The Yeast *RAD50* Gene Encodes a Predicted 153-kD Protein Containing a Purine Nucleotide-Binding Domain and Two Large Heptad-Repeat Regions

Eric Alani, S. Subbiah and Nancy Kleckner

Department of Biochemistry and Molecular Biology, Harvard University, Cambridge, Massachusetts 02138

Manuscript received December 24, 1988
Accepted for publication February 8, 1989

ABSTRACT

The *RAD50* gene of *Saccharomyces cerevisiae* is required for chromosome synapsis and recombination during meiosis and for repair of DNA damage during vegetative growth. The precise role of the *RAD50* gene product in these processes is not known. Most *rad50* mutant phenotypes can be explained by the proposal that the *RAD50* gene product is involved in the search for homology between interacting DNA molecules or chromosomes, but there is no direct evidence for this model. We present here the nucleotide sequence of the *RAD50* locus and an analysis of the predicted 153-kD RAD50 protein. The amino terminal region of the predicted protein contains residues suggestive of a purine nucleotide binding domain, most likely for adenine. The remaining 1170 amino acids consist of two 250 amino acid segments of heptad repeat sequence separated by 320 amino acids, plus a short hydrophobic carboxy-terminal tail. Heptad repeats occur in proteins such as myosin and intermediate filaments that form α-helical coiled coils. One of the two heptad regions in RAD50 shows similarity to the S-2 domain of rabbit myosin beyond that expected for two random coiled coil proteins.

THE *RAD50* gene of *Saccharomyces cerevisiae* plays an important role in both production of viable spores in meiosis and in the repair of DNA damage during vegetative growth (HAYNES and KUNZ 1981; GAME *et al.* 1980).

During meiosis, *RAD50* function is required for two interrelated events of meiotic prophase, chromosome synapsis and recombination, both of which are required for proper segregation of chromosomes at the first meiotic division (BAKER *et al.* 1976). In *rad50* mutants, the defect in chromosome synapsis is manifested cytologically by the absence of mature synaptonemal complexes (SC) (BYERS, as cited in MALONE, JORDAN and WARDMAN 1985). More specifically, axial elements form between each pair of sister chromatids but never come together to form mature tripartite SCs (asynapsis) (FARNET *et al.* 1988; R. PADMORE and N. KLECKNER, unpublished data).

The *rad50* defect in meiotic recombination is absolute and involves a block at a very early step in the recombination process as revealed by both phenotypic and epistasis analysis (MALONE and ESPOSITO 1981; MALONE 1983; MONTELONE, HOEKSTRA and MALONE 1988). *RAD50* appears to act at a very early time in meiosis; it is required for the appearance of a specific meiotic signal which occurs just at the conclusion of premeiotic DNA replication, and which manifests itself as a site specific double strand break or pair of double strand ends (L. CAO, E. ALANI and N. KLECKNER, unpublished data). The *rad50* defects in meiotic recombination and synaptonemal complex formation are sufficient to account for the fact that fewer than one percent of the spores from a *rad50* meiosis are viable (GAME *et al.* 1980).

During vegetative growth, *RAD50* function is required for repair of some types of DNA damage, including double strand breaks induced by ionizing radiation or the radiomimetic alkylating agent methylmethane sulfonate (MMS) (GAME and MORTIMER 1974; STRIKE, cited in HAYNES and KUNZ 1981; J. NITISS and M. RESNICK, personal communication). *rad50* mutants also exhibit a delay in the spindle elongation step of normal vegetative growth, a phenotype which suggests that they are defective in repair of spontaneous DNA damage as well (FARNET *et al.* 1988; MONTELONE, HOEKSTRA and MALONE 1988). *RAD50*-dependent DNA repair is thought to occur by a recombinational pathway (HAYNES and KUNZ 1981).

Both chromosome synapsis (SC formation) and meiotic recombination require that homologous chromosomes or DNA molecules find one another. Some models propose that synapsis is required for recombination; other models propose more complex relationships. An economical hypothesis would be that a single homology search is used for both processes. We have proposed that *RAD50* is required for such a homology search and that the *RAD50* dependent meiotic signal is a manifestation of this search (FARNET *et al.* 1988; L. CAO, E. ALANI and N. KLECKNER, unpublished data).

The relationship between meiotic and mitotic phenotypes of *rad50* mutants is also not established. However, the involvement of *RAD50* in DNA repair makes it seem more likely that it is directly involved in

meiotic chromosome metabolism rather than playing an indirect and/or regulatory role. The role proposed for *RAD50* in meiosis could also account for its role in DNA repair, since a recombinational repair pathway would also require recognition and interaction of homologous molecules. Arguments that *RAD50* is required for recognition of homologous DNA segments have also been put forward by J. NITISS and M. RESNICK (personal communication).

To further probe the nature of the *RAD50* gene product, we have carried out DNA sequence analysis of a previously cloned *RAD50* gene. This analysis, described below, has identified a 1312 amino acid open reading frame which corresponds to the genetically defined *RAD50* gene. The predicted 153-kD protein has three interesting features: (1) The amino terminal region of the predicted protein contains residues diagnostic of a purine nucleotide binding domain (ROSSMAN, MORAS and OLSEN 1974), probably for adenine (FRY, KUBY and MILDVAN 1986). (2) Two regions, one at amino acids 177–421 and another at amino acids 743–995, are segments of heptad repeat sequence characteristic of alpha helical coiled coil proteins such as myosin and intermediate filaments (MCLACHLAN and KARN 1982; STEINERT, STEVENS and ROOP 1985). (3) Heptad 743–995 also shows a significant similarity to the S2 domain of rabbit myosin (CAPONY and ELZINGA 1981).

MATERIALS AND METHODS

Strains: *Escherichia coli* stain MM294 (F^-, *end*A, *hsd*R, *sup*E44, *thi*A) was used for all plasmid manipulations (GUARENTE *et al.* 1980). JM101 (Δ*lac*, *thi*, *sup*E/F, *tra*D36, *pro*A, *pro*B, *laci*^{q}, *lacZ*Δ, *M*15) was used as a host to prepare single stranded DNA for dideoxy sequencing (MESSING 1983). *Saccharomyces cerevisiae* strain NKY551 is a diploid that contains a complete coding region deletion of *RAD50* (*Hin*dIII-*Sal*I deletion, Figure 1). NKY551 (**a**/α, *rad50*Δ::*hisG*/*rad50*Δ::*hisG*, *ura3*/*ura3*, *lys2*/*lys2*, *ho*::*LYS2*/*ho*::*Lys2*) was constructed by transforming NKY278 (**a**/α, *ura3*/*ura3*, *lys2*/*lys2*, *ho*::*LYS2*/*ho*::*LYS2* to Ura^+ with *Bgl*II and *Eco*RI digested pNKY83 (described below). The integration of this fragment resulted in a complete coding region deletion of *RAD50* (blot not shown). One such transformant, NKY492, was plated onto 5-fluoro-orotic (5-FOA) acid plates and a Ura^- strain, NKY546, was selected (ALANI, CAO and KLECKNER 1987). Tetrad dissection of NKY546 yielded spores which segregated 2:2, MMS^r, wild-type growth: MMS^s, slow growth (FARNET *et al.* 1988). Two *rad50*Δ haploid segregants from NKY546 were mated to form the homozygous *rad50*Δ strain NKY551. This strain sporulated at high frequency (50%) but yielded inviable spores.

Media: *E. coli* was grown in LB broth (MILLER 1972). Ampicillin was supplemented at 25 μg/ml. Yeast was grown in either YPD or minimal medium (SHERMAN, FINK and HICKS 1983). Minimal medium contained 0.7%; yeast nitrogen base, 2% agar, 2% glucose and 0.004% uracil. 5-FOA was purchased from SCM specialty chemical, Gainesville, Florida. 5-FOA plates were prepared as described previously (BOEKE, LACROUTE and FINK 1984).

Yeast transformations: All transformations were performed using standard methods (ITO *et al.* 1983).

Nucleic acid techniques: All restriction enzymes, T4 DNA ligase, T4 DNA polymerase and *BAL*31 were purchased from New England Biolabs and used according to the manufacturer's specifications. Plasmid DNA was isolated by a cleared lysate protocol (MANIATIS, FRITSCH and SAMBROOK 1982). DNA manipulations including isolation of single stranded DNA from M-13 origin plasmids was described previously. Chromosomal yeast DNA isolation and Southern blotting were described previously (HOLM *et al.* 1986; SOUTHERN 1975; MELTON *et al.* 1984).

