The EMBO Journal vol.9 no.9 pp.2775–2781, 1990

The molecular characterization of PRP6 and PRP9 yeast genes reveals a new cysteine/histidine motif common to several splicing factors

Pierre Legrain and André Choulika

Unité de Génétique Moléculaire des Levures, Department of Molecular Biology, Institut Pasteur, 28 rue du Docteur Roux, Paris 75724, France

Communicated by B.Dujon

***prp6* and *prp9* thermosensitive (ts) mutants are affected in pre-mRNA splicing and transport from the nucleus to the cytoplasm. *PRP6* and *PRP9* wild-type alleles have been sequenced. DNA sequence analysis reveals homologies in the 5′ and 3′ non-coding regions, suggesting a common regulation of gene expression. *PRP6* and *PRP9* genes encode a 899 amino acid and a 530 amino acid protein, respectively. The PRP6 protein has repeated motifs that evoke helix–loop–helix structures. Both PRP6 and PRP9 proteins have cysteine/histidine motifs loosely related to those found in zinc finger proteins. The substitution of some, but not all, of these residues by directed mutagenesis has a critical effect on the protein function. Homology searches reveal that two other proteins known to be involved in the nuclear splicing pathway—the yeast PRP11 and the human U1C proteins —contain similar sequences. The five cysteine/histidine motifs found in these four proteins display amino acid similarities in addition to the cysteine and histidine residues, indicating that they participate in biological structures or functions related to the splicing process. In addition, PRP6 and PRP9 exhibit leucine repeat motifs which may be implicated in protein interactions. The *prp6* and *prp9* ts mutations have been mapped and sequenced.**

Key words: sequence/splicing/U1 snRNP/yeast/zinc finger

Introduction

In eukaryotic cells, many primary transcripts are spliced before they are transported into the cytoplasm where they are translated into proteins. By *in vivo* and *in vitro* experiments, the major biochemical reactions involved in the splicing process have been characterized and important *cis*-acting sequences in the pre-mRNA have been determined (for review see Green, 1986). By *in vitro* experiments, it has been demonstrated that splicing occurs in large complexes called spliceosomes (Brody and Abelson, 1985) which are assembled step by step (Frendewey and Keller, 1985; Konarska and Sharp, 1986; Pikielny *et al.*, 1986; Cheng and Abelson, 1987; Legrain *et al.*, 1988; Seraphin and Rosbash, 1989). Several components of these complexes have been identified, among them small nuclear ribonucleoparticles (snRNPs) containing U1, U2, U4, U5 and U6 snRNAs (for review see Maniatis and Reed, 1987). Numerous peptides of higher eukaryotic snRNPs have been characterized, some of them being common to several snRNPs and others being specific for a given snRNP (for review see Lührmann, 1988). Soluble protein splicing factors, such as U2AF, which interact with snRNPs and/or pre-mRNA, have also been described in HeLa cells (Ruskin) *et al.*, 1988; Zamore and Green, 1989). In the yeast *Saccharomyces cerevisiae*, the availability of ts mutant strains specifically affected in pre-mRNA splicing (Hartwell *et al.*, 1970; Rosbash *et al.*, 1981; Vijayraghavan *et al.*, 1989) has allowed the cloning and the molecular characterization of several *trans*-acting splicing factors (PRP2, 3, 4, 8, 11, 16, 18; Vijayraghavan and Abelson, 1990; Whittaker *et al.*, 1990; for review see Woolford, 1989). Some of the PRP proteins are components of snRNPs, but the low content of snRNPs in the yeast nucleus, and also maybe an extensive evolutionary divergence with higher eukaryotes, has until now forbidden a clear identification of a yeast counterpart to a known mammalian snRNP protein. On the contrary, a putative mammalian homologue of the PRP8 protein has been identified (Anderson *et al.*, 1989; Pinto and Steitz, 1990), but its molecular characterization has not yet been completed, excluding any sequence comparison. *In vitro* analysis of the splicing pathway in yeast has also been performed, using either mutant cell extracts specifically heat-inactivated for one function (Lustig *et al.*, 1986; Cheng and Abelson, 1987) or extracts selectively depleted by biochemical or genetic means (Lossky *et al.*, 1987; Legrain *et al.*, 1988; Seraphin and Rosbash, 1989; for review see Woolford, 1989). However, up to now, two *prp* mutants, namely *prp6* and *prp9*, have not been analysed *in vitro* due to the failure to obtain mutant cell extracts which are specifically heat-inactivated (Lustig *et al.*, 1986).

In a previous study (Legrain and Rosbash, 1989), we showed by an *in vivo* assay using the translation of pre-mRNA as a measurement of the pre-mRNA escape from the splicing pathway, that in *prp6* and *prp9* ts mutants, unspliced pre-mRNAs exit from the nucleus at the non-permissive temperature. The present hypothesis is that these genes encode factors that stably commit pre-mRNAs in the splicing pathway *in vivo*. In the absence of functional PRP6 or PRP9 proteins, pre-mRNAs escape this pathway and are transported into the cytoplasm. In that respect, *PRP6* and *PRP9* gene products may be essential factors that discriminate between intron- and non-intron-containing transcripts (Legrain and Rosbash, 1989).

We report here sequences of *PRP6* and *PRP9* wild-type genes, and the identification of the ts mutations. The genes encode proteins which both contain cysteine/histidine motifs that are loosely related to zinc finger motifs of the TFIIIA type (for review see Vincent, 1986). A similar motif had been found in PRP11 protein (Chang *et al.*, 1988). By an extensive search among available protein sequences, we found that the human U1C protein (Sillekens *et al.*, 1988) displays a similar structural feature that allows the definition of a new family of structurally related proteins which are involved in RNA processing.

