Proc. Natl. Acad. Sci. USA
Vol. 87, pp. 1571–1575, February 1990
Biochemistry

Identification of five putative yeast RNA helicase genes

(gene family/degenerate oligonucleotides/polymerase chain reaction)

TIEN-HSIEN CHANG, JAIME ARENAS, AND JOHN ABELSON

Division of Biology 147-75, California Institute of Technology, Pasadena, CA 91125

Contributed by John Abelson, December 5, 1989

ABSTRACT **The RNA helicase gene family encodes a group of eight homologous proteins that share regions of sequence similarity. This group of evolutionarily conserved proteins presumably all utilize ATP (or some other nucleoside triphosphate) as an energy source for unwinding double-stranded RNA. Members of this family have been implicated in a variety of physiological functions in organisms ranging from *Escherichia coli* to human, such as translation initiation, mitochondrial mRNA splicing, ribosomal assembly, and germinal line cell differentiation. We have applied polymerase chain reaction technology to search for additional members of the RNA helicase family in the yeast *Saccharomyces cerevisiae*. Using degenerate oligonucleotide primers designed to amplify DNA fragments flanked by the highly conserved motifs V L D E A D and Y I H R I G, we have detected five putative RNA helicase genes. Northern and Southern blot analyses demonstrated that these genes are single copy and expressed in yeast. Several members of the RNA helicase family share sequence identity ranging from 49.2% to 67.2%, suggesting that they are functionally related. The discovery of such a multitude of putative RNA helicase genes in yeast suggests that RNA helicase activities are involved in a variety of fundamentally important biological processes.**

In cells, DNA and RNA molecules frequently undergo a variety of dynamic conformational changes. For example, the unwinding of double-stranded DNA helix is required for replication, recombination, and transcription, and the melting of secondary structure is required for the initiation of translation of many mRNAs. A number of proteins involved in controlling the helical structure of DNA molecules have been purified and extensively studied (reviewed in ref. 1). Among these are the DNA binding proteins, such as histones and single-stranded DNA binding proteins, and proteins with enzymatic activities, such as DNA helicases and topoisomerases.

In contrast, only recently have proteins that function in destabilizing RNA helices begun to receive attention. Mammalian eukaryotic translation initiation factor eIF-4A was the first of the RNA helicases to be characterized. eIF-4A has an intrinsic RNA-dependent ATPase activity and is capable of unwinding mRNA in an ATP-dependent fashion (2, 3). It is thought that eIF-4A catalyzes ATP hydrolysis and utilizes the energy generated to disrupt mRNA secondary structure. Recently, the RNA helicase activity as well as the RNA-dependent ATPase activity of both human p68 protein and simian virus 40 large tumor antigen have been unambiguously demonstrated (4, 5). In addition, an *Escherichia coli* protein SrmB with sequence homology to eIF-4A has been reported to bind RNA in the absence of ATP and to hydrolyze ATP in the presence of RNA (6).

More recently, strong sequence homology was detected between eIF-4A; p68; genes encoding SrmB and the newly characterized *Drosophila* gene *vasa* (7); yeast genes *TIF1*, *TIF2*, and *MSS116* (8, 9); as well as mouse *PL10* gene (10). Based on the amino acid sequence alignments of these proteins, an additional gene family, the RNA helicase gene family, consisting of both established and putative RNA helicase genes, was defined (11). Although the sequence conservation is spread out over a stretch of 420 amino acids, several sequence elements are especially striking. The sequence D X_4 A X_4 G K T, found in all eight proteins, is typical for the A motif of ATP binding proteins (12–14). The D E A D box, (V/I) L D E A D X_2 L, on the other hand, represents a special version of the B motif of ATP-binding proteins (14) and is uniquely present in members of the RNA helicase family. Equally conserved is the sequence H R I G R, which has been speculated to participate in polynucleotide binding and/or the unwinding activity (11).

To date, three putative RNA helicase genes have been reported in yeast. The *TIF1* gene was isolated as a suppressor of a mitochondrial missense mutation (15). A second gene, *TIF2*, is almost identical to *TIF1* within the open reading frame, except for five silent substitutions, and thus specifies the same protein (8, 15). TIF1 and TIF2 are considered to be the yeast translation initiation factor based on their high sequence homology to the mammalian eIF-4A (8, 15). The third putative RNA helicase gene in yeast, *MSS116*, was isolated in a genetic screen for nuclear mutants that specifically affect mitochondrial splicing (16). Mutations in this locus have been shown to directly affect the splicing of several introns of cytochrome *b* and cytochrome *c* oxidase subunit I primary transcripts (9). Thus, for the first time, an RNA helicase activity was implicated in mRNA splicing.