Plasmid constructions: All plasmids were derived from the following vectors (relevant genotype and/or description): YEP24 (*URA3*, 2μ; ROSE, GRISAFI and BOTSTEIN 1984; New England Biolabs Catalog 1986), YCP50 (*URA3*, *ARS1*, *CEN4*; STINCHCOMB, MANN and DAVIS 1982) pSG205 (7-kb *RAD50* *Hin*dIII fragment inserted into YCP50, from S. GOTTLIEB and R. ESPOSITO), pOL13 (pSG205 derivative with *RAD50* in opposite orientation), and M-13 sequencing vectors pGC-1, pGC-2, pOL10 (pGC-1 with the *Hin*dIII site destroyed) and pNKY47 (pOL10 with a *Hin*dIII linker at *Bam*HI site) (MYERS and MANIATIS 1985).

pNKY100 is a vector that was used to clone chromosomal sequences upstream of the 5′ *Hin*dIII site in *RAD50*. A 4.0-kb *Hin*dIII-*Sal*I subclone from pSG205 was inserted into the *Hin*dIII-*Sal*I backbone of pBR322 to make pNKY19. A *Hin*dIII *URA3* subclone isolated from YEP24 was inserted into the *Hin*dIII site of pNKY19 to form pNKY100. In pNKY100 *URA3* transcription is in the same direction as *RAD50* and *Eco*RI cuts only once in the plasmid, between *amp*′ and *URA3*.

pNKY101 is a vector bearing the *RAD50* gene and its upstream sequences. pNKY100 was linearized within the *RAD50* gene by an *Nsi*I restriction digest and then integrated into the *RAD50* locus of NKY501 (α, *his4*-713, *lys2*-20, *suf10*, *ura3*, Ura^- derivative of RE821, kindly supplied by R. ESPOSITO) by lithium acetate transformation (blot not shown). Chromosomal DNA from the integrant was isolated, digested with *Eco*RI, ligated and transformed into MM294. An *amp*′ plasmid bearing the *Hin*dIII-*Sal*I *RAD50* fragment plus 2.8 kb of additional upstream *RAD50* sequence was isolated and designated pNKY101.

pNKY74 is a pGC derived vector (MYERS and MANIATIS 1985) that contains an 8.8-kb *RAD50* clone. *RAD50* DNA upstream of the coding region was isolated on a 2 kb *Bgl*II-*Hin*dIII fragment from pNKY101 and inserted into the *Bgl*II and *Hin*dIII backbone of pNKY47 to form pNKY73. The *RAD50* locus, containing the coding region plus both 5′ and 3′ flanking sequence, was reconstructed by inserting the 7-kb *Hin*dIII *RAD50* fragment from pSG205 into pNKY73 to form the integrating vector pNKY74.

pNKY1070 is an *ARS1*, *CEN4* vector that contains 6.6 kb of *RAD50* sequence. pNKY74 was digested with *Eco*RI and a 6.6-kb *RAD50* fragment was inserted into the *Eco*RI site of YCP50.

pNKY83 was used to delete the *Hin*d-*Sal*I *RAD50* sequences from the chromosome in order to create complete chromosomal coding region deletions of *RAD50*. The 6.6-kb *RAD50* *Eco*RI fragment from pNKY74 was inserted into the *Eco*RI site of a pBR322 derived vector (*Pvu*II-*Eco*RI deletion) to form pNKY79. pNKY79 was digested with *Hin*dIII and *Sal*I (complete *RAD50* coding region deletion, filled in with T4 DNA polymerase and religated in the presence of *Bam*HI 8-bp linkers. Finally, a 3.8-kb *Bgl*II-*Bam*HI *hisG*-*URA3*-*hisG* fragment (ALANI, CAO and KLECKNER 1987) was inserted into the *Bam*HI linker site to form pNKY83.

pNKY5 and **pNKY6 are** *RAD50* vectors that were used

to sequence both strands of *RAD50* (strands A and B, respectively). pOL13 was digested with *Xho*I and *Eco*RI. A 9.7-kb *RAD50* fragment was inserted into the *Xho*I and *Eco*RI sites of pOL10 to form pNKY5. The 7-kb *Hin*dIII fragment from pSG205 was blunt ended with T4 DNA polymerase and inserted into the blunt ended *Sal*I site of pGC-2 to form pNKY6 (*Hin*dIII site is regenerated).

Dideoxy sequencing: A 4.0-kb *Hin*dIII-*Sal*I *RAD50* subclone that was shown to complement the *rad50-1* mutation was sequenced by the dideoxy approach (MESSING 1983). Two M13 origin vectors, pNKY5 and 6, which bear the *Hin*dIII-*Sal*I fragment in opposite orientations, were used as substrates for *BAL*31 deletion. The end points of 24 deletions in pNKY5 and 25 deletions in pNKY6 were sequenced (Figure 3). DNA sequences that were not uncovered by *BAL31* deletion were sequenced by cloning restriction fragments into appropriate M13 vectors. Using this approach the entire *Hin*dIII-*Sal*I fragment was sequenced on both strands. Sequences downstream of *Sal*I were sequenced on only one strand (pNKY5). Transcription studies indicated that the *RAD50* gene is transcribed from *Hin*dIII to *Sal*I (FARNET *et al.* 1988; W. RAYMOND and N. KLECKNER, unpublished data). Because aberrant *RAD50* transcripts are observed on *ARS CEN* vectors that contain the *Hin*dIII-*Sal*I clone, sequences upstream of the *Hin*dIII site were cloned (pNKY101). Upstream from the *Hin*dIII site 600 bp of DNA were sequenced on both strands by cloning restriction fragments into pGC-1 and pGC-2.

Computer analysis: All computer analysis was performed on a Digital VAX 8700 computer. *RAD50* amino acid sequence was compared to the NBRF protein data bank, release #15 (4931 sequences, 1,264,388 residues) using the FASTP algorithm described by LIPMAN and PEARSON (1985). Fourier analysis was performed by a method described by MCLACHLAN and KARN (1983). Heptad shuffle analysis using NEEDLEMAN-WUNSCH matrix analysis was described by S. SUBBIAH (unpublished data). The SIM matrix (similarity cutoff = 5 SD, gap penalties = 18 or 24 (MCLACHLAN 1971)) and UP matrix (similarity cutoff = 3 SD, gap penalties = 3 or 5 (DAYHOFF 1978)) were used in this analysis and the following *Z* values (LIPMAN and PEARSON 1985) were calculated using these matrices: *Z* regular = (similarity score − mean similarity score of scrambled sequences)/(standard deviation of scrambled similarity scores) (NEEDLEMAN and WUNSCH 1970). *Z* heptad = (similarity score − mean similarity score of heptad shuffled sequences)/(standard deviation of heptad shuffled similarity scores) (S. SUBBIAH, unpublished data). *Z* control = *Z* heptad analysis of aligned protein sequences scrambled once in register (S. SUBBIAH, unpublished data).

RESULTS AND DISCUSSION

A 9.5-kb region which includes the *RAD50* gene is diagrammed in Figure 1. RAD50 function is contained within a four kilobase region defined by *Hin*dIII and *Sal*I restriction sites; this segment complements the mitotic and meiotic defects of both point and deletion mutants in *RAD50* when present on an autonomously replicating yeast plasmid (*ARS1*, *CEN4*) (KUPIEC and SIMCHEN 1984; S. GOTTLIEB and R. E. ESPOSITO, personal communication; see MATERIALS AND METHODS). Essentially all of the sequences in this 4-kb segment are required for complementation; Tn*10*-LUK insertion mutations that abolish such complementation span the entire region (HUISMAN *et al.* 1987).