© Oxford University Press

```
 -89        TTCTATTAAG CAAAAAAAAA CGAAAAATTT CAGAAAATAT ACGGGTGAGT TGACATAAAG AGTTTAACAG CAAGAAAACG TCACATTTT

   1 ATG GAG AGG CCA TCT TTT TTG GAT CAA GAA CCA CCT GCA GGT TAC GTA CCA GGT ATT GGT CGT GGA GCC ACT GGA TTT TCA ACA AAA GAA
   1 Met glu arg pro ser phe leu asp gln glu pro pro ala gly tyr val pro gly ile gly arg gly ala thr gly phe ser thr lys glu

  91 AAG CAA GTG GTT AGT AAT GAT GAC AAA GGA AGA AGA ATA CCG AAA AGG TAC CGT GAA AAT TTG AAC AAC CAT CTT CAA AGC CAA CCG AAA
  31 lys gln val val ser asn asp asp lys gly arg arg ile pro lys arg tyr arg glu asn leu asn asn his leu gln ser gln pro lys

 181 GAT GAT GAA GAT GAT GAA GCT GCA AAT GTA TTC AAA ACG CTT GAA TTG AAA TTA GCA CAA AAG AAA AAG AAA AGA GCT AAT GAA AAG GAT
  61 asp asp glu asp asp glu ala ala asn val phe lys thr leu glu leu lys leu ala gln lys lys lys lys arg ala asn glu lys asp

 271 GAT GAC AAT TCA GTT GAT TCT TCA AAC GTG AAA CGG CAA TTT GCC GAT TTG AAA GAA TCA TTA GCT GCT GTA ACG GAG AGT GAG TGG ATG
  91 asp asp asn ser val asp ser ser asn val lys arg gln phe ala asp leu lys glu ser leu ala ala val thr glu ser glu trp met

 361 GAT ATT CCG GAT GCC ACA GAT TTT ACA AGA AGA AAC AAG AGA AAT AGA ATT CAA GAG CAA TTA AAC AGA AAA ACT TAT GCT GCA CCG GAT
 121 asp ile pro asp ala thr asp phe thr arg arg asn lys arg asn arg ile gln glu gln leu asn arg lys thr tyr ala ala pro asp

 451 TCG CTA ATA CCT GGG AAT GTT GAT TTA AAT AAA TTA ACG GAA GAA CGA GAA AAA TTA TTG CAA TCT CAA ATA GAT GAG AAT CTT GCA CAA
 151 ser leu ile pro gly asn val asp leu asn lys leu thr glu glu arg glu lys leu leu gln ser gln ile asp glu asn leu ala gln

 541 TTA ACG AAG AAT GCA AGT AAC CCT ATA CAG GTT AAT AAA CCG AAC GCT GCT ACC GAT GCC CTA AGT TAC TTA AAG GAC TTA GAA AAC GAT
 181 leu thr lys asn ala ser asn pro ile gln val asn lys pro asn ala ala thr asp ala leu ser tyr leu lys asp leu glu asn asp

 631 AGA GTA AAT TCT CTC TCA GAC GCA ACG TTA GAA GAT TTA CAG AAA ATG CGC ACA ATT TTA AAG TCA TAC AGA AAG GCC GAT CCA ACA AAT
 211 arg val asn ser leu ser asp ala thr leu glu asp leu gln lys met arg thr ile leu lys ser tyr arg lys ala asp pro thr asn

 721 CCA CAG GGT TGG ATA GCT TCT GCC AGA TTA GAA GAA AAG GCA AGA AAA TTT TCA GTA GCA AAA AAA ATA ATA GAA AAT GGT TGC CAA GAG
 241 pro gln gly trp ile ala ser ala arg leu glu glu lys ala arg lys phe ser val ala lys lys ile ile glu asn gly CYS GLN GLU

 811 TGC CCT CGA AGC TCC GAT ATT TGG CTA GAA AAC ATT AGA CTA CAC GAA TCT GAT GTT CAC TAC TGT AAA ACA TTA GTG GCA ACG GCA ATA
 271 CYS PRO ARG SER SER ASP ILE TRP LEU GLU ASN ILE ARG LEU HIS GLU SER ASP VAL HIS TYR CYS lys thr leu val ala thr ala ile

 901 AAT TTT AAT CCA ACG TCT CCG CTT CTT TGG TTC AAA GCT ATT GAT TTG GAA AGC ACA ACG GTT AAC AAA TAT AGA GTA GTG AGA AAA GCA
 301 asn phe asn pro thr ser pro leu leu trp phe lys ala ile asp leu glu ser thr thr val asn lys tyr arg val val arg lys ala

 991 CTG CAA GAG ATT CCT CGA GAT GAG GGC CTA TGG AAG CTA GCT GTC AGT TTT GAA GCT GAC AAA GCG CAA GTT ATA AAA ATG TTA GAG AAA
 331 leu gln glu ile pro arg asp glu gly leu trp lys leu ala val ser phe glu ala asp lys ala gln val ile lys met leu glu lys

1081 GCC ACA CAA TTT ATT CCA CAA AGT ATG GAT CTC TTG ACT GCA TAT ACT AAT TTG CAA AGC TAT CAT AAT GCT AAA ATG ACT TTG AAT TCC
 361 ala thr gln phe ile pro gln ser met asp leu leu thr ala tyr thr asn leu gln ser tyr his asn ala lys met thr leu asn ser

1171 TTC AGA AAA ATC CTT CCG CAA GAA CCG GAA ATT TGG ATT ATC TCT ACA CTC TTG GAA GAA CGA AAT AAC CCA GAT ATA CCT GTA GAT AAA
 391 phe arg lys ile leu pro gln glu pro glu ile trp ile ile ser thr leu leu glu glu arg asn asn pro asp ile pro val asp lys

1261 CTA GTT AGT TTG CTC AAG GAG GGT TTA TTG GAA CTC TCT AAA AAT GGG TAC AAA GCG ACC TTG TCA GCA TGG TTG AAA CGT GCA GAG GCT
 421 leu val ser leu leu lys glu gly leu leu glu leu ser lys asn gly tyr lys ala thr leu ser ala trp leu lys arg ala glu ala

1351 CTA AAT GAT GCG CCC AAT TCA AAT TTA ACC TGT CAA GCC ATC GTT TAC GCT ATA TTA GAA TGG TTA AGA GAA AGT GGC GAG TAT GAG TCT
 451 leu asn asp ala pro asn ser asn leu thr cys gln ala ile val tyr ala ile leu glu trp leu arg glu ser gly glu tyr glu ser