ATP is required for pre-mRNA splicing *in vitro* in both the mammalian and yeast systems (reviewed in ref. 17). This requirement of ATP occurs in both steps of the splicing reaction and in several steps of the spliceosome assembly (18–20). The reason for the ATP requirement is currently unknown, but it is likely that multiple molecules of ATP must be hydrolyzed for each splicing event. We are interested in the possibility that RNA helicase may function in pre-mRNA splicing and that this may explain the ATP requirement. The most dramatic event in the assembly of the spliceosome, which would appear to require an RNA helicase, occurs midway through the assembly process. The small nuclear RNAs U4 and U6 are base-paired together in the same small nuclear ribonucleoprotein particle (21). There are 21 base pairs (bp) joining the two RNAs, and in the absence of protein the melting temperature of the complex is 53°C (21). Prior to formation of the active spliceosome complex, U4 either leaves the complex or is markedly destabilized, so it cannot be detected when the spliceosome is analyzed by native polyacrylamide gel electrophoresis (18, 20, 22).

Although the function of MSS116 in mitochondrial self-splicing is unknown, it is likely that it acts to generate an enzymatically active RNA structure. RNA helicases may

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: eIF, eukaryotic initiation factor; PCR, polymerase chain reaction.

also act to alter structure in nuclear pre-mRNA splicing. Motivated by these considerations, we have undertaken a project to enumerate and characterize the RNA helicases in yeast. By using polymerase chain reaction (PCR) technology, we have identified five additional members of this gene family in yeast.

MATERIALS AND METHODS

Materials. *Thermus aquaticus* DNA polymerase (*Taq* polymerase) was obtained from Perkin–Elmer/Cetus. Restriction enzymes were obtained from New England Biolabs. Oligonucleotides were synthesized on an Applied Biosystems DNA synthesizer and purified by urea polyacrylamide gel electrophoresis. GeneScreen hybridization transfer membrane was from NEN Research Products.

PCR. Yeast genomic DNA was prepared as described (23). Typically, 500 ng of yeast genomic DNA was used in a 100-μl PCR reaction mixture containing 10 mM Tris·HCl (pH 8.3), 1.5 mM $MgCl_2$, 0.1% gelatin, 200 μM each deoxynucleoside triphosphate, 3.5 μM each primer, and 2.5 units of *Taq* polymerase. Reaction mixtures were denatured at 94°C for 10 min and then chilled on ice immediately before the addition of deoxynucleotides and enzyme. Amplification of DNA sequences was achieved by 30 cycles of DNA denaturation (94°C, 1 min), primer annealing (45°C, 1 min), and primer extension (72°C, 2 min), followed by a 10-min incubation at 72°C after the final cycle was completed. PCRs were all performed on a Cetus DNA thermocycler.

Cloning and Sequencing of the PCR Products. The DNA products of the PCR were extracted with phenol/chloroform, ethanol precipitated, and then digested with *Bam*HI and *Spe* I at 37°C for 2 hr. The digested DNAs were then electrophoresed on a 0.7% low melting point agarose gel, and the desired product was excised for cloning into the Bluescript KS vector (Stratagene). Initially, clones were sorted by single lane (T-lane) sequencing using double-stranded plasmid DNAs as templates. Subsequently, the complete sequences of the cloned PCR products were obtained by combined single-stranded and double-stranded DNA sequencing using KS and T7 primers.

Analyses of Genomic Sequences. For Southern blot analyses, 3.6 μg of yeast genomic DNA was digested to completion by either *Eco*RI, *Pst* I, or *Bgl* II and fractionated on a 0.8% agarose gel. GeneScreen membrane was used to bind nucleic acids in these experiments. The hybridization probes were generated by using purified insert DNAs from various PCR clones as templates for random-primer labeling with [α-^{32}P]dATP as described (24). Filters were prehybridized for 2 hr at 42°C in 50% formamide containing 6× SSC (1× SSC = 0.15 M NaCl/0.015 M sodium citrate, pH 7.0), 5× Denhardt's solution (1× Denhardt's solution = 0.02% Ficoll/0.02% polyvinylpyrrolidone/0.02% bovine serum albumin), and 0.1% SDS and were hybridized for 18 hr at 42°C in the same solution with ^{32}P-labeled DNA probes. Filters were washed to a final stringency of 0.2× SSC/0.1% SDS at 65°C before being subjected to autoradiography.

Analysis of mRNA. For the Northern blotting, yeast total RNA, poly(A)$^-$ RNA, and poly(A)$^+$ RNA were prepared as described (25). RNAs were treated with glyoxal before being electrophoresed on a 1% agarose gel for Northern blotting (26). Conditions used in Northern blot analyses were the same as described in the section above.

