., 1980) and higher eukaryotes (Hagenbüchle et al., 1978) and could be involved in binding of mRNA by ribosomes.

The sequence of the 353 nucleotides downstream of the TGA codon contains 26.3% A, 38.5% T, 15.9% G, and 19.3% C (Figure 2). In *S. cerevisiae*, the sequence TAAATAA A/G has been observed 28 - 33 nucleotides upstream from the 3' mRNA terminus of various genes (Bennetzen and Hall, 1982). Henikoff *et al.* (1983) have suggested that the sequence TTTTTATA is required for transcription termination in yeast, and Zaret and Sherman (1982) have identified a sequence TAG . . . . . TAGT or TATGT . . (AT-rich) . . TTT in the 3' mRNA terminus of various yeast genes and have implicated it in transcription termination and polyadenylation. The *RAD10* 3'-flanking region contains several similar sequences.

## Deletion of the RAD10 gene

We previously showed that a yeast genomic disruption of the *RAD10* gene, made by insertion of the *URA3* gene at the *XbaI* site at position +472 (Figure 2; Prakash *et al.*, 1985) was viable but exhibited much enhanced u.v. sensitivity compared with the *rad10-1* and *rad10-2* mutants. We have now deleted the entire *RAD10* gene in the yeast genome, from *SaII* to *BgIII*, which represents the entire nucleotide sequence shown in Figure 2 plus  $\sim 0.4$  kb of DNA downstream of nucleotide +986 at the *Eco*RV site, and replaced it with the yeast *LEU2* gene and find that it



Fig. 4. Nucleotide sequence of various portions of the RAD10 gene. (A) Nucleotide sequence of the DNA strand complementary to the DNA strand which is identical to the mRNA, from positions +524 to +538. The additional G is seen at +532. (B) Nucleotide sequence of the DNA strand identical to the mRNA from positions +505 to +513, indicating the GAC GAT sequence at position +508 to +513. (C) Nucleotide sequence of the DNA strand identical to the mRNA from positions +557 to +571, indicating the presence of an A nucleotide at +563.

also has no effect on viability. The u.v. sensitivity of the rad10 deletion is similar to that reported by us for the rad10 disruption (Prakash *et al.*, 1985). We and others have previously shown that disruptions of the *RAD1* and *RAD2* genes do not affect viability (Higgins *et al.*, 1983b, 1984; Naumovski and Friedberg, 1984), whereas disruptions and deletions of the *RAD3* gene are recessive lethals (Higgins *et al.*, 1983; Naumovski and Friedberg, 1983; Reynolds *et al.*, 1985).

## Discussion

The *RAD10* open reading frame encodes a protein of 210 amino acids of mol. wt. 24 310. A computer search of the data bank in the National Biomedical Research Foundation library was carried out according to Lipman and Pearson (1985). However, no significant homologies with the RAD10 protein could be identified. The RAD10 protein is slightly basic, containing 13.3% basic and 11.0% acidic residues, but the distribution of charged residues along the protein is not random (Figure 2). The amino acids 1-77 contain 12 acidic and nine basic residues, the next 82 amino acids from 78 to 159 contain three acidic and 17 basic residues and the last 51 amino acids from 160 to 210 contain eight acidic and two basic residues.

In addition to the high density of basic residues, the amino acids 78 - 159 contain eight of the 10 tyrosine residues present in the RAD10 protein. This region of the RAD10 protein could be involved in DNA binding through ionic interactions and by intercalation of tyrosine residues between the bases of DNA. Tyrosine intercalation has been shown to contribute to the binding of bacteriophage fd gene 5 protein to single-stranded DNA (Anderson et al., 1975; McPherson et al., 1979). The gene 32 protein of bacteriophage T4 binds tightly and cooperatively to single-stranded DNA and functions in DNA replication, recombination and repair. In the amino-terminal half of the gene 32 protein, the amino acids between 72 and 116 contain six of the eight tyrosine residues in gene 32 protein (Williams *et al.*, 1981). Studies by Anderson and Coleman (1975) have suggested that five of the tyrosyl residues participate in binding of gene 32 protein with DNA.