1441 GAG TTG AAT AAT GTT GAT CAG ATA TTA GAA AAA ATG CCA CAC TCA AAG GTA CAA ATT GCT GTC TTA AAA AAG CTT ATT CAG TGG GAT CCT
 481 glu leu asn asn val asp gln ile leu glu lys met pro his ser lys val gln ile ala val leu lys lys leu ile gln trp asp pro

1531 TGT GAT ACA GTT CTT TGG TCT AGA CTG AAA ATG GCC ACT GAA AGC TAC CAT AAA ATT GAA GAG TTA TTA GCA TTT TTC CAG GAG CTG CTA
 511 cys asp thr val leu trp ser arg leu lys met ala thr glu ser tyr his lys ile glu glu leu leu ala phe phe gln glu leu leu

1621 TTT CAG ACC AAG AAT AGT GAT GAT ATA CGA GCA AAT ATG AGG GAG AAA AGC CCT GGC TTG TTA ATG ATG TAT GTA AGC GAA TAT TGG AAG
 541 phe gln thr lys asn ser asp asp ile arg ala asn met arg glu lys ser pro gly leu leu met met tyr val ser glu tyr trp lys

1711 GCC CAA AAA GGG GAT ACT AGG CAA ACA CTA GTT TTG ATT GAC CAG ATT ATA GAT TTC GCC CCG CAT AAT TTG GAT TTA CGC TTT TTC AAG
 571 ala gln lys gly asp thr arg gln thr leu val LEU ile asp gln ile ile asp PHE ala pro his asn leu asp LEU arg phe phe lys

1801 ATA AAG TTA TTA GGT CGT TCA CTA CAA CTT GAT GAA TTA CGA GAT TTT TTT CAG CAA ACT TTC TCC TCT TTA GAG GAT TTT AAG ATC AGT
 601 ile lys LEU leu gly arg ser leu gln LEU asp glu leu arg asp phe PHE gln gln thr phe ser ser LEU glu asp phe lys ile ser

1891 GGC ACG GAA AGA TTA TAT TAT AAA TAC GTA AAC TTT CTG CGG TAC CAA GAT CTG AAT GAA GAG GCT ATA AAA TTC TTG AAT GAG AGA TGT
 631 gly thr glu arg leu tyr tyr lys tyr val asn phe leu arg tyr gln asp leu asn glu glu ala ile lys phe leu asn glu arg CYS
                                                      ----------------------------
1981 TTG AAA TCA TTT CCC ATC TGC CAC AAA TTT TTT TTA CAG CTG GGT CAA ATT TAT CAT TCC ATG GGC AAT ATT GAA ATG AGT AGA GAA ACC
 661 LEU LYS SER PHE PRO ILE CYS HIS LYS PHE PHE LEU GLN LEU GLY GLN ILE TYR HIS ser met gly asn ile glu met ser arg glu thr
                                                                               A
2071 TAT TTG TCT GGT ACA AGG TTA GTG CCC AAT TGC CCT TTA TTA TGG GTT TCC CTA TCA AAG ATT GAC GAG ATT GAT CTA AAA AAT CCA GTA
 691 tyr leu ser gly thr arg leu val pro asn cys pro leu leu trp val ser leu ser lys ile asp glu ile asp leu lys asn pro val

2161 AGG GCA AGA TCA ATT TTA GAT AGA GGA TTG TTA AAA AAT CCT GAC GAT GTA TTA TTT TAC ATT GCT AAA ATC CAA ATG GAA ATA AGA CTT
 721 arg ala arg ser ile leu asp arg gly leu leu lys asn pro asp asp val leu phe tyr ile ala lys ile gln met glu ile arg leu

      A                                                   A                       C
2251 GGT AAC TTA GAT CAG GCG GAG TTA CTC GTC ACA CAG GCA TTG CAA AAG TTT CCA AGC AAT GCT TTA CTT TGG GTG GAG CAA ATC AAG CTG
 751 gly asn leu asp gln ala glu leu leu val thr gln ala leu gln lys phe pro ser asn ala leu leu trp val glu gln ile lys leu
     asp                                                 thr
                                 A
2341 TTT AAG CAT GGA AAC AAA AGT TCG TTA AAA AAA ACA ATT TTT CAA GAT GCT TTA AGA AGG ACA CAA AAC GAT CAT CGC GTT CTT TTG GAG
 781 phe lys his gly asn lys ser ser leu lys lys thr ile phe gln asp ala leu arg arg thr gln asn asp his arg val leu leu glu
                                                                                                                 A
2431 ATT GGA GTA TCC TTT TAT GCA GAA GCG CAA TAT GAA ACA TCA TTA AAA TGG TTA GAA AGA GCT CTG AAA AAG TGC TCG CGT TAC GGA GAT
 811 ile gly val ser phe tyr ala glu ala gln tyr glu thr ser leu lys trp leu glu arg ala leu lys lys cys ser arg tyr gly asp
                                                                                                                 glu
2521 ACA TGG GTT TGG CTA TTT AGG ACA TAT GCA AGG TTA GGC AAG GAT ACT GTT GAT CTC TAC AAT ATG TTC GAT CAA TGT GAG CCT ACT TAC
 841 thr trp val trp leu phe arg thr tyr ala arg leu gly lys asp thr val asp leu tyr asn met phe asp gln cys glu pro thr tyr

2611 GGA CCC GAA TGG ATA GCC GCC TCC AAG AAC GTA AAA ATG CAA TAC TGC ACA CCT AGA GAG ATT TTA TTG CGC TTG ATG AAT GAC AAA TAA
 871 gly pro glu trp ile ala ala ser lys asn val lys met gln tyr cys thr pro arg glu ile leu leu arg leu met asn asp lys TRM
                                                 -----------------------
2701 AGTAATTCTA TCTTGCGTTG TTTTCTACTT AGCCTGCGCG TATATTTATA TGTATGTATG TATGTATACA
```