Sequence Comparison. Pairwise sequence comparisons for identity and similarity were performed on a VAXstation using the BESTFIT program in the GCG sequence analysis software package developed by Genetics Computer Group of University of Wisconsin.

RESULTS AND DISCUSSION

The eight protein sequences known to be members of the RNA helicase family share pronounced sequence homology (9, 11). In particular, several blocks of amino acid sequence were found to be highly conserved in the RNA helicases in organisms ranging from *E. coli* to human. Of these, blocks I and II (for block assignments, see refs. 11 and 12) were implicated in nucleotide binding and appeared to be present in a number of the ATP-utilizing enzymes (12). In contrast, blocks Ia, V, and VI are unique to the RNA helicases (9, 12). We reasoned that any protein that contains both the nucleotide-binding site and the putative RNA helicase conserved sequences is likely to be an RNA helicase. Computer search in the data bank for proteins containing the block II and block VI peptide sequences yielded only members of this gene family and thus qualified blocks II and VI as suitable probes for identifying new RNA helicase genes.

Identification of Additional Members of the RNA Helicase Family in Yeast. To identify additional members of the RNA helicase gene family, degenerate oligonucleotides covering

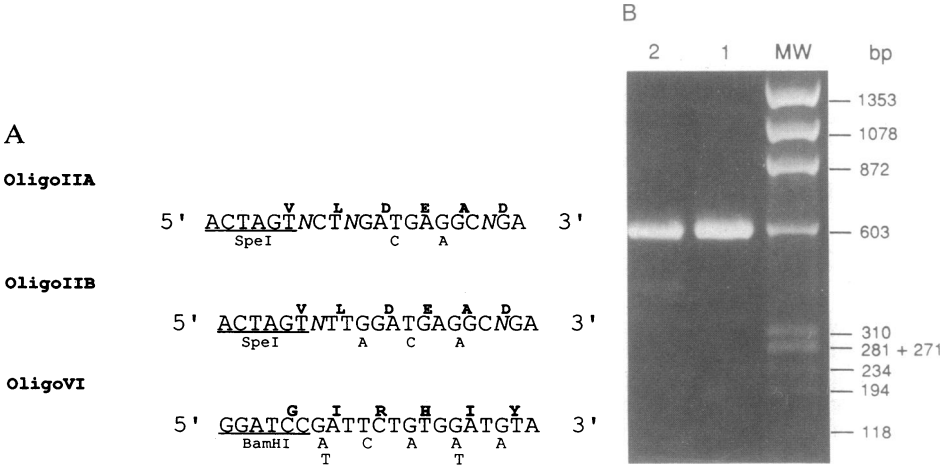

FIG. 1. PCRs using degenerate oligonucleotides as primers. (*A*) Sequences of the oligonucleotides used in PCRs. Nucleotide sequences were derived from the amino acid sequences (single-letter code) as indicated. The oligoVI is the complement of the coding sequence that specifies those amino acids. The italicized *N* represents all four possible nucleotides. Restriction sites (underlined) have been built into oligoIIA and -IIB (*Spe* I), as well as oligoVI (*Bam*HI). (*B*) Gel electrophoresis of the amplified PCR products. After 30 cycles, PCR products were electrophoresed on a 1% agarose gel. Lane 1, reaction using oligoIIA and oligoVI; lane 2, reaction using oligoIIB instead of -IIA. Size standards are *Hae* III-digested ϕX174 DNA.

all possible codons for the peptide sequences V L D E A D (block II) and Y I H R I G (block VI) were designed for use in the PCR (Fig. 1*A*). Two oligonucleotides, OligoIIA and OligoIIB, were synthesized for the V L D E A D sequence, since leucine is encoded by two different groups of triplet codons.

Initially, we tested three different primer annealing temperatures, 35°C, 40°C, and 45°C, for PCRs, and obtained apparently identical results. We therefore selected the most stringent temperature, 45°C, for subsequent PCRs. After 30 cycles of PCR amplification using yeast genomic DNA as template, a predominant product of ≈600 bp as well as some minor products could be clearly detected in agarose gel electrophoresis (Fig. 1*B*). Identical results were obtained in reactions using either OligoIIA or OligoIIB, although the yields varied (Fig. 1*B*, lanes 1 and 2). In both cases, the size of the predominant product, ≈600 bp, was in good agreement with the estimated sizes (between 550 and 650 bp) of the gene fragments encompassed by the V L D E A D and Y I H R I G sequences in the eight known members of the gene family. For the purpose of identifying sequences amplified in the PCR reactions, PCR products were digested with restriction enzymes *Bam*HI and *Spe* I, and the major product (i.e., the ≈600-bp DNA fragment) was isolated and cloned into plasmid vector.