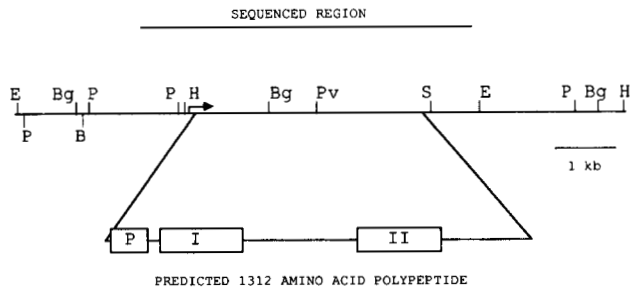


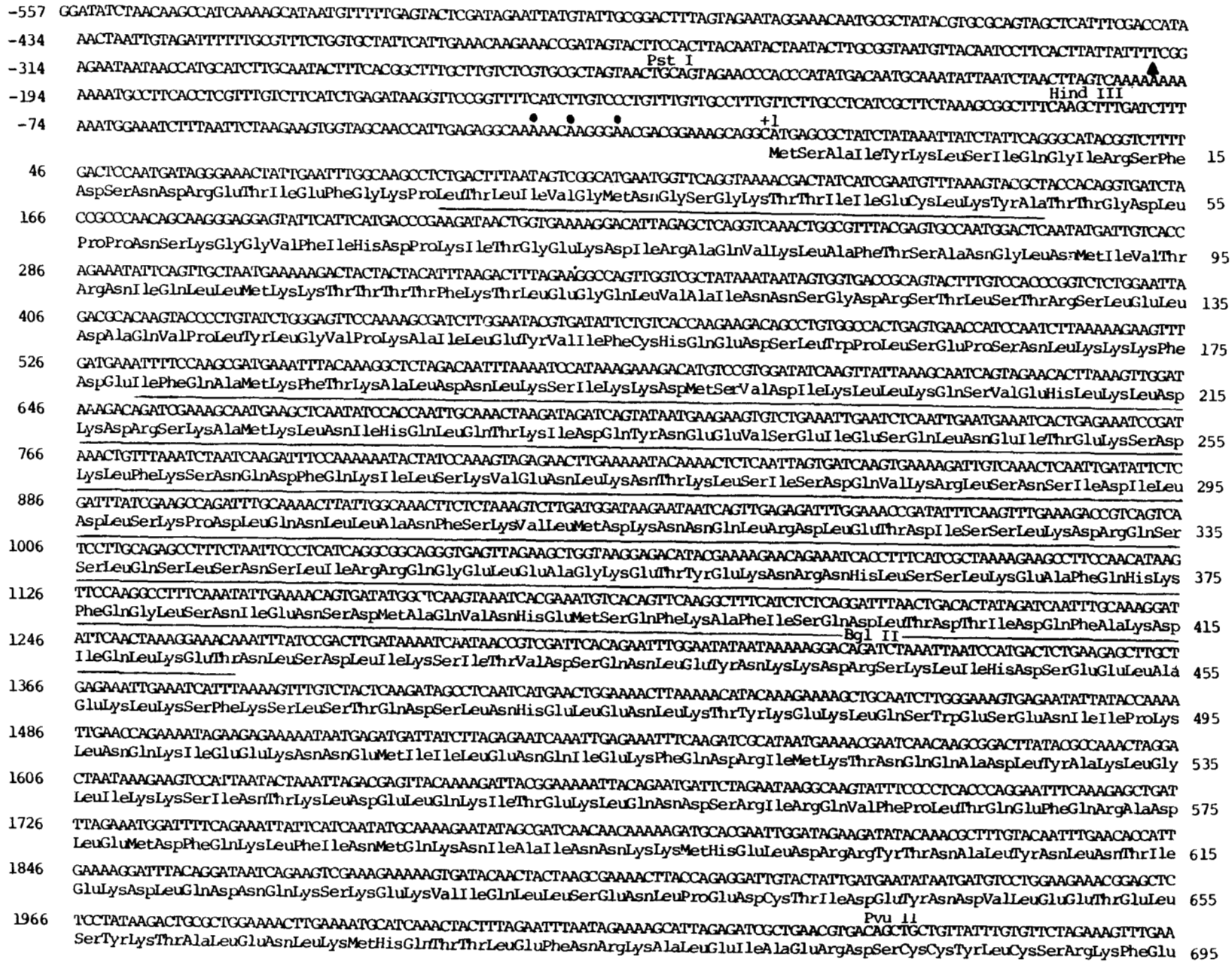
FIGURE 1.—The 9.5-kb *RAD50* locus, 4.775 kb of which was sequenced. The point of initiation and the direction of transcription is indicated by the arrow. E = *Eco*RI, P = *Pst*I, Bg = *Bgl*II, B = *Bam*HI, H = *Hin*dIII, Pv = *Pvu*II, S = *Sal*I. P = purine nucleotide binding domain similarity; I = heptad repeat I (amino acids 177–421); II = heptad repeat II (amino acids (743–995).

Additional information outside of this minimal segment is required for normal expression of the *RAD50* gene. When expressed as a *Hin*dIII to *Sal*I fragment on an *ARS CEN* vector, both the size and quantity of the *RAD50* transcripts differ from that observed for a chromosomal *RAD50* gene (W. RAYMOND and N. KLECKNER, unpublished data). However, a normal chromosomal transcription pattern is obtained with a larger 6.6-kb *Bgl*II to *Eco*RI *RAD50* subclone (W. RAYMOND and N. KLECKNER, unpublished data).

To define the *RAD50* coding region in more detail, the DNA sequence was determined for a 4775-bp region that includes the *Hin*dIII-*Sal*I fragment (Figure 2). This analysis, described in Figure 3 and MATERIALS AND METHODS, reveals a continuous 1312 amino acid open reading frame (ORF) that is fully contained in the minimal *RAD50 Hin*dIII-*Sal*I segment (Figure 2). The corresponding protein begins with a methionine encoded by bp 558–560 and terminates with a stop codon at bp 4492–4494 and has a predicted M_r of 153 kD. Immunoblot analysis of yeast proteins has subsequently identified a single polypeptide of the expected M_r which has been correlated genetically with *RAD50* function (W. RAYMOND and N. KLECKNER unpublished data).

The RAD50 protein predicted from the DNA sequence is composed of 17.5% acidic, 15.9% basic, 30.9% hydrophobic and 36.0% polar residues; nine glycosylation consensus sequences (NXS, $X \neq P$) are present. KYTE-DOOLITTLE (1983) analysis suggests that the protein is predominantly hydrophilic; the longest hydrophobic region is a 35 amino acid segment near the carboxy terminus of the protein (amino acids 1205–1232, corresponding to bp 4170–4256 in Figure 2) which contains 20 hydrophobic residues.

Nucleotide binding domain: Initial inspection of the RAD50 amino acid sequence reveals a short glycine-rich sequence that has previously been found in the sequences of many adenine and guanine nucleotide binding proteins (Figure 4; FRY, KUBY and MILD-