Fig. 1. *PRP6* gene sequence. Nucleotide and amino acid sequences of the *PRP6* locus are shown. Underlined nucleotides are sequences related to transcriptional signals (Zaret and Sherman, 1982; Henikoff and Cohen, 1984; Hahn *et al.*, 1985). Peptidic segments rich in basic amino acid residues are underlined once. The PW repeated motifs are underlined twice. Cysteine and histidine rich regions are in bold capital letters. Leu and Phe residues of the leucine repeat motif are in capital and underlined letters. The location of oligonucleotides used for PCR reactions are indicated by dotted lines above nucleotides. Mutant positions and their amino acid translations are given above the nucleotidic sequence and below the peptidic one.

Results

Sequences of the PRP6 and PRP9 genes

The cloning of *PRP6* and *PRP9* wild-type genes was made by complementation of the ts alleles with a yeast genomic library. The precise location of the genes in the cloned inserts was made by linker insertion introducing termination codons in the six frames (P.Legrain *et al.*, in preparation). A double strand sequencing strategy was used, taking advantage of the inserted linkers by using them as sequencing primers in

```
-583                                                                                          CCG TCCGTGGGAA ATTCAAGCAT AAACAAATGA

-550 AAGCGCCTAA AATGCCAGAT AAACACAGAG ACAATTATTA TTCGCAAAAG AAGAAGGTGG AAAAGGCTCT ACAAAGTGGA ATTTCCGTCA AGGGATATAA TAACGCTCCG

-440 GGTCTGCGAA GCGAATTGAA GTCCACTGAG CAAATTAGAA AGGACAGGAT TATAGCAGAG AAAAAACGTG CGAAGAATGC TCGTCCTTCC AAAAAGCGTA AATTTTAGAT

-330 AGTTTAAAAC CATCTAGATC TACAGATATA TAGTTTACAT ATGTACCATA GCAAGAAAAC AAAAATAGAT AGCGAGTCCG TTCTTCATAA TTCTTTGTAG TTCTTTGTGA

-220 TTTCAGTATT TTGTCGATGT AAGTCATGTT CAATATTTTT TTTGTAGCGC ATTTTATTCT AACTTTCCGA TGAGGCAACA AAGAAGTGTT AAGTATCGGC TGACGAAGTG
                                                                                                                ------------------
-110 CAATAAAAAG TACTTAGTTA TACGGTAGAT CAAGAACCTT CGGTGAGCCA AGCGAGGGTG TATACATTCA TTTAATTTTG CTTTAATAGG AGTAGGAGGG CATCATCTCA

   1 ATG AAT TTA CTT GAA ACA AGG AGG TCC TTG TTG GAG GAG ATG GAG ATT ATT GAA AAT GCC ATA GCA GAA AGA ATT CAG CGG AAT CCA GAG
   1 Met asn leu leu glu thr arg arg ser leu leu glu glu met glu ile ile glu asn ala ile ala glu arg ile gln arg asn pro glu

  91 TTA TAT TAC CAC TAT ATA CAA GAA TCG AGC AAG GTG TTT CCT GAT ACT AAA CTG CCT AGA TCA TCG TTG ATT GCA GAG AAT AAA ATA TAC
  31 leu tyr tyr his tyr ile gln glu ser ser lys val phe pro asp thr lys leu pro arg ser ser leu ile ala glu asn lys ile tyr

                                                                   A
 181 AAG TTT AAA AAG GTT AAG AGG AAG AGA AAA CAG ATA ATT TTG CAG CAA CAT GAG ATA AAT ATT TTT CTT CGA GAC TAC CAA GAG AAA CAA
  61 lys phe lys lys val lys arg lys arg lys gln ile ile leu gln gln his glu ile asn ile phe leu arg asp tyr gln glu lys gln
                                                                    lys

 271 CAA ACT TTT AAT AAA ATC AAT CGT CCA GAA GAG ACA CAG GAG GAT GAC AAG GAT TTG CCT AAT TTC GAA AGA AAA CTA CAA CAG CTA GAG
  91 gln thr phe asn lys ile asn arg pro glu glu thr gln glu asp asp lys asp LEU pro asn phe glu arg lys LEU gln gln leu glu

 361 AAG GAA CTG AAA AAT GAA GAT GAG AAC TTT GAA TTG GAT ATC AAC TCT AAA AAA GAC AAA TAC GCT TTA TTC TCA TCT TCT TCT GAT CCA
 121 lys glu LEU lys asn glu asp glu asn PHE glu leu asp ile asn ser lys lys asp lys tyr ala leu phe ser ser ser ser asp pro
                                                     ------------------------
 451 TCG AGG CGC ACA AAT ATA TTG TCT GAC AGA GCT CGA GAC CTA GAC TTA AAT GAA ATA TTT ACT AGA GAT GAG CAA TAT GGT GAA TAT ATG
 151 ser arg arg thr asn ile leu ser asp arg ala arg asp leu asp leu asn glu ile phe thr arg asp glu gln tyr gly glu tyr met

 541 GAG CTG GAA CAA TTT CAT TCT TTA TGG TTG AAT GTA ATT AAA CGG GGC GAT TGT TCA CTG CTT CAA TTT CTC GAC ATC CTA GAA TTA TTT
 181 glu leu glu gln phe his ser leu trp leu asn val ile lys arg gly asp cys ser leu leu gln phe leu asp ile leu glu leu phe

 631 TTG GAC GAC GAG AAA TAT TTG CTA ACC CCA CCG ATG GAT CGC AAG AAT GAT AGA TAC ATG GCC TTT TTG CTA AAG TTG AGC AAA TAT GTA
 211 leu asp asp glu lys tyr leu leu thr pro pro met asp arg lys asn asp arg tyr met ala phe leu leu lys leu ser lys tyr val

 721 GAA ACT TTT TTC TTC AAA AGT TAT GCT TTG CTT GAC GCT GCG GCA GTT GAA AAT CTA ATC AAA TCT GAC TTC GAA CAT TCA TAC TGT AGG
 241 glu thr phe phe phe lys ser tyr ala leu leu asp ala ala ala val glu asn leu ile lys ser asp phe glu HIS SER TYR CYS ARG

 781 GGA TCT CTT CGG TCC GAG GCA AAA GGT ATC TAT TGC CCT TTT TGT TCG AGG TGG TTC AAG ACA TCT TCC GTT TTC GAA AGC CAT TTA GTA
 261 GLY SER LEU ARG SER GLU ALA LYS GLY ILE TYR CYS PRO PHE CYS SER ARG TRP PHE LYS THR SER SER VAL PHE GLU SER HIS LEU VAL

 901 GGG AAA ATT CAT AAG AAA AAT GAA TCT AAA AGA AGA AAT TTT GTG TAC TCT GAA TAT AAA CTG CAT CGG TAT TTG AAA TAT TTA AAT GAT
 301 GLY LYS ILE HIS lys lys asn glu ser lys arg arg asn phe val tyr ser glu tyr lys leu his arg tyr leu lys tyr leu asn asp

 991 GAA TTT TCT CGA ACG AGA AGT TTT GTT GAA AGA AAA CTG GCA TTT ACT GCA AAT GAA AGA ATG GCA GAA ATG GAT ATC TTA ACA CAG AAG
 331 glu phe ser arg thr arg ser phe val glu arg lys leu ala phe thr ala asn glu arg met ala glu met asp ile leu thr gln lys

1081 TAT GAA GCA CCT GCA TAT GAT TCG ACG GAA AAA GAG GGG GCC GAA CAA GTG GAT GGT GAG CAG AGA GAT GGT CAA CTG CAA GAA GAG CAC
 361 tyr glu ala pro ala tyr asp ser thr glu lys glu gly ala glu gln val asp gly glu gln arg asp gly gln leu gln glu glu his

1171 CTC TCT GGT AAA TCG TTT GAC ATG CCA TTG GGT CCG GAT GGA TTG CCT ATG CCA TAC TGG CTA TAC AAA CTG CAT GGG CTT GAC AGA GAG
 391 leu ser gly lys ser phe asp met pro leu gly pro asp gly leu pro met pro tyr trp leu tyr lys leu HIS GLY LEU ASP ARG GLU

1261 TAT CGC TGC GAA ATT TGT TCG AAT AAA GTT TAT AAT GGG CGA CGC ACT TTT GAA AGA CAT TTC AAC GAA GAA AGA CAT ATT TAT CAC TTG
 421 TYR ARG CYS GLU ILE CYS SER ASN LYS VAL TYR ASN GLY ARG ARG THR PHE GLU ARG HIS PHE ASN GLU GLU ARG HIS ILE TYR HIS LEU

1351 CGA TGC CTT GGT ATC GAA CCT TCT TCA GTA TTC AAG GGC ATA ACC AAA ATT AAG GAG GCA CAA GAG CTC TGG AAA AAT ATG CAG GGG CAG
 451 ARG CYS leu gly ile glu pro ser ser val phe lys gly ile thr lys ile lys glu ala gln glu leu trp lys asn met gln gly gln

1441 TCA CAG TTG ACA TCT ATT GCA GCA GTT CCC CCA AAG CCT AAT CCT TCA CAA CTA AAA GTT CCT ACA GAA TTA GAA CTA GAA GAA GAA GAC
 481 ser gln leu thr ser ile ala ala val pro pro lys pro asn pro ser gln leu lys val pro thr glu leu glu leu glu glu glu asp

1531 GAA GAA GGA AAT GTA ATG AGT AAG AAG GTC TAC GAT GAA CTT AAG AAG CAA GGT TTG GTG TGA AATCTCTCTG TGATATGTAT ATTTGTTTGA
 511 glu glu gly asn val met ser lys lys val tyr asp glu leu lys lys gln gly leu val TRM

1624 TAGATAGCAG TTGTATGTAT TTATACTATT CTTAGTAATG ACCTTATTAC CCGTCTTGTA GAATTGAAAG GCGAAAAACA ATTTGAGGAC CCATATAATG GAAGAAAACG

1734 GACGTGAAAG TTAAAT
```

Fig. 2. *PRP9* gene sequence. Nucleotide and amino acid sequences of the *PRP9* locus are shown. A representation similar to that of Figure 1 is used.

both directions. Additional oligonucleotides were used to complete the sequence (see Materials and methods). Each of these two sequences encodes one single open reading frame (ORF) (Figures 1 and 2). Upstream of the ORFs, several AT rich regions similar to the TATA elements analysed by Hahn *et al.* (1985) were found (Figures 1 and 2). Also, downstream of the ORFs, sequences similar to those described by Zaret and Sherman (1982) or Henikoff and Cohen (1984) as elements involved in transcription termination were found. More significantly, however, when compared with each other, upstream and downstream sequences of PRP6 and PRP9 display homologies (Figure 3). Within the 30 nucleotides preceding the initiation codons, two stretches of five and 16 nucleotides are homologous in the two sequences: five out of five and nine out of 16 identities, respectively, the other seven nucleotides being conserved pyrimidines or purines. Downstream of the termination codon, a stretch of 33 nucleotides contains 22 identities between the two sequences.