Plasmid DNAs were isolated from 150 randomly picked transformants and their inserts were sorted by T-lane DNA sequencing. Based on the T-lane patterns, eight different groups of sequences could be categorized, and they were subsequently designated as *CA1–CA8*. Assuming the primer annealing is highly specific under our PCR conditions, we expected to be able to retrieve at least one of the three partial sequences corresponding to the documented yeast genes—namely, *MSS116* (9), *TIF1*, and *TIF2* (8). Indeed, when the complete DNA sequences of clones *CA1–CA8* were obtained and their amino acid sequences deduced, all three of the previously determined sequences were present. Yeast translation initiation factor genes *TIF1* and *TIF2* are almost identical except for five silent substitutions (8), and they were represented by clone *CA2*. *MSS116*, a yeast nuclear gene involved in mitochondrial splicing (9) was represented by clone *CA7*. The frequency of generating the *CA1–CA8* clones in this experiment varied widely. Of the 150 randomly picked clones, 4 were identified as *CA1*, 2 as *CA2*, 8 as *CA3*, 27 as *CA4*, 67 as *CA5, -6* (see below), 4 as *CA7*, and 38 as *CA8*. This could be due to different complementarity of the degenerate oligonucleotides to their target sequences. Clones *CA5* and *CA6* were identical, with the exception of one nucleotide transition (C to T), which generated an amino acid substitution (threonine to isoleucine) adjacent to the Y I H R I G sequence. Since this variation was distributed evenly in all the *CA5, -6* clones we have sequenced, and since our Southern analyses indicated that there is only one *CA5, -6* gene in the yeast genome (see below), this variation may reflect a natural allelic polymorphism in yeast, a very early deamination event (from cytosine to uridine) occurring in the PCR reaction, or simply an error of the *Taq* polymerase.

Protein Sequence Comparison of Members of the RNA Helicase Family. In Fig. 2, we compiled a list of 15 sequences of the RNA helicase family between the V L D E A D and Y I H R I G boxes. Of these, 9 have been previously published [p68 (human), eIF-4A1 (mouse), eIF-4A2 (mouse), PL10 (mouse), vasa (*Drosophila*), TIF1 (yeast), TIF2 (yeast), MSS116 (yeast), and SrmB (*E. coli*)], 5 were discovered in this work [CA1, CA3, CA4, CA5, -6, and CA8], and 1 is unpublished [SPB4, a yeast protein involved in ribosomal biogenesis (A. B. Sachs and R. Davis, personal communication)]. In addition, PRP5, a yeast protein essential for nuclear pre-mRNA splicing, can also be included in the RNA helicase family (the PRP5 sequence will be published elsewhere; G. Dalbadie-McFarland and J. Abelson, unpublished data).