-557 GGATATCTAACAAGCCATCAAAAGCATAATGTTTTTGAGTACTCGATAGAATTATGTATTGCGGACTTTAGTAGAATAGGAAACAATGCGCTATACGTGCGCAGTAGCTCATTTCGACCATA
-434 AACTAATTGTAGATTTTTTGCGTTTCTGGTGCTATTCATTGAAACAAGAAACCGATAGTACTTCCACTTACAATACTAATACTTGCGGTAATGTTACAATCCTTCACTTATTATTTTCGG
Pst I
-314 AGAATAATAACCATGCATCTTGCAATACTTTCACGGCTTTGCTTGTCTCGTGCGCTAGTAACTGCAGTAGAACCCACCCATATGACAATGCAAATATTAATCTAACTTAGTCAAAAAAAA
-194 AAAATGCCTTCACCTCGTTTGTCTTCATCTGAGATAAGGTTCCGGTTTTCATCTTGTCCCTGTTTGTTGCCTTTGTTCTTGCCTCATCGCTTCTAAAGCGGCTTTCAAGCTTTGATCTTT
Hind III
-74 AAATGGAAATCTTTAATTCTAAGAAGTGGTAGCAACCATTGAGAGGCAAAAACAAGGGAACGACGGAAAGCAGGCATGAGCGCTATCTATAAATTATCTATTCAGGGCATACGGTCTTTT
+1
MetSerAlaIleTyrLysLeuSerIleGlnGlyIleArgSerPhe 15
46 GACTCCAATGATAGGGAAACTATTGAATTTGGCAAGCCTCTGACTTTAATAGTCGGCATGAATGGTTCAGGTAAAACGACTATCATCGAATGTTTAAAGTACGCTACCACAGGTGATCTA
AspSerAsnAspArgGluThrIleGluPheGlyLysProLeuThrLeuIleValGlyMetAsnGlySerGlyLysThrThrIleIleGluCysLeuLysTyrAlaThrThrGlyAspLeu 55
166 CCGCCCAACAGCAAGGGAGGAGTATTCATTCATGACCCGAAGATAACTGGTGAAAAGGACATTAGAGCTCAGGTCAAACTGGCGTTTACGAGTGCCAATGGACTCAATATGATTGTCACC
ProProAsnSerLysGlyGlyValPheIleHisAspProLysIleThrGlyGluLysAspIleArgAlaGlnValLysLeuAlaPheThrSerAlaAsnGlyLeuAsnMetIleValThr 95
286 AGAAATATTCAGTTGCTAATGAAAAAGACTACTACTACATTTAAGACTTTAGAAGGCCAGTTGGTCGCTATAAATAATAGTGGTGACCGCAGTACTTTGTCCACCCGGTCTCTGGAATTA
ArgAsnIleGlnLeuLeuMetLysLysThrThrThrThrPheLysThrLeuGluGlyGlnLeuValAlaIleAsnAsnSerGlyAspArgSerThrLeuSerThrArgSerLeuGluLeu 135
406 GACGCACAAGTACCCCTGTATCTGGGAGTTCCAAAAGCGATCTTGGAATACGTGATATTCTGTCACCAAGAAGACAGCCTGTGGCCACTGAGTGAACCATCCAATCTTAAAAAGAAGTTT
AspAlaGlnValProLeuTyrLeuGlyValProLysAlaIleLeuGluTyrValIlePheCysHisGlnGluAspSerLeuTrpProLeuSerGluProSerAsnLeuLysLysLysPhe 175
526 GATGAAATTTTCCAAGCGATGAAATTTACAAAGGCTCTAGACAATTTAAAATCCATAAAGAAAGACATGTCCGTGGATATCAAGTTATTAAAGCAATCAGTAGAACACTTAAAGTTGGAT
AspGluIlePheGlnAlaMetLysPheThrLysAlaLeuAspAsnLeuLysSerIleLysLysAspMetSerValAspIleLysLeuLeuLysGlnSerValGluHisLeuLysLeuAsp 215
646 AAAGACAGATCGAAAGCAATGAAGCTCAATATCCACCAATTGCAAACTAAGATAGATCAGTATAATGAAGAAGTGTCTGAAATTGAATCTCAATTGAATGAAATCACTGAGAAATCCGAT
LysAspArgSerLysAlaMetLysLeuAsnIleHisGlnLeuGlnThrLysIleAspGlnTyrAsnGluGluValSerGluIleGluSerGlnLeuAsnGluIleThrGluLysSerAsp 255
766 AAACTGTTTAAATCTAATCAAGATTTCCAAAAAATACTATCCAAAGTAGAGAACTTGAAAAATACAAAACTCTCAATTAGTGATCAAGTGAAAAGATTGTCAAACTCAATTGATATTCTC
LysLeuPheLysSerAsnGlnAspPheGlnLysIleLeuSerLysValGluAsnLeuLysAsnThrLysLeuSerIleSerAspGlnValLysArgLeuSerAsnSerIleAspIleLeu 295
886 GATTTATCGAAGCCAGATTTGCAAAACTTATTGGCAAACTTCTCTAAAGTCTTGATGGATAAGAATAATCAGTTGAGAGATTTGGAAACCGATATTTCAAGTTTGAAAGACCGTCAGTCA
AspLeuSerLysProAspLeuGlnAsnLeuLeuAlaAsnPheSerLysValLeuMetAspLysAsnAsnGlnLeuArgAspLeuGluThrAspIleSerSerLeuLysAspArgGlnSer 335
1006 TCCTTGCAGAGCCTTTCTAATTCCCTCATCAGGCGGCAGGGTGAGTTAGAAGCTGGTAAGGAGACATACGAAAAGAACAGAAATCACCTTTCATCGCTAAAAGAAGCCTTCCAACATAAG
SerLeuGlnSerLeuSerAsnSerLeuIleArgArgGlnGlyGluLeuGluAlaGlyLysGluThrTyrGluLysAsnArgAsnHisLeuSerSerLeuLysGluAlaPheGlnHisLys 375
1126 TTCCAAGGCCTTTCAAATATTGAAAACAGTGATATGGCTCAAGTAAATCACGAAATGTCACAGTTCAAGGCTTTCATCTCTCAGGATTTAACTGACACTATAGATCAATTTGCAAAGGAT
PheGlnGlyLeuSerAsnIleGluAsnSerAspMetAlaGlnValAsnHisGluMetSerGlnPheLysAlaPheIleSerGlnAspLeuThrAspThrIleAspGlnPheAlaLysAsp 415
Bgl II
1246 ATTCAACTAAAGGAAACAAATTTATCCGACTTGATAAAATCAATAACCGTCGATTCACAGAATTTGGAATATAATAAAAAGGACAGATCTAAATTAATCCATGACTCTGAAGAGCTTGCT
IleGlnLeuLysGluThrAsnLeuSerAspLeuIleLysSerIleThrValAspSerGlnAsnLeuGluTyrAsnLysLysAspArgSerLysLeuIleHisAspSerGluGluLeuAla 455
1366 GAGAAATTGAAATCATTTAAAAGTTTGTCTACTCAAGATAGCCTCAATCATGAACTGGAAAACTTAAAAACATACAAAGAAAAGCTGCAATCTTGGGAAAGTGAGAATATTATACCAAAA
GluLysLeuLysSerPheLysSerLeuSerThrGlnAspSerLeuAsnHisGluLeuGluAsnLeuLysThrTyrLysGluLysLeuGlnSerTrpGluSerGluAsnIleIleProLys 495
1486 TTGAACCAGAAAATAGAAGAGAAAAATAATGAGATGATTATCTTAGAGAATCAAATTGAGAAATTTCAAGATCGCATAATGAAAACGAATCAACAAGCGGACTTATACGCCAAACTAGGA
LeuAsnGlnLysIleGluGluLysAsnAsnGluMetIleIleLeuGluAsnGlnIleGluLysPheGlnAspArgIleMetLysThrAsnGlnGlnAlaAspLeuTyrAlaLysLeuGly 535
1606 CTAATAAAGAAGTCCATTAATACTAAATTAGACGAGTTACAAAAGATTACGGAAAAATTACAGAATGATTCTAGAATAAGGCAAGTATTTCCCCTCACCCAGGAATTTCAAAGAGCTGAT
LeuIleLysLysSerIleAsnThrLysLeuAspGluLeuGlnLysIleThrGluLysLeuGlnAsnAspSerArgIleArgGlnValPheProLeuThrGlnGluPheGlnArgAlaAsp 575
1726 TTAGAAATGGATTTTCAGAAATTATTCATCAATATGCAAAAGAATATAGCGATCAACAACAAAAAGATGCACGAATTGGATAGAAGATATACAAACGCTTTGTACAATTTGAACACCATT
LeuGluMetAspPheGlnLysLeuPheIleAsnMetGlnLysAsnIleAlaIleAsnAsnLysLysMetHisGluLeuAspArgArgTyrThrAsnAlaLeuTyrAsnLeuAsnThrIle 615
1846 GAAAAGGATTTACAGGATAATCAGAAGTCGAAAGAAAAAGTGATACAACTACTAAGCGAAAACTTACCAGAGGATTGTACTATTGATGAATATAATGATGTCCTGGAAGAAACGGAGCTC
GluLysAspLeuGlnAspAsnGlnLysSerLysGluLysValIleGlnLeuLeuSerGluAsnLeuProGluAspCysThrIleAspGluTyrAsnAspValLeuGluGluThrGluLeu 655
Pvu II
1966 TCCTATAAGACTGCGCTGGAAAACTTGAAAATGCATCAAACTACTTTAGAATTTAATAGAAAAGCATTAGAGATCGCTGAACGTGACAGCTGCTGTTATTTGTGTTCTAGAAAGTTTGAA
SerTyrLysThrAlaLeuGluAsnLeuLysMetHisGlnThrThrLeuGluPheAsnArgLysAlaLeuGluIleAlaGluArgAspSerCysCysTyrLeuCysSerArgLysPheGlu 695