The secondary structures of the fd gene 5, T4 gene 32, and yeast RAD10 proteins show similarity. The amino acid residues 12-49 of fd gene 5 protein, involved in DNA binding, occur in three stranded anti-parallel  $\beta$ -sheet (McPherson *et al.*, 1979). The Chou-Fasman (1978) predicted secondary structure of gene 32 protein from residues 72 to 116 shows three short  $\beta$ -sheet regions and several  $\beta$ -turns (Williams *et al.*, 1981). The Chou-Fasman (1978) predicted secondary structure of the middle basic region of RAD10 protein from residues 78 to 159 consists predominantly of  $\beta$ -sheets and  $\beta$ -turns.

### Materials and methods

#### Sequencing strategy

DNA segments of varying lengths were obtained from the ~1.5-kb Sal1-EcoRV RAD10 fragment by digestion with restriction enzymes having either a 4-base or 6-base recognition sequence, or by ligation of BamHI linkers following Bal31 digestion from the EcoRV site (Figure 1). The EcoRV site itself was also converted to a BamHI site by attachment of BamHI linkers (Boehringer<sup>4</sup>Mannheim) to facilitate cloning in the M13 derivative phages. These DNA segments were then inserted into phages M13mp18 and M13mp19 (Norrander et al., 1983) for DNA sequencing by the dideoxy method (Sanger et al., 1977) using deoxy-adenosine 5'-( $\alpha$ -[<sup>35</sup>S]thio)triphosphate as described (Biggin et al., 1983).

#### Generation of genomic rad10 deletions

A genomic deletion of the *RAD10* gene was constructed by the replacement of the chromosomal *RAD10* gene by gene conversion using the yeast *LEU2* gene flanked by *RAD10* region DNA (Rothstein, 1983). A 6.9-kb *BamHI-XhoI* fragment which contains the 1.9-kb *SalI-BgIII* fragment with *rad10* complementing

function (Prakash *et al.*, 1985) was used as the starting material and cloned into *BamHI-SalI* digested pBR322, to generate the plasmid pDD40. The 1.9-kb *SalI-BgIII RAD10* fragment of pDD40 was then removed and replaced with a 2.5-kb *SalI-BgIII RaD10* fragment containing the *LEU2*, gene, generating the plasmid pDD41. A 5.6-kb *BamHI-PvuII* fragment from pDD41 was used to transform the Rad<sup>+</sup> strain DBY747 (*MATa his3-* $\Delta I$  *leu2-3 leu2-112 trp1-289 ura3-52*) to Leu<sup>+</sup>. This 5.6-kb *BamHI-PvuII* fragment contains ~ 1 kb of *RAD10* region DNA upstream of the *SalI* site at -505 (Prakash *et al.*, 1985 and Figure 2), followed by 2.5 kb of DNA containing the *LEU2* gene, followed by ~ 2.2 kb of DNA 3' to the *BgIII* site lying ~ 0.4 kb downstream of the *Eco*RV site at +984 (Prakash *et al.*, 1985 and Figure 2). All Leu<sup>+</sup> transformants obtained with this 5.6-kb *BamHI-PvuII* fragment were Rad<sup>-</sup> and allelic to *rad10*.

#### Acknowledgements

We thank Alan Morrison for a critical reading of the manuscript. This work was supported by grants CA35035 from NIH and DE-FG02-85ER60281 from DOE.