The *PRP6* gene encodes a 899 amino acid long protein with an estimated mol. wt of 104 kd. This protein is highly charged (29.8% charged residues). Three sequences rich in basic residues are found in the N-terminal region of the protein (Figure 1). The PRP6 protein exhibits a proline-X_5-tryptophan motif (PW) found nine times in the protein. Segments of ~30 residues including the PW motif in the middle can be aligned and are shown in Figure 4. From this alignment, a loose consensus was built and was then used to screen for other repeats in the PRP6 protein. The best scores of homology with the consensus were in the nine PW repeats and another segment directly located between the fourth and the fifth repeat (Figure 4). Other regions of the PRP6 protein gave lower scores (data not shown). This strongly suggests that these segments define homologous structural domains. A leucine repeat motif (leucine occurring every seven residues, Landschulz *et al.*, 1988; see Figure 1) is also found between residues 582 and 624 and contains five leucines and two phenylalanines. Finally, two cysteine-

```
Upstream regions:

PRP6:   -30  GAGTTTAAC    AGCAAGAAAACGTCACATTTT ATG
                 =====-    ==-=-==---=-===-
PRP9:   -30     CTTTAATAGGAGTAGGAGGGCATCATCTCA  ATG

Downstream regions:

PRP6:      TAA ...17 nt...  TTGTTTTCTACTTAGCCTGCGCGTATATTTATATGTA
                            ======  ==  ====   -=--=-========-
PRP9:      TGA ...22 nt...  TTGTTTGATAGATAGCAGTTGTATGTATTTATACTAT
```

Fig. 3. Comparison of untranslated *PRP6* and *PRP9* gene sequences. Both upstream and downstream regions surrounding the *PRP6* and *PRP9* ORFs are shown. Identities (=) and purine or pyrimidine conservation (−) are indicated. The initiation and termination codons are underlined.

```
225     - KMRTILKSYRKA  DPTNPQG   WIASARLEEKARKFS
259       VAKKIIENGCQE  CPRSSDI   WLENIRLHESDVH
291       YCKTLVATAINF  NPTSPLL   WFKAIDLESTTV
322       NKYRVVRKALQE  IPRDEGL   WKLAVSFEADKA
353*      QVIKMLEKATQF  IPQSMDL   LTAYTNLQSYHN
385        AKMTLNSFRKI  LPQEPEI   WIISTLLEERNNP-
497     - VQIAVLKKLIQW  DPCDTVL   WSRLKMATESYHK-
686     - MSRETYLSGTRL  VPNCPLL   WVSLSKIDEIDLK-
755     - QAELLVTQALQK  FPSNALL   WVEQIKLFKHGNK-
855     - DTVDLYNMFDQC  EPTYGPE   WIAASKNVKMQYC-

CONS    -  AK-LL--A-Q-  -PT-P-L   WI-A-KLEE---K
               V
```

Fig. 4. PRP6 repetitive motif. The various PW motifs are aligned and the positions of the first residue of each are indicated. The repeat marked with an asterisk lacks the tryptophan residue but is homologous to the other repeats (see text). The consensus used to check for repeats in the PRP6 sequence and in the databases is shown below (residues present at least three times were retained). (−) indicates that the motifs are not contiguous. The representation of each repeat in three parts underlines the hypothesis of a helix−turn−helix (or helix−loop−helix) structure with proline rich segment in the middle (see text). Note that the Cys/His motif is included in the second PW repeat.

and histidine-containing regions with five and four such residues (268−292 amino acids; 660−679 amino acids, respectively; see Figure 1) are present, the first one being reminiscent of zinc finger structures (see below).

The *PRP9* gene codes for a 530 residue long protein with an estimated mol. wt of 63 kd. This protein is also highly charged (35.6% charged residues). One sequence rich in basic residues is found in the N-terminal region of the protein (amino acids 66−70, Figure 2). A leucine repeat motif, containing one phenylalanine residue, is located between amino acids 109 and 130. Finally, the protein also exhibits two segments containing cysteines and histidines (256−304; 415−432; see Figure 2), both of them being related to zinc finger structures (see below).

The two sequences have been compared with the protein identification resource (National Biomedical Research Foundation, NBRF) and the PGtrans database from the EMBL (Heidelberg) and we found no significant homology to any protein. However the presence of patterns similar to zinc finger structures prompted us to analyse these sequences more carefully.

In vitro mutagenesis on the cysteine/histidine containing regions

The PRP6 protein contains 22 cysteine or histidine residues, nine of which being clustered in two regions totalling 45 residues (Figure 1). Similarly, the PRP9 protein has 19 such residues and 13 are clustered in two regions of 39 and 38 residues (Figure 2). By visual inspection, structures similar to zinc fingers of the TFIIIA type can be drawn for three of them (Figure 5A). The more C-terminal cysteine/histidine containing region of PRP6 does not display a similar structure and was not studied further. *In vitro* mutagenesis has been performed on cysteine and histidine residues that are located at the conserved positions in both PRP6 and PRP9 proteins (Table I). Cysteines have been substituted for serines and histidines for leucines. Mutated genes were re-introduced into ts strains on a centromeric replicative vector (YCp50) and growth at 37°C was tested. In such cells, both the ts protein and the *in vitro* mutated protein are co-expressed. By this assay, none of the four mutated positions in PRP6 affects the complementation of the ts phenotype. In PRP9, mutation in the first histidine of the first region or mutation of any position of the second region completely abolishes the ability of the mutated gene to complement the ts phenotype (Table I).

Mapping of the ts mutations in the *prp6* and *prp9* mutants

First, *prp6* and *prp9* ts mutations were mapped by deletion: non-complementing plasmids derived from pPL1 or pPL4 were obtained by deletion. They were re-introduced as replicative plasmids in YCp50 vector and assayed for their ability to generate temperature-resistant cells by recombination. Results are presented in Table II: for PRP6, only the *Pvu*II deletion (ΔP) allowed the obtention of colonies growing at 37°C; for PRP9, transformation with two deleted plasmids (ΔEV and ΔSc) led to the production of colonies growing at 37°C. These experiments defined regions in the *PRP6* ORF (2016−2697 nucleotides) and the *PRP9* ORF (1−403 nucleotides) where the ts mutations should be located (Table II). Then, using a polymerase chain reaction (PCR) methodology, these fragments of genomic DNAs derived from *prp6* and *prp9* ts strains were cloned and sequenced (see Figures 1 and 2). Sequence comparisons showed one nucleotide difference between *prp9* ts and the wild-type allele (nucleotide 232, changing a glutamate residue into lysine, see Figure 2) and six nucleotide differences between *prp6* ts and its wild-type allele (Figure 1); in this latter case, three differences are silent, whereas the others change two glycines into aspartate (nucleotide 2252) or glutamate (nucleotide 2516) and an alanine into threonine (nucleotide 2287). This high number of nucleotide divergences between two alleles (1% of sequenced positions) could reflect polymorphism among strains which originate from different laboratories.