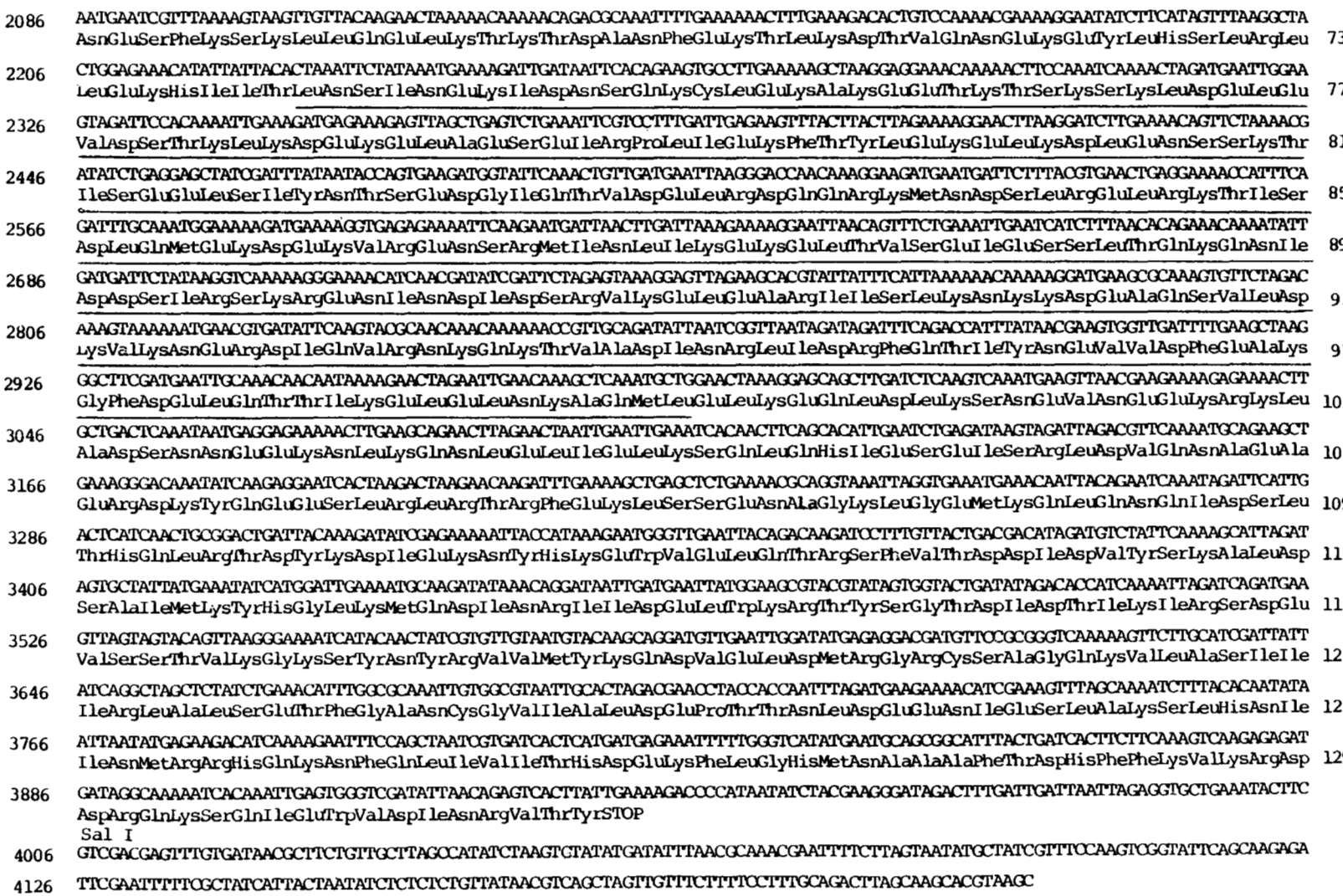


FIGURE 2.—The 4775-bp DNA sequence and predicted 1312 amino acid sequence of the *RAD50* region from NKY501. Coordinates of DNA and protein sequences are indicated on the left and right, respectively. The open reading frame begins with a methionine at bp 1 and ends with a TGA stop codon at bp 4307. The next in frame methionine is at bp 103. Major transcription start points (−26, −22, −17), determined by S1 nuclease and primer extension analysis (W. RAYMOND and N. KLECKNER, unpublished results) are indicated by the filled circles. A 12-bp poly(dA-dT) track, similar to that implicated in constitutive basal expression of *HIS3* (STRUHL 1985), is found at −202 to −191 and is indicated by the filled triangle. Similarities to purine nucleotide binding domains are found at amino acids 29–50 (FRY, KUBY and MILDVAN 1986). Heptad repeat I and heptad repeat II are found at 177–421 and 743–995, respectively. The *RAD50* gene from a different strain, SK-1, was cloned and sequenced from bp −449 to 215 (E. ALANI and N. KLECKNER, unpublished results). The following changes were observed: bp −320 T to A, bp −317 C to T, −166 C to G and −135 C to T.

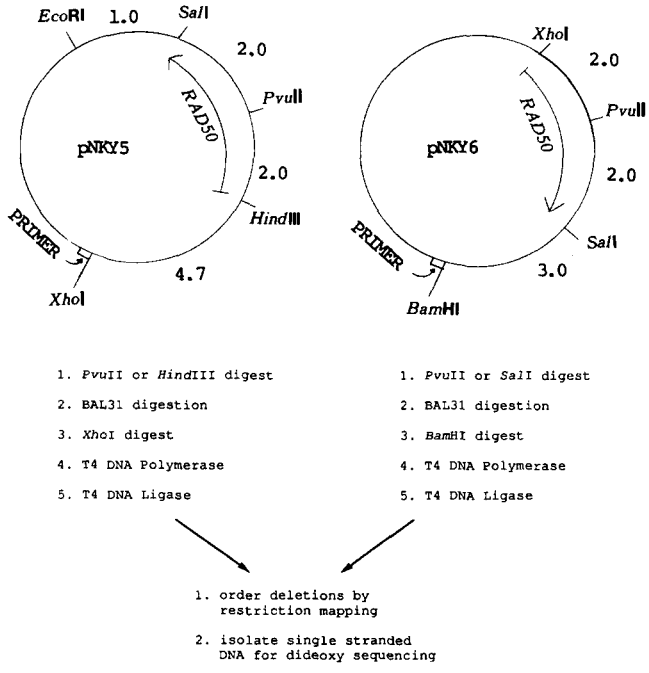


FIGURE 3.—*RAD50* sequencing strategy. pNKY5 and pNKY6 contain *Col*E1 and M13 origins of replication, an ampicillin resistance determinant, a site used for binding of an oligonucleotide primer for DNA sequencing, and the *RAD50* gene, inserted in opposite orientations in the two plasmids. Sequence of both strands of the *RAD50* *Hind*III-*Sal*I segment was obtained as described in this figure and in the MATERIALS AND METHODS.

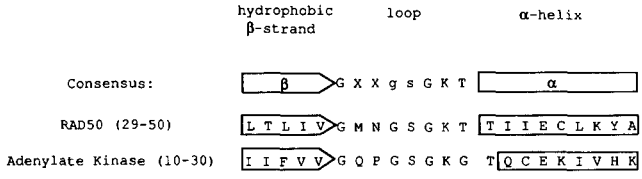


FIGURE 4.—RAD50 protein shows similarity to purine nucleotide binding proteins. Structural analysis of adenylate kinase and sequence comparisons between ATP and GTP binding proteins has led to a consensus sequence for purine nucleotide binding pockets (FRY, KUBY and MILDVAN 1986; ROSSMAN, MORAS and OLSEN 1974; BRADLEY *et al.* 1987). This pocket is composed of a hydrophobic β-strand and an α-helix which flank a glycine rich flexible loop (top line). The corresponding structures in adenylate kinase, based on crystal structure and RAD50, based on CHOU-FASMAN analysis, are shown (FRY, KUBY and MILDVAN 1986; CHOU and FASMAN 1978; RALPH, WEBSTER and SMITH 1987). In the consensus sequence, upper case and lower case letters indicate strongly and less strongly conserved residues, respectively, and X indicates nonconserved residues.

VAN, 1986; WALKER *et al.* 1982; DEVER, GLYNIAS and MERRICK 1987). In adenylate kinase, this sequence specifies a flexible loop whose terminal lysine residue appears to interact with the α-phosphoryl group of Mg-ATP (FRY, KUBY and MILDVAN 1986; DREUSICKE and SCHULZ 1986). In a number of ATP and GTP binding proteins, this loop is suggested to mediate conformational changes in the protein brought about in some cases by interactions with additional proteins (FRY, KUBY and MILDVAN 1986; REINSTEIN, BRUNE and WITTINGHOFER 1988).

In addition, RAD50 protein is predicted by ARIADNE, a computer matching program that utilizes a descriptor pattern for the ATP binding motif (LATHROP, WEBSTER and SMITH 1987; RALPH, WEBSTER and SMITH 1987), and CHOU-FASMAN (1978) analysis to have specific secondary structure features common to the ATP binding pockets of many other nucleotide binding proteins. These include a hydrophobic β-strand and an α-helix which flank the flexible loop (ROSSMAN, MORAS and OLSEN 1974; BRADLEY *et al.* 1987; Figure 4). Preliminary analysis reveals several reasonable candidates for a magnesium coordination site (BRADLEY *et al.* 1987; T. WEBSTER and M. BRADLEY, personal communication).

RAD50 protein probably binds an adenine nucleotide. It lacks two consensus sequence features that are specific to the known set of closely related guanine nucleotide binding proteins (DEVER, GLYNIAS and MERRICK 1987).

ATP binding is a common feature of DNA repair and recombination proteins from *E. coli* (RecA, RecBCD, RecN, UvrABC, UvrD), *S. typhimurium* (MutS) and *S. cerevisiae* (RAD3) (SMITH 1988 for review, BACKENDORF *et al.* 1986; HABER *et al.* 1988; NAUMOVSKI and FRIEDBERG 1986). However, in the event that RAD50 is involved structurally rather than or in addition to having an enzymatic role in recombination, the role of ATP in RAD50 function may be

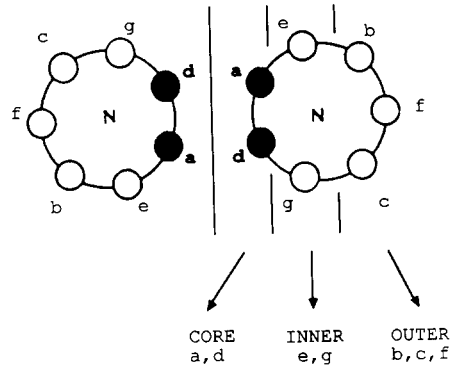


FIGURE 5.—Cross section of α-helical coiled coil when viewed from the amino end of parallel helices (N). Filled circles indicate core hydrophobic residues. (Adapted from MCLACHLAN and KARN 1983.)

different from its role in these other proteins (see below).