A striking homology among these 15 sequences is observed when they are optimally aligned (Fig. 2). Over a stretch of 217 amino acids, the calculated identity scores range from 27.2% (SPB4 versus CA8) to 73.8% (eIF-4A1 versus TIF1), with an average of 36.1%. With such a strong homology, these sequences appear to belong to a single protein family. The eIF-4A and TIF proteins are thought to be carrying out identical functions during translation initiation based on their sequence homology (8). We also noticed that there is an intriguing overall identity among p68, PL10, vasa, and CA1 (from 49.2% to 67.4%). Although these proteins are found in different organisms, their strong homology with each other suggests that they are functionally closely related. In particular, PL10, a protein involved in mouse spermatogenesis

```
                                  10        20        30        40        50        60        70        80        90       100
                                  *         *         *         *         *         *         *         *         *         *
eIF-4A1/A2   Mouse      VLDEADEMLSR.GFKDQIYDIFQKL.......NSNTQVVLLSATMPSD.VLEVTKKFMRDPIRILVKKEELTLEG........IRQFYINVER...EEWK
Tif1/2/CA2   Yeast      ILDEADEMLSS.GFKEQIYQIFTLL.......PPTTQVVLLSATMPND.VLEVTTKFMRNPVRILVKKDELTLRG........IKQFYVNVEE...EEYK
p68          Human      VLDEADRMLDM.GFEPQIRKIVDQI.......RPDRQTLMFSATWPKE.VRQLAEDFLKDYIHINIGALELSANHN.......ILQIVDVCH....DVEK
vasa         Drosophila VLDEADRMLDM.GFSEDMRRIMTHVTM.....RPEHQTLMFSATFPEE.IQRMAGEFLKNYVFVAIGIVGGACSD........VKQTIYEVNK...YAKR
SrmB         E.coli     ILDEADRMLDM.GFAQDIEHIAGET.......RWRKQTLLFSATLEGDAIQDFAERLLE..DPVEVSANPSTRERKK......IHQWYYRADD...LEHK
CA3          Yeast      VLDEADRMLEK.GFEEDIKNIIRETD......ASKRQTLMFTATWPKE.VRELASTFMNNPIKVSIGNTDQLTANKR......ITQIVEVVDP...RGKE
CA4          Yeast      VLDEADRCLDM.GFKKTLDAIVSTL.......SPSRQTLLFSATQSQS.VADLARLSLTDYKTVGTHDVMDGSVNKEASTPETLQQFYI.......EVPL
CA5/6        Yeast      VLDEADNMLDQGLGDQC..IRVKRFL.....PKDTQLVLFSATFADA.VRQYAKKIVPNANTLELQTNEVNVDA........IKQLYMDCKNEADKFDV
CA8          Yeast      VLDEADKMIDL.GFEDQVTNILTKVDINADS.AVNRQTLMFTATMTPV.IEKLAAGYMQKPVYATIGVETGSEPL........IQQVVEYADNDEDKFKK
PL10         Mouse      VLDEADRMLDM.GFEPQIRRIVEQDTMPP...KGVRHTMMFSATFPKE.IQMLARDFLDEYIFLAVGRVGSTSEN........ITQKVVWVEE...ADKR
CA1          Yeast      VLDEADRMLDM.GFEPQIRHIVEECDMPS...VENRQTLMFSATFPVD.IQHLARDFLDNYIFLSVGRVGSTSEN........ITQRILYVDD...MDKK
MSS116/CA7   Yeast      VLDEADRLLEI.GFRDDLETISGILNEKNSKSADNIKTLLFSATLDDK.VQKLANNIMNKKECLFLDTVDKNEPEAHER....IDQSVVISEKF..ANSI
SPB4         Yeast      VMDEADRLLDM.SFIKDTEKILRLL.......PKQRRTGLFSATMRSA.GSDIFKTGLRNPVRITVNSKNQAPSS........LKLNYCVVNPAEKLQLL

                                 110       120       130       140       150       160       170       180       190       200       210   217
                                  *         *         *         *         *         *         *         *         *         *         *     *
eIF-4A1/A2   LDTLCDLY....ETLTITQAVIFIN.TRRKVDWLTEKMHAR...DFTVSAMHGDMDQKERDVIMREFRSGSSRVLITT.DLLARGIDVQQVSLVINYDLPTNREN......YIHRIG
Tif1/2/CA2   YECLTDLY....DSISVTQAVIFCN.TRRKVEELTTKLRND...KFTVSAIYSDLPQQERDTIMKEFRSGSSRILIST.DLLARGIDVQQVSLVINYDLPANKEN......YIHRIG
p68          DEKLIRLM.EEIMSEKENKTIVFVE.TKRRCDELTRKMRRD...GWPAMGIHGDKSQQERDWVLNEFKHGKAPILIAT.DVASRGLDVEDVKFVINYDYPNSSED......YIHRIG
vasa         SKLIEILS....EQADGT..IVFVE.TKRGADFLASFLSEK...EFPTTSIHGDRKQSQREQALRDFKNGSMKVLIAT.SVASRGLDIKNIKHVINYDMPSKIDD......YVHRIG
SrmB         TALLVHLL....KQPEATRSIVFVRNKRERVHELANWLREA...GINNCYLEGEMVQGKENEAIKRLTEGRVNVLVAT.DVAARGIDIPDVSHVFNFDMPRSGDT......YLHRIG
CA3          RKLLELLKKYHSGPKKNEKVLIFAL.YKKEAARVERNLKYN...GYNVAAIHGDLSQQQRTQALNEFKSGKSNLLLAT.DVAARGLDIPNVKTVINLTFPLTVAD......YIHRIG
CA4          ADKLDILF.SFIKSHLKCKMIVFLSSSKQV.HFVYETFRKMQP.GISLMHLHGRQKQRARTETLDKFNRAQQVCLFAT.DVVARGIDFPAVDWVVQVDCPEDVDT......YIHRIG
CA5/6        LTELYGLM.......TIGSSIIFVA.TKKTANVLYGKLKSE...GHEVSILHGDLQTQERDRLIDDFREGRSKVLITT.NVLARGIDIPTVSMVVNYDLPTLANGQADPATYIHRIG
CA8          LKPIVAKY........DPPIIIFIN.YKQTADWLAEKFQKET..NMKVTILHGSKSQEQREHSLQLFRTNKVQIMIRT.NVAARGLDIPNVSLVVNFQISKKMDD......YIHRIG
PL10         SFLLDLLN....ATGKDSLILVFVE.TKKGADSLEDFLYHE...GYACTSIHGDRSQRDREEALHQFRSGKSPILVAT.AVAARGLDISNVDHVINFDLPSDIEE......YVHRIG
CA1          SALLDLLS.....AEHKGLTLIFVE.TKRMADQLTDFLIMQ...NFKATAIHGDRTQAERERALSAFKANVADILVAT.AVAARGLDIPNVTHVINYDLPSDIDD......YIHRIG
MSS116       FAAVEHIKKQIKERDSNYKAIIFAP.TVKFTSFLCSILKNEFKKDLPILEFHGKITQNKRTSLVKRFKKDESGILVCT.DVGARGMDFPNVHEVLQIGVPSELAN......YIHRIG
SPB4         VSILNNYK.FKKCIVYFPTCVSVSY.FYSFIQYLGKRNILVN..EVEIFSLHGKLQTSARTKTLTAFTDSLSNSVLFTTDVAARGIDIPDVDLVIQLDPPTNTDM......FMHRCG
```