Discussion

PRP6 and *PRP9* genes encode an 899 amino acid and a 530 amino acid protein, respectively. Searches in databases reveal no significant homologies to any known protein. Upstream and downstream of the ORFs, the two genes display similarities that are not found in other yeast genes (Figure 3). This observation suggests that the *PRP6* and *PRP9* genes may have a common regulation for transcriptional and/or translational expression. This finding is reinforced by the mapping of the 5′ end of the most abundant *PRP9* mRNA which is located at position −11 relative to the initiation codon (E.Schwob and R.Martin, personal communication); this localization falls into the homologous upstream region.

PRP6 PW repetitive motif

The consensus made from the nine PW repeats found in PRP6 reveals an additional repeat which lacks the

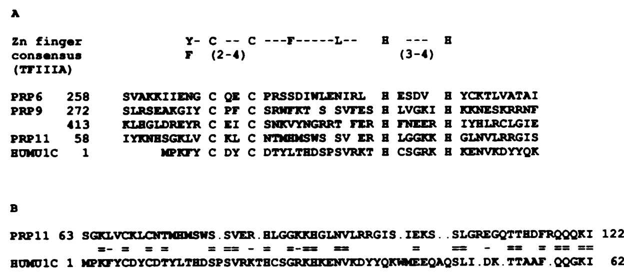

Fig. 5. Alignment of cysteine and histidine containing regions. (**A**) Sequences from PRP6, PRP9 and PRP11 (Chang *et al.*, 1988) and human U1C (Sillekens *et al.*, 1988) are aligned and positions of the first residue of the alignment are given on the left. Above is indicated the consensus sequence found in zinc finger proteins such as TFIIIA (Vincent, 1986). (**B**) An alignment of PRP11 and human U1C proteins is proposed between the positions shown on both sides of the sequences. Identical (=) or homologous (−) residues are indicated.

tryptophan residue but is exactly located between the fourth and the fifth repeat (Figure 4). The first six repeats are contiguous and define a 184 residue long domain which thus seems highly structured. The same consensus was used to screen protein databases. No protein contains repeats that have similar scores to those of the PW repeats in the PRP6 protein. The absence of proline residues on both sides of the PW motifs suggests some similarities with the helix−turn−helix motif (reviewed by Struhl, 1989) or with the amphipathic helix−loop−helix (HLH) motif described by Murre *et al.* (1989). The middle part (i.e. the proline-containing segment) can form a β turn or a loop similar to the Ω loop described by Leszczynski and Rose (1986) in which residues such as Pro, Asp, Asn, Ser and Thr are often found (Figure 4). On both sides of the proline rich segment, α helical structures are predicted by secondary structure search programs. Short amphipathic helices can be drawn but they are difficult to align from one repeat to another. These similarities suggest that the PRP6 PW motif may have a structure related to the HLH motif. If this is the case, one would predict that helices are buried in the protein and the loops are exposed at the surface of the protein (Leszczynski and Rose, 1986). It has been shown that such regions are necessary for DNA binding in some proteins (Murre *et al.*, 1989), but the HLH structure by itself could be an important element for the tertiary structure of proteins and for intermolecular interactions (Leszczynski and Rose, 1986). In that respect, further work will be needed to establish the respective importance of helices and loops of the PW motif.

Leucine repeats

PRP6 and PRP9 proteins both contain one leucine repeat motif which includes phenylalanine residues (Figures 1 and 2) and these motifs can be drawn on typical 4/3 α helices (O'Shea *et al.*, 1989; Figure 6). PRP6 displays a hydrophobic side surrounded by positively charged residues whereas PRP9 displays a similar hydrophobic alignment with negatively charged residues on both sides. Leucine repeat motifs are implicated in homo and hetero dimerization processes (for review see Struhl, 1989). Since *prp6* and *prp9* ts mutants displayed similar *in vivo* phenotypes on pre-mRNA splicing and export to the cytoplasm, we have suggested that PRP6 and PRP9 proteins could interact in the same complex (Legrain and Rosbash, 1989). The dimerization of PRP6 with PRP9, using their leucine repeat

Table I. *In vitro* mutagenesis of cysteine[a] and histidine[a] residues in PRP6 and PRP9 genes

	Growth at 37°C		Growth at 37°C		Growth at 37°C
PRP6 Cys268	+	PRP9 Cys282	+	PRP9 Cys423	−
Cys271	+	Cys285	+	Cys426	−
His285	+	His298	−	His440	−
His290	+	His304	+	His446	−

[a]Cysteine and histidine residues have been changed into serine and leucine residues, respectively.

Table II. Mapping by deletion of *PRP6* ts and *PRP9* ts mutations

			Exp. 1[a]	Exp. 2
PRP6	ORF length	2967 nucleotide		
	Δ*Bam*HI	1530−2697[b]	0/141	0/172
	Δ*Cla*I	1−2697	0/96	ND
	Δ*Eco*RI	1170−2697	0/150	0/188
	Δ*Hin*dIII	1516−2697	0/99	0/72
	Δ*Pvu*II	1−2016	6/56	47/238
Mutation mapping		2016−2697		
PRP9	ORF length	1590 nucleotide		
	Δ*Eco*RV	403−1062	3/70	2/75
	Δ*Nde*I	1−1093	0/170	0/142
	Δ*Sac*I	485−1413	13/100	5/247
Mutation mapping		1−403		

[a]Results are expressed as number of colonies which exhibit growth at 37°C (as papilla) compared to the total number of colonies grown at 25°C.
[b]Nucleotide numbering indicates the starting and ending point of deletions compared to ORF numbering.
ND, not done.

motifs, can be proposed (Figure 6). In addition to the binding of hydrophobic residues to each other, the lateral axes with either acidic (PRP9) or basic (PRP6) residues may contribute to the interaction. In any case, PRP6 protein displays several features (PW motif repeats, leucine repeat and Cys/His motif) that suggest that PRP6 has numerous interactions with other macromolecules (proteins or nucleic acids) as could be expected for a snRNP component or a protein involved in large complexes such as spliceosomes.

Cysteine and histidine containing motifs

Both PRP6 and PRP9 proteins contain two regions rich in cysteine and histidine residues, compared with the other parts of the proteins. Three of them can be aligned, some of the cysteines and histidines being at conserved positions (Figure 5A). The comparison with zinc finger proteins of the TFIIIA type (Vincent, 1986) shows that PRP6 and PRP9 do not share residues other than cysteines and histidines with the consensus sequence (Figure 5A). However, a systematic search for similar motifs in other proteins reveals that at least two proteins have sequences similar to the one displayed by PRP6 and PRP9 (Figure 5A): PRP11 (Chang *et al.*, 1988; the authors noticed the presence of a zinc finger-related structure) and the human U1C (Sillekens *et al.*, 1988) which are both implicated in the splicing process (see below).