Heptad repeat sequences and similarity to myosin S2 domain: Two regions within the RAD50 amino acid sequence, designated heptad I (177–421) and heptad II (743–995) display a heptad repeat pattern found in proteins such as myosin, keratin and tropomyosin (MCLACHLAN and KARN 1982, 1983; COHEN and PARRY 1986). The heptad repeat is a seven amino acid sequence with primarily hydrophobic amino acids at the first and fourth positions and a high percentage of charged amino acids at the other positions. When present in a repeated motif, this unit forms an alpha helical structure (Figure 5). Two heptad-containing α-helices can coil around one another in a regular helical structure that is stabilized by repeated interactions between hydrophobic residues and specific, oppositely charged residues on the two chains (Figure 5) (MCLACHLAN and KARN 1982). The RAD50 heptad repeat regions were identified and analyzed by three separate methods, discussed below. These approaches also reveal a significant similarity between heptad II and the S-2 fragment of rabbit myosin heavy chain.

The RAD50 protein sequence was compared to other protein sequences in the NBRF database using the FASTP computer algorithm of LIPMAN and PEARSON (1985). This analysis provides a similarity score for each such comparison. A significant match is indicated if the difference between the similarity score of a match of interest and the average similarity score is more than 10 SD ($Z \geq 10$). When comparisons are made under conditions were significant gaps in matching sequences were allowed, eight protein sequences show significant homology to RAD50 protein. These include rabbit skeletal muscle myosin (heavy chain S-2 fragment), rabbit cardiac muscle myosin (α and β heavy chains), and several muscle tropomyosins from chicken, rabbit and rat. When stricter penalties are assigned for gaps, only the rabbit S-2 myosin fragment continues to show strong similarity to RAD50 (Z = 12.5). RAD50 protein and the rabbit S-2 myosin fragment are 22% identical over the entire 259 amino acid rabbit myosin fragment (CAPONY and ELZINGA 1981, Figure 6).

The five proteins initially identified by FASTP analysis are all known to contain extensive regions that form α-helical coiled coils. Therefore, a systematic search for heptad repeats in the RAD50 protein was carried out using Fourier analysis. A computer program based on the analysis of MCLACHLAN and KARN (1983) was written to search for repeats of seven residue sequences containing hydrophobic amino acids at the first and fourth positions (MCLACHLAN and KARN 1983). The RAD50 protein sequence was divided into eleven blocks of 120 amino acids each; each block was analyzed individually. This approach eliminates the possibility that small heptad containing regions might be obscured by large non-heptad regions; it is sensitive enough to detect heptad-containing sequences only twenty to thirty amino acids long (three or four heptads).

The Fourier analysis reveals that two large regions of the protein contain significant amounts of heptad repeat sequence: three consecutive blocks of 120 amino acids in the amino-terminal third of the protein and three consecutive 120 amino acid blocks in the carboxy-terminal third of the protein. Visual inspection of these two regions identified by Fourier analysis reveals that each contain a single, virtually uninterrupted series of heptad repeats; both the end points of each region and rare one or two amino acid deletions in the heptad pattern (skip residues) could be unambiguously identified. The amino-terminal region heptad I spans amino acids 177–421; it consists of thirty-five heptad repeats, one of which contains a skip residue. Heptad II spans amino acids 743–995; it consists of thirty-seven heptads, three of which contain one skip residue and one contains two residues. Given the rise of an alpha helix at 1.485 Å per amino acid (FRASER and MACRAE 1973), the length of alpha helical coiled coil structures involving heptad I and heptad II, if present as fully extended rods, would be 364 Å and 385 Å, respectively. However, other coiled coil interactions exist which do not result in extended rod structures (BANNER, KOKKINIDIS and TSERNOGLOU 1987). Interestingly, heptad II and the entire S-2 fragment of rabbit myosin are exactly co-extensive, with 37 heptads each.

The amino acid compositions of the two RAD50 heptad regions and the S-2 domain of rabbit myosin are very similar to those of previously analyzed α-helical coiled coil regions. These three regions and the relevant portions of nematode myosin, keratin and tropomyosin are compared in Table 1 (MCLACHLAN and KARN 1983; PARRY *et al.* 1977; STONE and SMILLIE 1978; KARN, BRENNER and BARNETT

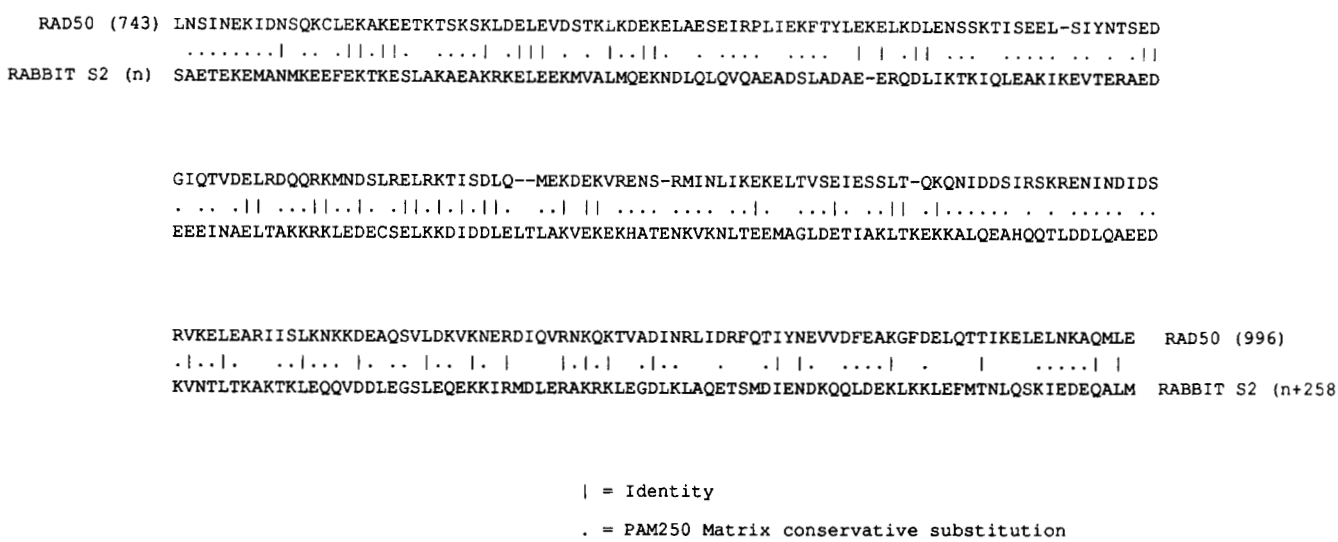


FIGURE 6.—Amino acid alignment of rabbit S2 fragment of myosin (CAPONY and ELZINGA 1981) and amino acids 743–996 of the RAD50 protein. | = amino acid identity, • = Conservative substitution according to PAM 250 matrix (DAYHOFF 1978).

TABLE 1

Amino acid distributions in *RAD50* heptad I and heptad II compared to distributions in coiled coil regions of other proteins

Distribution	Protein (No. of heptads)					
	RAD50 heptad I (35)	*RAD50* heptad II (37)	Rabbit myosin (S2) (37)	Nematode myosin (152)	α-Keratin (29)	Tropomyosin (40)
Core (a, d)						
% hydrophobic	61	63	61	56	65	88
% charged	25	16	23	18	17	8
Inner, outer (b, c, e, f, g)						
% hydrophobic	16	19	8	7	17	17
% charged	34	49	55	53	45	63

Hydrophobic amino acids were: methionine, tyrosine, leucine, tryptophan, phenylalanine, valine and isoleucine; charged amino acids were: lysine, arginine, histidine, aspartic acid and glutamic acid. References for the protein sequences were as follows: rabbit myosin S2 domain (CAPONY and ELZINGA 1981), nematode myosin (KARN, BRENNER and BARNETT 1983), α-keratin (PARRY *et al.* 1977), and tropomyosin (STONE and SMILLIE 1978).

1983). All five regions show the distribution of hydrophobic and charged amino acids characteristic of coiled coil proteins: high percentages of hydrophobic residues at the first and fourth positions of the heptad (56–88%), and at the other positions, low percentages of hydrophobic residues (8–19%) and a high percentage of charged amino acid residues (34–63%) (Figure 5, Table 1).

Significance of similarity with rabbit myosin S-2 domain: The similarity between RAD50 heptad II and the rabbit S-2 myosin fragment (Figure 6) is suggestive. Further analysis indicates the existence of a deeper relationship between the two proteins than expected simply on the basis of their heptad nature. The RAD50 and the rabbit S-2 fragment protein sequences were examined by a computer comparison program designed to eliminate the contributions of the repeating heptad pattern itself. This program, a modified NEEDLEMAN-WUNSCH analysis that incorporates a "heptad shuffle," is described below (Table 2; S. SUBBIAH, unpublished data).