FIG. 2. Sequence alignment of members of the RNA helicase family (single-letter code). The alignment was based primarily on Linder *et al.* (11), with additional manual optimization for sequence homology. Boldface type indicates either identities or strong similarities at the corresponding positions. Sequences are numbered starting from the V L D E A D sequence, with gaps included.

(10), and CA1 are 67.4% identical and 85.8% similar and thus may have similar functions. In this regard, it is interesting that clone *CA1* appears to be identical in sequence to a yeast gene, *SPP81*. Genetic analysis suggests that SPP81 interacts with PRP8, another yeast protein required for nuclear pre-mRNA splicing (D. Jamieson and J. Beggs, personal communication).

From the sequence alignment, we further observed that there are several highly conserved amino acids previously unnoticed. For instance, Arg-7, Asp-10, Gln-37, Ala-54, Gln-87, and Leu-104 are all highly conserved in their respective positions. The divergence away from the highly conserved Y (I/L/V) H R I G sequence in PRP5 as Y V H T T G (G. Dalbadie-McFarland and J. Abelson, unpublished data) and SPB4 as F M H R C G (A. B. Sachs and R. Davis, personal communication) offered a possible explanation for their absence in our collection. The regions of high amino acid conservation have similar spacing in all the proteins, suggesting that these sequences have similar or identical functions. Interestingly, PRP5 has a large insertion in the middle of the sequence (G. Dalbadie-McFarland and J. Abelson, unpublished data), whereas CA5, -6 contains a 6-amino acid insertion between Gly-205 and block VI sequence. These and other more divergent regions may reflect individual functional differences.

Expression and Copy Number of the Newly Identified Genes. To determine whether the newly identified putative RNA helicase genes are expressed in yeast, we performed hybridization analyses using the cloned DNAs as probes. In all cases in which results could be obtained, a single species of RNA was detected by each probe (Fig. 3). The estimated sizes of the transcripts recognized by each individual probe are 2.6 kilobases (kb) (*CA1*), 1.5 kb (*CA2*), 2.1 kb (*CA3*), 2.7 kb (*CA4*), and 1.85 kb (*CA5, -6*). The expressed RNAs appeared to be polyadenylylated, since they were highly enriched in the poly(A)$^+$ fraction. Thus, it is likely that these transcripts actually represent mRNAs encoding the putative RNA helicases. Judging from the intensity of the hybridization signals, *TIF1, -2* encode messages that are highly abundant, while the abundance of the *CA1* message is much lower. The failure to detect mRNAs encoded by *MSS116* and *CA8* suggests that their messages may be either unstable or stringently expressed in the cells. We have also determined the copy number of the yeast RNA helicase genes by Southern analyses (Fig. 4). In three different genomic restriction digests, all PCR clones, except *TIF1, -2* (*CA2* probe) and *CA1* (in *Pst* I digest only), hybridized to a single restriction fragment, indicating that they are most likely encoded by single-copy genes. As expected, the *TIF1, -2* cloned DNA

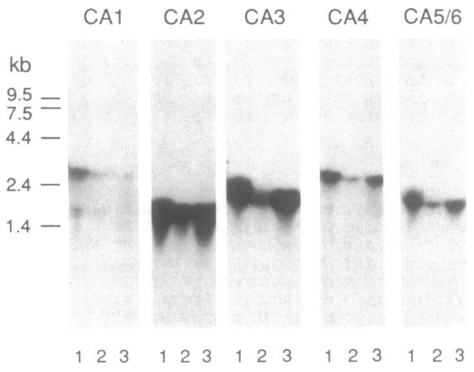

FIG. 3. Northern blot analyses of members of the RNA helicase gene family. Total yeast RNA (lane 1), poly(A)$^-$ RNA (lane 2), and poly(A)$^+$ RNA (lane 3) were denatured by glyoxal and electrophoresed on an agarose gel for Northern blotting. Probes used for each of the hybridization experiments are listed on the top, and size markers (RNA ladder obtained from BRL) are indicated.