Cysteine or histidine residues in the PRP6 and PRP9 motifs have been substituted by *in vitro* mutagenesis. For

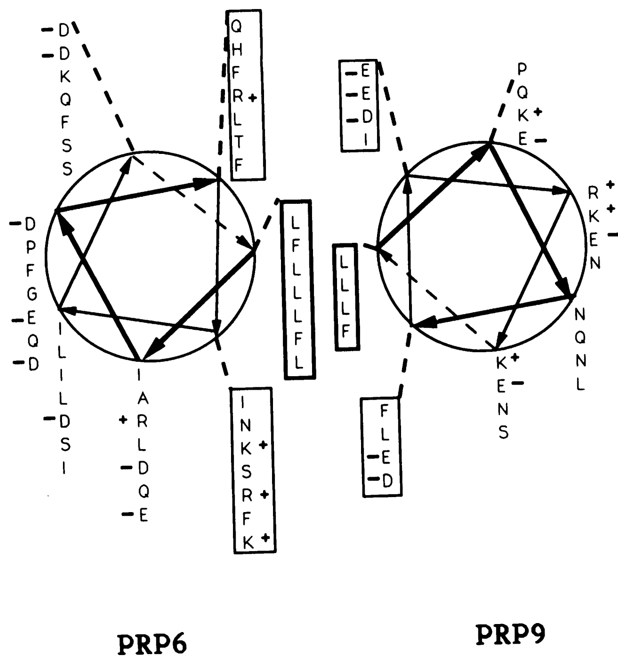

Fig. 6. Leucine repeat motifs and PRP6 and PRP9 proteins. Sequences of residues 582–630 of PRP6 and residues 109–136 of PRP9 are presented on helical wheels to reflect a putative α helix configuration (Landschulz *et al.*, 1988). However the presence of one proline residue close to the extremities of both helices is quite unusual for helical structures. Hydrophobic axes and positively charged (PRP6) or negatively charged (PRP9) residues on surrounding axes are boxed. Helices are presented in parallel configuration, but anti-parallel configuration is also possible.

five positions in the PRP9 protein, the substitution destroys the function of the protein (Table I). For the other positions, namely three out of the four residues of the N-proximal motif in PRP9 and the four residues in the PRP6 motif, the mutated genes did restore the growth at 37°C. Several hypotheses can be made to explain this apparently ambiguous result. First, the motifs contain other cysteine or histidine residues at non-conserved positions (see Figures 1 and 2) which could participate in the function carried out by the motifs. The substitution of other residues or the combination of two or several substitutions may be necessary to abolish this function. Secondly, the function of the protein could be largely affected, but not sufficiently to forbid the growth of the cells at 37°C. Thirdly, the complementation assay allows a biochemical complementation between the ts protein and the *in vitro* mutated protein. It is possible that these latter proteins are not functional *per se*, but that they can complement either in a homodimer or as two independent functional domains (intragenic complementation). So, the function of the mutated *prp6* or *prp9* genes should be assayed in the absence of any other PRP6 or PRP9 protein, respectively, to eliminate the possible biochemical complementation. However, the fact that the substitution of any residue of the C-proximal motif of PRP9 abolishes the restoration of a wild-type phenotype demonstrates that essential residues for the function of PRP9 have been identified.

A careful examination of the five aligned motifs reveals that, in addition to the conserved cysteines and histidines, some other residues are preferentially but not always found at certain positions (Figure 5A). Moreover, similarities between U1C and PRP11 proteins expand on both sides of the motif: the two proteins can be aligned over 60 residues (Figure 5B) which suggests that these sequences define a similar functional domain. Other parts of the proteins, such as the strikingly proline rich domain of U1C which does not exist in PRP11, may be involved in interactions that implicate factors which have largely diverged between yeast and mammalian cells. In that respect, one should keep in mind that the yeast U1 snRNA is noticeably larger than and different from its mammalian counterpart (Kretzner *et al.*, 1987) and that very little is known about yeast U1 snRNP proteins. It is possible that PRP11 and U1C are functionally homologous proteins.

It is not known at what step of the splicing pathway the PRP6 and PRP9 proteins play a role. However, in a previous work, we showed that they were implicated, as well as U1 snRNP, in stably committing pre-mRNAs to the splicing pathway (Legrain and Rosbash, 1989). So the structural similarities found between PRP6, PRP9, PRP11 and U1C may indicate that they participate in a common function and that PRP6 and PRP9 interact together with U1 snRNP. PRP6 and PRP9 could be involved in the recognition process of the intron in the pre-mRNA. Several works have shown that U1 snRNP binding to the pre-mRNAs is an early step in the *in vitro* or *in vivo* splicing pathway (Ruby and Abelson, 1988; Zillmann *et al.*, 1988; Legrain and Rosbash, 1989; Seraphin and Rosbash, 1989).

The present work reveals that the Cys_2His_2 zinc finger structure found in DNA binding proteins has a similar counterpart in some proteins involved in RNA processing; these new Cys/His motifs are less closely related to those found in DNA binding proteins than to each other and could be involved in the binding of the proteins to RNA. A consensus has been described for many RNA binding proteins, including hnRNP and several snRNP proteins (for review see Bandziulis *et al.*, 1989). None of the four proteins sharing the cysteine/histidine motif described here, displays this consensus. There is no evidence for direct binding of PRP6 or PRP9 to either pre-mRNAs or snRNAs. However, it has been shown that PRP11 is associated with 30S complexes and spliceosomes and can be immunoprecipitated with anti-Sm antibodies, suggesting that PRP11 could be a snRNP constitutive peptide (Chang *et al.*, 1988). Moreover, U1C is a peptide specifically associated with the U1 snRNP and plays a crucial part in the efficient binding of the U1 snRNP to the 5′ splice site of pre-mRNAs (Heinrichs *et al.*, 1990). The relationship between PRP11 and U1C remains to be clarified.

Mapping of prp6 and prp9 ts mutations

prp6 and *prp9* ts mutations have been cloned and sequenced. In both cases, mutations are located outside the leucine repeat or the cysteine/histidine regions. In the case of *prp6* mutation, one of the three amino acid residues divergent from the wild-type sequence is located in one of the PW repeats which cover more than one-third of the protein (Figure 1). It is striking that the mapping of the ts mutations reveals functional domains whose importance was not underlined by the sequence analysis. Further analyses should give us more information on these mutated domains. The availability of cloned *PRP6* and *PRP9* wild-type genes and of their thermosensitive alleles allows us to address the questions of PRP6 and PRP9 functions by *in vivo* and *in vitro* analyses.

Materials and methods

Strains and plasmids

SpJ6.66 (*prp6* ts) and Jm664 (*prp9* ts) have been described in a previous work (Legrain and Rosbash, 1989). Conditions for yeast cultures have already been described (Legrain and Rosbash, 1989). YCp50 plasmid (Rose *et al.*, 1987) and pBluescript SK^+ vector (Stratagene, CA) were used for the different cloning steps according to standard procedures (Maniatis *et al.*, 1982).

Sequencing strategy

Double strand sequencing was performed either on YCp50 or SK^+ derived plasmids. Sequencing procedures with various sequencing kits were followed according to the manufacturers. Several primers have been synthesized to complete sequences on both strands and to overlap all linker insertion sites.