Standard NEEDLEMAN-WUNSCH (1970) analysis allows the optimal alignment of two sequences to yield a similarity score. To determine the significance of that optimal alignment, both sequences are scrambled at the same time and a NEEDLEMAN-WUNSCH similarity score is determined after each scrambling. From this process an average random score and its standard deviation is determined. Then the difference between the similarity score of the optimal alignment and the average random score of the distribution is assessed in terms of standard deviation units. This calculation yields a value called *Z* regular.

In this type of analysis, two heptad-containing proteins should always show greater similarity to one another than to a scrambled sequence in which the constraints of the heptad sequence are removed. This was verified by creating and then comparing two nonidentical amino acid sequences each consisting of 30 "perfect" heptads; *Z* regular for these two sequences is close to (UP matrix, DAYHOFF 1978) or above (SIM matrix, MCLACHLAN 1971) the cutoff for significant

TABLE 2

Detailed comparison of *RAD50* heptad I (HEPI), *RAD50* heptad II (HEPII), nematode myosin S-2 domain (NEM), and rabbit myosin S-2 domain (RAB) protein sequences

	UP matrix analysis		SIM matrix analysis	
Comparison	Z-heptad	Z-regular	Z-heptad	Z-regular
1. HEPII-RAB	4.21	5.25	4.82	8.50
2. HEPII-NEM	2.19	2.93	2.72	4.91
3. RAB-NEM	21.13	25.44	20.78	25.94
4. HEPI-HEPII	1.04	1.37	3.07	4.06
5. HEPI-RAB	1.31	1.77	0.49	2.20
6. HEPI-NEM	−0.83	−0.21	0.99	3.49
7. R1-R2	−1.61	2.76	−1.24	11.36
8. R1-R3	19.54	37.82	13.28	35.45

Z-regular and Z-heptad calculations based on the UP and SIM matrices are described in the MATERIALS AND METHODS. Underlined sequences indicate similarity at or above the cutoff values for the UP (cutoff = 3) or SIM (cutoff = 5) comparisons. R1 and R2 are unrelated synthetic 30 heptad segments, each 210 amino acids long. R3 is a 30 heptad segment, 210 amino acids long that is 33% identical to R1. References are as follows: Rabbit myosin S2 domain (CAPONY and ELZINGA 1981) and nematode myosin (KARN, BRENNER and BARNETT 1983).

homologies (Table 2). This defect was eliminated by performing the usual significance analysis on a set of control sequences scrambled in such a way that the heptad nature was always retained. Specifically, amino acids at the hydrophobic core positions of the reference sequence were scrambled separately from amino acids at other positions (MATERIALS AND METHODS, the heptad shuffle). The comparison of a sequence of interest with this control set gives a value called *Z* heptad. Additional calculations show that the heptad shuffle does not dramatically change the similarity scores obtained either for two non-heptad proteins of similar amino acid composition or for two heptad-containing proteins with a significant percentage of amino acid identities (S. SUBBIAH, unpublished data).

For each pairwise combination of RAD50 heptads I and II and the S-2 domains of rabbit and nematode myosins, alignments were made by the standard NEEDLEMAN-WUNSCH procedure. For each pair, alignments were made using two different matrices to define amino acid relatedness (UP and SIM, MATERIALS AND METHODS). The significance of each alignment was evaluated both with and without the heptad shuffle (Table 2). Heptad I exhibits no significant similarity to any of the other sequences regardless of the relatedness matrix and significance analysis used. In contrast, heptad II and the rabbit S-2 domain shows significant relatedness under all conditions, although in one case the level of relatedness was slightly less than the number of standard deviations usually taken as a cutoff point (MCLACHLAN 1971; DAYHOFF 1978).

Biological importance of coiled coil regions: Proteins that contain coiled coil regions participate in a variety of biological processes including muscle movement, anti-sense RNA pairing, nuclear envelope formation and viral pathogenesis (STEINERT, STEVENS and ROOP 1985; MCLACHLAN and KARN 1982; BANNER, KOKKINIDIS and TSERNOGLOU 1987; AEBI *et al.* 1986; MCKEON, KIRSCHNER and CAPUT 1986). Some of these proteins show an additional feature that is also observed in the RAD50 protein sequence; the nuclear and cytoplasmic intermediate filament protein sequences contain long heptad repeat regions that are interspersed with non-heptad spacers (STEINERT, STEVENS and ROOP 1985). The RAD50 protein also shows a 320 amino acid non-heptad spacer region between heptad I and heptad II. However, the length of the RAD50 spacer is much larger than those present in intermediate filament proteins (STEINERT, STEVENS and ROOP 1985).

In the absence of any specific information about the role of RAD50 function, we can only speculate as to the role of its coiled coil regions. Three intriguing and somewhat unusual possibilities have arisen thus far. One possibility would be that the RAD50 protein stretches across the synaptonemal complex. The length of each coiled coil region is approximately the distance between an axial element and the central element, so the protein might extend from one axial element to the other, with the two heptad containing regions spanning the two regions that separate the axial and central elements. A second possibility, suggested by the similarity of heptad II to the S2 domain of myosin would be that RAD50 is part of a contractile system which actively brings homologous chromosomes together once homology is sensed. A third possibility, raised by the apparent similarity in overall structure between RAD50 protein and kinesin (J. GELLES, personal communication), would be that RAD50 is involved in translocation of DNA molecules along microtubules, or vice versa. Kinesin is an ATP dependent translocator that travels along microtubules. Electron microscopy suggests that each kinesin molecule consists of a globular domain at one end plus two rod-like elements connected by a flexible joint (AMOS 1987). The globular domain is in contact with microtubules and presumably encodes the ATP binding activity; the arrangement of the rod-like elements is suggestive of two α-helical coiled coils interrupted by a spacer (AMOS 1987). This organization is similar to that of RAD50, with its amino terminal purine nucleotide binding domain and two separated heptad repeat regions. However, computer comparisons between RAD50 and a partial sequence of Drosophila kinesin do not reveal any significant similarity beyond that expected from their common coiled coil and purine nucleotide binding features (L. GOLDSTEIN, personal communication). Biochemical purification of the RAD50 protein and genetic analysis of *rad50* mutants should be useful in understanding the

role of RAD50 protein in mitotic DNA repair and in meiotic recombination.

We are grateful to R. E. ESPOSITO for providing the *RAD50* clone, to WENDY RAYMOND, PATRICK ERRADA and LIANG CAO for help in organizing sequence information and technical advice, to TSE MING YANG for the cloning of *RAD50* upstream sequences and to WILL GILBERT for his generous help in computer analysis. We are also thankful to MARGARET BRADLEY and TERRY WEBSTER for advice on mapping purine nucleotide binding domains, to members of the KLECKNER laboratory for critical reading of the manuscript, and to JIM HENLE for expert manuscript preparation. E. A. was funded by a National Institutes of Health genetics training grant awarded to the Department of Biochemistry and Molecular Biology. N. K. was funded by National Science Foundation grant DCB-8711551.

LITERATURE CITED

AEBI, U., J. COHN, L. BUHLE and L. GERACE, 1986 The nuclear lamina is a meshwork of intermediate-type filaments. Nature **323:** 560–564.

ALANI, E., L. CAO and N. KLECKNER, 1987 A method for gene disruption that allows repeated use of *URA3* selection in the construction of multiply disrupted yeast strains. Genetics **116:** 541–545.

AMOS, L. A., 1987 Kinesin from pig brain studied by electron microscopy. J. Cell. Sci. **87:** 105–111.

BACKENDORF, C., H. SPAINK, A. P. BARBEIRO and P. VAN DE PUTTE, 1986 Structure of the *uvr*B gene of *Escherichia coli*. Homology with other DNA repair enzymes and characterization of the *uvr*B5 mutation. Nucleic Acids Res. **14:** 2877–2890.

BAKER, B. S., A. T. C. CARPENTER, M. S. ESPOSITO, R. E. ESPOSITO and L. SANDLER, 1976 The genetic control of meiosis. Annu. Rev. Genet. **10:** 53–134.

BANNER, D. W., M. KOKKINIDIS and D. TSERNOGLOU, 1987 Structure of the ColE1 Rop protein at 1.7 Å resolution. J. Mol. Biol. **196:** 657–675.

BOEKE, J., F. LACROUTE and G. FINK, 1984 A positive selection for mutants lacking orotidine-5′ phosphate decarboxylase activity in yeast: 5-fluoro-orotic acid resistance. Mol. Gen. Genet. **197:** 345–346.

BRADLEY, M. K., T. F. SMITH, R. H. LATHROP, D. M. LIVINGSTON and T. A. WEBSTER, 1987 Consensus topography in the ATP binding site of the simian virus 40 and polyomavirus large tumor antigens. Proc. Natl. Acad. Sci. USA **84:** 4026–4030.