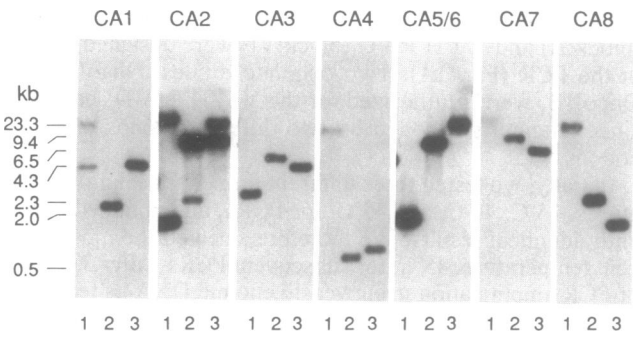

FIG. 4. Southern blot analyses of members of the RNA helicase gene family. Yeast genomic DNA was digested with *Pst* I (lane 1), *Eco*RI (lane 2), and *Bgl* II (lane 3) and separated on a 0.8% agarose gel for Southern blotting. Probes used in each of the hybridization experiments are listed on the top. Also indicated are the DNA size standards, the *Hin*dIII-digested λ DNA.

(the *CA2* probe) recognized two restriction fragments, confirming that there are two *TIF* genes in the yeast genome (8, 15). The predicted internal *Pst* I site in the *CA1* clone (Fig. 4, lane 1) was verified by DNA sequencing and restriction mapping (data not shown).

Implications and Speculations. Although several RNA helicases have in common the catalytic activity of unwinding double-stranded RNAs, their biological functions may be quite different. For instance, mammalian eIF-4A and its counterparts in yeast, TIF1 and TIF2, are essential in translation initiation (2, 3, 8), whereas p68 is a major nuclear antigen from a human hepatoma cell line and is thought to regulate cell growth and division (27). The functions of other putative RNA helicases are also rather diversified: MSS116 is involved in yeast mitochondrial mRNA splicing (9), PL10 is a protein expressed specifically during spermatogenesis in mouse (10), the *Drosophila* vasa protein is required for both oocyte formation and the specification of the posterior structures of the embryo (7), and the SrmB protein can suppress a mutation in the L24 ribosomal protein, which is essential in the assembly of 50S ribosomal subunits in *E. coli* (6). Since members of the RNA helicase family share homology only in the middle regions of their sequences (11), the sequence divergences outside the conserved regions probably reflect differences in their functional roles. The five additional RNA helicases we have discovered are therefore likely to have a variety of functions.

There are apparently at least 10 RNA helicase genes in yeast. Our failure to find *SPB4* and *PRP5* sequences in the PCR amplified DNAs leaves open the possibility that additional members exist. This large number of the putative RNA helicase genes in a single organism raises a number of intriguing questions: (*i*) Are they all essential for the survival of the organism? (*ii*) What are their substrate specificities? (*iii*) How do they exercise their physiological functions? (*iv*) What are the roles of the highly conserved motifs present in these proteins? In the case of the yeast *TIF1* and *TIF2* genes, inactivation of either gene by gene disruption has no effect on cell viability or mitochondrial functions (15). However, simultaneous inactivation of both genes is lethal to the cell (15), suggesting that the other RNA helicases in yeast are unable to substitute for the translation initiation function of TIF1 and TIF2.

The sequence conservation of the RNA helicase family spanning from eubacteria to eukaryotes suggests that RNA helicase activities are of fundamental importance to living cells, and evolved very early. During early stages of molecular evolution, it is likely that RNA catalyzed its own replication (28). Indeed, an RNA polymerase activity has been demonstrated in the *Tetrahymena* type I intron RNA

(29, 30). The separation of the strands following any self-replication (31, 32) may well have been a rate-limiting step. Thus, it is likely that a very early event in evolution, once protein synthesis had to evolve, would have been the development of an RNA helicase activity. In present day organisms, RNA helicases play a number of specialized roles. However, since the basic features of their enzymatic mechanism are highly specified for the destabilization of RNA helices, the basic structure of these enzymes is therefore highly conserved in evolution.