In vitro mutagenesis

In vitro mutagenesis was performed according to Kunkel (1985). Single strand DNA was recovered from SK^+ derived plasmids grown in a *dut*$^-$, *ung*$^-$ *Escherichia coli* strain. 17–19 nucleotide long oligonucleotides were used to introduce the mutated position. Sequencing four clones from one *in vitro* mutagenesis was sufficient to recover at least one mutant (more often three or four). Sequence of the mutant position was systematically controlled after subcloning in the yeast expression vector.

Polymerase chain reactions

Yeast genomic DNA was prepared from SpJ6.66 and Jm664 strains from 2 ml cultures by standard procedures (including one phenol extraction and vortexing with glass beads). Oligonucleotides were designed with additional cloning sites at the ends. PCR reactions were performed with 3–10 μg genomic DNA or 5 ng plasmid DNA for the control reaction with wild-type DNA. A Perkin-Elmer Cetus thermocycler was used and reactions were set up as recommended by the manufacturer with 2.5 mM $MgCl_2$. Reaction mixtures were subjected to repeated cycles (15 s at 94°C, 1 min at 55°C and 2 min at 72°C), 20–40 times, without noticeable difference in the recovery of amplified DNA (>1 μg). This suggests that, in these conditions, 20 cycles are sufficient to reach the maximum DNA synthesis. It should be pointed out that the higher the input of genomic DNA, the lower the heterogeneity of PCR-amplified DNA.

Computer analysis

An extensive use of the various menus of DNA Strider™ was made in the course of this work (Marck, 1988). In addition, sequence comparisons were made to nucleotidic (Genbank and EMBL) and protein (NBRF and PGtrans) databases through the central computer unit at the Pasteur Institute.

Acknowledgements

We are greatly indebted to B.Dujon for his encouragement and our numerous stimulating discussions. We thank Y.Esnault who contributed to some experiments during a training program at the University Pierre et Marie Curie, E.Schwob and R.Martin for communicating unpublished results and M.Rosbash, B.Seraphin and B.Dujon's laboratory personnel for helpful discussions and critical reading of the manuscript. This work was partly supported by the CNRS (URA 1149) and the University Pierre et Marie Curie (Paris, France).

References

Anderson,G.J., Bach,M., Lührmann,R. and Beggs,J.D. (1989) *Nature*, **342**, 819–821.
Bandziulis,R.J., Swanson,M.S. and Dreyfuss,G. (1989) *Genes Dev.*, **3**, 431–437.
Brody,E. and Abelson,J. (1985) *Science*, **228**, 963–967.
Chang,T.-H., Clark,M.W., Lustig,A.J., Cusick,M.E. and Abelson,J. (1988) *Mol. Cell. Biol.*, **8**, 2379–2393.
Cheng,S.C. and Abelson,J. (1987) *Genes Dev.*, **1**, 1014–1027.
Frendewey,D. and Keller,W. (1985) *Cell*, **42**, 355–367.
Green,M.R. (1986) *Annu. Rev. Genet.*, **20**, 671–708.
Hahn,S., Hoar,E.T. and Guarente,L. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 8562–8566.
Hartwell,L.H., McLaughlin,C.S. and Warner,J.R. (1970) *Mol. Gen. Genet.*, **109**, 42–56.
Heinrichs,V., Bach,M., Winkelmann,G. and Lührmann,R. (1990) *Science*, **247**, 69–72.
Henikoff,S. and Cohen,E.H. (1984) *Mol. Cell. Biol.*, **4**, 1515–1520.
Konarska,M.M. and Sharp,P.A. (1986) *Cell*, **46**, 845–855.
Kretzner,L., Rymond,B.C. and Rosbash,M. (1987) *Cell*, **50**, 593–602.
Kunkel,T.A. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 488–492.
Landschulz,W.H., Johnson,P.F. and McKnight,S.L. (1988) *Science*, **240**, 1759–1764.
Legrain,P. and Rosbash,M. (1989) *Cell*, **57**, 573–583.
Legrain,P., Seraphin,B. and Rosbash,M. (1988) *Mol. Cell. Biol.*, **8**, 3755–3760.
Leszczynski,J.F. and Rose,G.D. (1986) *Science*, **234**, 849–855.
Lossky,M., Anderson,G.J., Jackson,S.P. and Beggs,J. (1987) *Cell*, **51**, 1019–1026.
Lührmann,R. (1988) In Birnstiel,M.L. (ed.), *Structure and Function of Major and Minor Small Nuclear Ribonucleoprotein Particles.* Springer-Verlag KG, Heidelberg, pp. 71–99.
Lustig,A.J., Lin,R.-J. and Abelson,J. (1986) *Cell*, **47**, 953–963.
Maniatis,T. and Reed,R. (1987) *Nature*, **325**, 673–678.
Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning. A Laboratory Manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Marck,Ch. (1988) *Nucleic Acids Res.*, **16**, 1829–1836.
Murre,C., Schonleber, McCaw,P. and Baltimore,D. (1989) *Cell*, **56**, 777–783.
O'Shea,E.K., Rutkowski,R. and Kim,P.S. (1989) *Science*, **243**, 538–542.
Pikielny,C.W., Rymond,B.C. and Rosbash,M. (1986) *Nature*, **324**, 341–345.
Pinto,A.L. and Steitz,J.A. (1990) *Proc. Natl. Acad. Sci. USA*, **86**, 8742–8746.
Rosbash,M., Harris,P.K.W., Woolford,J.L., Jr and Teem,J.L. (1981) *Cell*, **24**, 679–686.
Rose,M.D., Novick,P., Thomas,J.H., Botstein,D. and Fink,G.R. (1987) *Gene*, **60**, 237–243.
Ruby,S.W. and Abelson,J. (1988) *Science*, **242**, 1028–1035.
Ruskin,B., Zamore,P.D. and Green,M.R. (1988) *Cell*, **52**, 207–219.
Seraphin,B. and Rosbash,M. (1989) *Cell*, **59**, 349–358.
Sillekens,P.T.G., Beijer,R.P., Habets,W.J. and van Venrooij,W.J. (1988) *Nucleic Acids Res.*, **16**, 8307–8321.
Struhl,K. (1989) *Trends Biochem. Sci.*, **14**, 137–140.
Vijayraghavan,U. and Abelson,J. (1990) *Mol. Cell. Biol.*, **10**, 324–332.
Vijayraghavan,U., Company,M. and Abelson,J. (1989) *Genes Dev.*, **3**, 1206–1216.
Vincent,A. (1986) *Nucleic Acids Res.*, **14**, 4385–4391.
Whittaker,E., Lossky,M. and Beggs,J.D. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 2216–2219.
Woolford,J.L., Jr (1989) *Yeast*, **5**, 439–457.
Zamore,P.D. and Green,M.R. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 9243–9247.
Zaret,K.S. and Sherman,F. (1982) *Cell*, **28**, 563–573.
Zillmann,M., Zapp,M.L. and Berget,S.M. (1988) *Mol. Cell. Biol.*, **8**, 814–821.

Received on May 3, 1990

Note added in proof

The nucleotide sequence data reported here will appear in the EMBL, GenBank and DDBT nucleotide sequence databases with the accession numbers X53465 and X53466 for *PRP6* and *PRP9*, respectively.