#### References

- Anderson, R.A. and Coleman, J.E. (1975) *Biochemistry (Wash.)*, 14, 5485-5491.
  Anderson, R.A., Nakashima, Y. and Coleman, J.E. (1975) *Biochemistry (Wash.)*, 14, 907-917.
- Baim, S.B., Goodhue, C.T., Pietras, D.F., Eustice, D.C., Labhard, M., Friedman, L.R., Hampsey, D.M., Stiles, J.I. and Sherman, F. (1985) in Calendar, C. and Gold, L. (eds.), Sequence Specificity in Transcription and Translation, UCLA Symposia on Molecular and Cellular Biology, Vol. 30, Alan R. Liss, Inc., NY, in press.
- Bennetzen, J.L. and Hall, B.D. (1982) J. Biol. Chem., 257, 3018-3025.
- Benoist, C., O'Hare, K., Breathnach, R. and Chambon, P. (1980) Nucleic Acids
- Res., 8, 127-142. Biggin, M.D., Gibson, T.J. and Hong, G.F. (1983) Proc. Natl. Acad. Sci. USA, 80, 3963-3965.
- Breathnach, R. and Chambon, P. (1981) Annu. Rev. Biochem., 50, 349-383.
- Casadaban, M.J., Chou, J. and Cohen, S.N. (1980) J. Bacteriol., 143, 971-980.
- Chou, P.Y. and Fasman, G.D. (1978) Adv. Enzymol., 47, 45-148. Grosschedl, R. and Birnstiel, M.L. (1980) Proc. Natl. Acad. Sci. USA, 77,
- 1432-1436. Hagenbüchle, O., Santer, M., Steitz, J.A. and Mans, R.J. (1978) Cell, 13, 551-563.
- Henikoff,S., Kelly,J.D. and Cohen,E.H. (1983) *Cell*, **33**, 607-614.
- Higgins, D.R., Prakash, S., Reynolds, P., Polakowska, R., Weber, S. and Prakash, L. (1983a) Proc. Natl. Acad. Sci. USA, 80, 5680-5684.
- Higgins, D.R., Prakash, S., Reynolds, P. and Prakash, L. (1983b) Gene, 26, 119-126.
- Higgins, D.R., Prakash, L., Reynolds, P. and Prakash, S. (1984) Gene, 30, 121-128.
- Jachymczyk, W.J., von Borstel, R.C., Mowat, M.R.A. and Hastings, P.J. (1981) Mol. Gen. Genet., 182, 196-205.
- Kozak, M. (1981) Nucleic Acids Res., 9, 5233-5252.
- Kozak, M. (1984) Nucleic Acids Res., 12, 857-872.
- Lipman, D.J. and Pearson, W.R. (1985) Science (Wash.), 227, 1435-1441.
- Magaña-Schwenke, N., Henriques, J.A.P., Chanet, R. and Moustacchi, E. (1982) Proc. Natl. Acadd. Sci. USA, **79**, 1722-1726.
- Mathis, D.J. and Chambon, P. (1981) Nature, 290, 310-315.
- McPherson, A., Jurnak, F.A., Wang, A.J., Molineux, I. and Rich, A. (1979) J. Mol. Biol., 134, 379-400.
- Miller, R.D., Prakash, L. and Prakash, S. (1982a) Mol. Cell. Biol., 2, 939-948.
- Miller, R.D., Prakash, L. and Prakash, S. (1982b) Mol. Gen. Genet., 188, 235-239.
- Nagpal, M.L., Higgins, D.R. and Prakash, S. (1985) Mol. Gen. Genet., 199, 59-63.
- Naumovski, L. and Friedberg, E.C. (1983) Proc. Natl. Acad. Sci. USA, 80, 4818-4821.
- Naumovski, L. and Friedberg, E.C. (1984) Mol. Cell. Biol., 4, 290-295.
- Norrander, J., Kempe, T. and Messing, J. (1983) Gene, 26, 101-106.
- Prakash,L., Dumais,D., Polakowska,R., Perozzi,G. and Prakash,S. (1985) Gene, 34, 55-61.
- Reynolds, R.J. and Friedberg, E.C. (1981) J. Bacteriol., 146, 692-704.
- Reynolds, P., Higgins, D.R., Prakash, L. and Prakash, S. (1985) Nucleic Acids Res., 13, 2357-2372.
- Rothstein, R.J. (1983) Methods Enzymol., 101, 202-211.
- Rubtsov, P.M., Musakhanov, M.M., Zakharyev, V.M., Krayev, A.S., Skyrabin, K.G. and Bayev, A.A. (1980) Nucleic Acids Res., 8, 5779-5794.
- Sanger, F., Nicklen, S. and Coulson, A. (1977) Proc. Natl. Acad. Sci. USA, 74, 5463-5467.
- Weiss, W.A. and Friedberg, E.C. (1985) EMBO J., 4, 1575-1582.
- Wilcox, D.R. and Prakash, L. (1981) J. Bacteriol., 148, 618-623.
- Williams, K.R., LoPresti, M.B. and Setoguchi, M. (1981) J. Biol. Chem., 256, 1754-1762.
- Zaret,K.S. and Sherman,F. (1982) Cell, 28, 563-573.
- Received on 22 August 1985; revised on 24 September 1985