CAPONY, J. P., and M. ELZINGA, 1981 The amino acid sequence of a 34,000 dalton fragment from S-2 of myosin. Biophys. J. **33:** 148a.

CHOU, P. Y., and G. D. FASMAN, 1978 Empirical predictions of protein conformation. Annu. Rev. Biochem. **47:** 251–276.

COHEN, C., and D. A. PARRY, 1986 α-Helical coiled coils—a widespread motif in proteins. Trends Biochem. Sci. **11:** 245–248.

DAYHOFF, M., 1978 *Atlas of Protein Sequence and Structure*, Vol. 5. National Biomedical Research Foundation, Silver Spring, Md.

DEVER, T. E., M. GLYNIAS and W. MERRICK, 1987 GTP-binding domain: three consensus sequence elements with distinct spacing. Proc. Natl. Acad. Sci. USA **84:** 1814–1818.

DREUSICKE, D., and G. E. SCHULZ, 1986 The glycine-rich flexible loop of adenylate kinase forms a giant anion hole. FEBS Lett. **208:** 301–304.

FARNET, C., R. PADMORE, L. CAO, W. RAYMOND, E. ALANI and N. KLECKNER, 1988 The *RAD50* gene of *S. cerevisiae*. UCLA Symp. Mol. Cell. Biol. **83:** 201–215.

FRASER, R. D., and T. P. MACRAE, 1973 *Conformations in Fibrous Proteins*. Academic Press, New York.

FRY, D., S. KUBY and A. S. MILDVAN, 1986 ATP-binding site of adenylate kinase: mechanistic implications of its homology with ras-encoded p21, F1-ATPase, and other nucleotide-binding proteins. Proc. Natl. Acad. Sci. USA **83:** 907–911.

GAME, J. C., and R. MORTIMER, 1974 A genetic study of X-ray sensitive mutants in yeast. Mutat. Res. **24:** 281–282.

GAME, J. C., T. J. ZAMB, R. J. BRAUN, M. RESNICK and R. M. ROTH, 1980 The role of radiation (*rad*) genes in meiotic recombination in yeast. Genetics **94:** 51–68.

GUARENTE, L., G. LAURER, T. ROBERTS and M. PTASHNE, 1980 Improved methods for maximizing expression of a cloned gene: a bacterium that synthesizes rabbit β-globin. Cell **20:** 543–553.

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

HAYNES, R., and B. KUNZ, 1981 *The Molecular Biology of the Yeast Saccharomyces cerevisiae: Life Cycle and Inheritance*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

HOLM, C. W., W. MEEKS-WAGNER, W. FANGMAN and D. BOTSTEIN, 1986 A rapid efficient method for isolating DNA from yeast. Gene **42:** 169–173.

HUISMAN, O., W. RAYMOND, K. FROEHLICH, P. ERRADA, N. KLECKNER, D. BOTSTEIN and M. A. HOYT, 1987 A Tn*10-lacZ-kan*R-*URA3* gene fusion transposon for insertion mutagenesis and fusion analysis of yeast and bacterial genes. Genetics **116:** 191–199.

ITO, H., Y. FUKUDA, K. MURATA and A. KIMURA, 1983 Transformation of intact yeast cells treated with alkali cations. J. Bacteriol. **153:** 163–168.

KARN, J., S. BRENNER and L. BARNETT, 1983 Protein structural domains in the *Caenorhabditis elegans unc-54* myosin heavy chain gene are not separated by introns. Proc. Natl. Acad. Sci. USA **80:** 4253–4257.

KUPIEC, M., and G. SIMCHEN, 1984 Cloning and mapping of the *RAD50* gene of *Saccharomyces cerevisiae*. Mol. Gen. Genet. **193:** 525–531.

KYTE, J., and R. F. DOOLITTLE, 1983 A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. **157:** 105–132.

LATHROP, R., T. WEBSTER and T. F. SMITH, 1987 ARIADNE: pattern directed inference and hierarchical abstraction in protein structure recognition. Commun. ACM **30:** 909–921.

LIPMAN, D. J., and W. R. PEARSON, 1985 Rapid and sensitive protein similarity searches. Science **227:** 1435–1441.

MALONE, R., 1983 Multiple mutant analysis of recombination in yeast. Mol. Gen. Genet. **189:** 405–412.

MALONE, R., and R. E. ESPOSITO, 1981 Recombinationless meiosis in *Saccharomyces cerevisiae*. Mol. Cell. Biol. **1:** 891–901.

MALONE, R., K. JORDAN and W. WARDMAN, 1985 Extragenic revertants of *rad50*, a yeast mutation causing defects in recombination and repair. Curr. Gen. **9:** 453–461.

MANIATIS, T., E. FRITSCH and J. SAMBROOK, 1982 *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

MCKEON, F. D., M. W. KIRSCHNER and D. CAPUT, 1986 Homologies in both primary and secondary structure between nuclear envelope and intermediate filament proteins. Nature **319:** 463–468.

MCLACHLAN, A., 1971 Tests for comparing related amino-acid sequences, cytochrome *c* and cytochrome c_{551}. J. Mol. Biol. **61:** 409–424.

MCLACHLAN, A., and J. KARN, 1982 Periodic charge distributions in the myosin rod amino acid sequence match cross-bridge spacings in muscle. Nature **299:** 226–231.

MCLACHLAN, A., and J. KARN, 1983 Periodic features in the

amino acid sequence of nematode myosin rod. J. Mol. Biol. **164:** 605–616.

MELTON, D., P. KRIEG, M. REBAGLIATI, T. MANIATIS and M. GREEN, 1984 Efficient *in vitro* synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. Nucleic Acids Res. **12:** 7035–7056.

MESSING, J., 1983 New M13 vectors for cloning. Methods Enzymol. **101:** 20–78.

MILLER, J., 1972 *Experiments in Molecular Genetics.* Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

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

MYERS, R., and T. MANIATIS, 1985 A general method for saturation mutagenesis of cloned DNA fragments. Science **229:** 242–247.

NAUMOVSKI, L., and E. C. FRIEDBERG, 1986 Analysis of the essential and excision repair functions of the *RAD3* gene of *Saccharomyces cerevisiae* by mutagenesis. Mol. Cell. Biol. **6:** 1218–1227.

NEEDLEMAN, S., and C. WUNSCH, 1970 A general method applicable to the search for similarities in the amino acid sequence of two proteins. J. Mol. Biol. **48:** 443–453.

New England Biolabs Catalog, 1986 New England Biolabs, Beverly, Mass.

PARRY, D. A. D., W. G. CREWTHER, R. D. B. FRASER and T. P. MACRAE, 1977 Structure of α-keratin: structural implication of the amino acid sequences of the type I and type II chain segments. J. Mol. Biol. **113:** 449–454.

RALPH, W., T. WEBSTER and T. F. SMITH, 1987 A modified Chou and Fasman protein structure algorithm. Comput. Appl. Biosci. **3:** 211–216.

REINSTEIN, J., M. BRUNE and A. WITTINGHOFER, 1988 Mutations in the nucleotide binding loop of adenylate kinase of *Escherichia coli.* Biochemistry **27:** 4712–4720.

ROSE, M., P. GRISAFI and D. BOTSTEIN, 1984 Structure and function of the yeast *URA3* gene: expression in *E. coli.* Gene **29:** 113–124.

ROSSMAN, M. G., D. MORAS and R. W. OLSEN, 1974 Chemical and biological evolution of a nucleotide-binding protein. Nature **250:** 194–199.

SHERMAN, F., G. FINK and J. HICKS, 1983 *Methods in Yeast Genetics.* Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

SMITH, G. R., 1988 Homologous recombination in procaryotes. Microbiol. Rev. **52:** 1–28.

SOUTHERN, E. M., 1975 Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98:** 503–517.

STEINERT, P. M., A. C. STEVENS and D. R. ROOP, 1985 The molecular biology of intermediate filaments. Cell **42:** 411–419.

STINCHCOMB, D., C. MANN and R. W. DAVIS, 1982 Centromeric DNA from *Saccharomyces cerevisiae.* J. Mol. Biol. **158:** 157–179.

STONE, D., and L. B. SMILLIE, 1978 The amino acid sequence of rabbit skeletal α-tropomyosin. J. Biol. Chem. **253:** 1137–1148.

STRUHL, K., 1985 Naturally occurring poly (dA-dT) sequences are upstream promoter elements for constitutive transcription in yeast. Proc. Natl. Acad. Sci. USA **82:** 8419–8423.

WALKER, J., M. SARASTE, M. RUNSWICK and N. GAY, 1982 Distantly related sequences in α- and β-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. EMBO J. **1:** 945–951.

Communicating editor: J. W. DRAKE