Note Added in Proof. We have learned that an essential gene in yeast, *DED1* (33), also appears to be a member of the RNA helicase family. The deduced amino acid sequence of *DED1* is highly homologous to those of *PL10*, *vasa*, and *CA1* (B. Rahe and J. Pringle, personal communication).

We thank N. Gautam and T. Wilkie for valuable discussions in using PCR techniques, and A. B. Sachs for communication of SPB4 sequence prior to publication. We also thank C. Ho, D. McPheeters, and C. O'Day for critical reading of the manuscript. T.-H.C. is supported by a Postdoctoral Fellowship from Merck Sharp & Dohme Research Laboratories. This work is supported by a grant awarded to J. Abelson from the National Institutes of Health.

1. Geider, K. & Hoffmann-Berling, H. (1981) *Annu. Rev. Biochem.* **50,** 233–260.
2. Ray, B. K., Lawson, T. G., Kramer, J. C., Cladaras, M. H., Grifo, J. A., Abramson, R. D., Merrick, W. C. & Thach, R. E. (1985) *J. Biol. Chem.* **260,** 7651–7658.
3. Abramson, R. D., Dever, T. E., Lawson, T. G., Ray, B. K., Thach, R. E. & Merrick, W. C. (1987) *J. Biol. Chem.* **262,** 3826–3832.
4. Hirling, H., Scheffner, M., Restle, T. & Stahl, H. (1989) *Nature (London)* **339,** 562–564.
5. Scheffner, M., Knippers, R. & Stahl, H. (1989) *Cell* **57,** 955–963.
6. Nishi, K., Morel-Diville, F., Hershey, J. W. B., Leighton, T. & Schnier, J. (1989) *Nature (London)* **336,** 496–498.
7. Lasko, P. F. & Ashburner, M. (1988) *Nature (London)* **335,** 611–617.
8. Linder, P. & Slonimski, P. P. (1988) *Nucleic Acids Res.* **21,** 10359.
9. Seraphin, B., Simon, M., Boulet, A. & Faye, G. (1989) *Nature (London)* **337,** 84–87.
10. Leroy, P., Alzari, P., Sasson, D., Wolgemuth, D. & Fellous, M. (1989) *Cell* **57,** 549–559.
11. Linder, P., Lasko, P., Ashburner, M., Leroy, P., Nielsen, P., Nishi, K., Schnier, J. & Slonimski, P. P. (1989) *Nature (London)* **337,** 121–122.
12. Hodgman, T. C. (1988) *Nature (London)* **333,** 22–23.
13. Hodgman, T. C. (1988) *Nature (London)* **334,** 478.
14. Walker, J. E., Saraste, M., Runswick, M. J. & Gay, N. J. (1982) *EMBO J.* **1,** 945–951.
15. Linder, P. & Slonimski, P. P. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 2286–2290.
16. Seraphin, B., Boulet, A., Simon, M. & Faye, G. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 6810–6814.
17. Padgett, R. A., Grabowski, P. J., Kornaska, M. M., Seiler, S. & Sharp, P. A. (1986) *Annu. Rev. Biochem.* **55,** 1119–1150.
18. Cheng, S.-C. & Abelson, J. (1987) *Genes Dev.* **1,** 1014–1027.
19. Kornarska, M. M. & Sharp, P. A. (1986) *Cell* **46,** 845–855.
20. Pikielny, C. W., Rymond, B. C. & Rosbash, M. (1986) *Nature (London)* **324,** 341–345.
21. Brow, D. A. & Guthrie, C. (1988) *Nature (London)* **334,** 213–218.
22. Lamond, A. I., Kornarska, M. M., Grabowski, P. J. & Sharp, P. A. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 411–415.
23. Johnston, J. R. (1984) in *Yeast, a Practical Approach*, eds. Compbell, I. & Duffus, J. H. (IRL, Oxford), p. 18.
24. Feinberg, A. P. & Vogelstein, B. (1983) *Anal. Biochem.* **132,** 6–13.
25. Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. & Struhl, K. (1987) in *Current Protocols in Molecular Biology* (Wiley, New York).
26. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY).
27. Ford, M. J., Anton, I. A. & Lane, D. P. (1988) *Nature (London)* **332,** 736–738.
28. Gilbert, W. (1986) *Nature (London)* **319,** 618.
29. Kruger, K., Grabowski, P. J., Zaug, A., Sands, J., Gottschling, D. E. & Cech, T. R. (1982) *Cell* **31,** 147–157.
30. Zaug, A. J. & Cech, T. R. (1986) *Science* **231,** 470–475.
31. Sharp, P. A. (1985) *Cell* **42,** 397–400.
32. Cech, T. R. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 4360–4363.
33. Struhl, K. (1985) *Nucleic Acids Res.* **13,** 8587–8601.