

# The yeast PRP2 protein, a putative RNA-dependent ATPase, shares extensive sequence homology with two other pre-mRNA splicing factors

Jiann-Hwa Chen and Ren-Jang Lin

Department of Microbiology, University of Texas at Austin, Austin, TX 78712–1095, USA

Submitted September 21, 1990

EMBL accession no. X55936

We have shown previously that the *PRP2(RNA2)* gene product of the yeast, *Saccharomyces cerevisiae* is required to convert the pre-mRNA to linear exon 1 and lariat intron-exon 2 in the spliceosome (1). Here we report the nucleotide sequence of a 3177 bp EcoRI/BamHI DNA fragment within which the *PRP2* gene is located (2–4; Figure 1). An open reading frame encoding a protein of 876 amino acids is found (Figure 1). The predicted molecular weight of 100 kDa is consistent with that of the PRP2 protein detected immunologically (5, 6). Surprisingly, the PRP2 protein sequence shares extensive homology with two other yeast pre-mRNA splicing factors, PRP16 (7) and PRP22 (8; J. Abelson, personal communication). These proteins share about 25% overall amino acid sequence identity, and about 42% within a stretch of approximately 400 residues which contains a putative RNA-dependent ATPase domain (7, 9; Figure 1). This is very intriguing since each of the three PRP proteins is involved in a distinct ATP-dependent step in pre-mRNA splicing. Furthermore, the DEAD sequence found in ATP-dependent RNA helicases (9) is replaced by DEAH in these PRP proteins. A putative zinc finger motif, which might be involved in RNA binding, is also found in the PRP2 protein sequence (10; Fig.

1). We are currently characterizing the PRP2 protein in vitro for these properties.

## ACKNOWLEDGEMENTS

We thank J. Woolford for the PRP2 plasmid, as well as J. Abelson (PRP22) and J. Beggs (a fragment of the PRP2) for sharing sequence data prior to publication. This work is supported by grants from the NIH and the Welch Foundation to R.-J.L.

## REFERENCES

- Lin,R.-J., Lustig,A.J. and Abelson,J. (1987) *Genes and Development* **1**, 7–18.
- Last,R.L., Stavenhagen,J.B. and Woolford,J.L.Jr. (1984) *Mol. Cell. Biol.* **4**, 2396–2405.
- Lee,M.G., Young,R.A. and Beggs,J.D. (1984) *EMBO J.* **3**, 2825–2830.
- Soltyk,A., Tropak,M. and Friesen,J.D. (1984) *J. Bacteriol.* **160**, 1093–1100.
- Last,R.L. and Woolford,J.L.Jr. (1986) *J. Cell Biol.* **103**, 2103–2112.
- Lee,M.G., Lane,D.P. and Beggs,J.D. (1986) *Yeast* **2**, 56–67.
- Burgess,S., Couto,J.R. and Guthrie,C. (1990) *Cell* **60**, 705–717.
- Vijayraghavan,U., Company,M. and Abelson, J. (1989) *Genes and Development* **3**, 1206–1216.
- Linder,P., Lasko,P.F., Ashburner,M., Leroy,P., Nielsen,P.J., Nishi,K., Schnier,J. and Slonimski,P.P. (1989) *Nature* **337**, 121–122.
- Evans,R.M. and Hollenberg,S.M. (1988) *Cell* **52**, 1–3.

1 AAGCTTCGGTAAAGCTGGAAAAGCAGTTAGTAGCTACCGCCGACAATTAAAAGTGATAATTAACATTGTTAACAGTGTATGTTAATGCTGGTATCTACAGGTCATACCCGATGACAGAG  
204 ATGTCAGTATTACATCCTGAAACCGTGTAGCCGAAAGTAGTGAACGACAGTGTAGACTTACAGCAAGCATAGCTGATGAAACGATGCAACCCATGGGCAACCGTGTAGATAAG  
H S S I T M T G R R E V T P D H A V R I E Y E V T R S L G E E E D E K R E A K D K 45  
339 AACCTGGCTTCAGCTCAAAGCATGACGGCTATGACCCCCATAAAGTTTCTTAACTAGCGAACCGAGCGAAAGATGGATCACAAAGGTGATGCATCAAGATGTAGAA  
N S A L Q M K C R S F T S T M G P F K H N E L Q G E O C S C K E S Y D E 90  
474 AAAATTAACGCTAAATAATGAACTTACTAACAAAAGAGGTCTTACTAGGTGACTGAGGTGACACTTAACTACCTCTTAACTTAACTAGGATTTACTCATAAAAATAAGAATCCAAAGAGAAT  
K I T S M V E D I D L R C Y A A T V Y A M S R D P D G V D V V E V M R E K T 135  
609 SACCAATAATACAGAAAACGGATGTGGAGAAGAACAAATGGAAATTCTACCGCCGAAAGCAGTGTACCTGATGATAACATTCGGGACTCTGTAGCAGACAGCTGAACTG  
6 D E N H M D P D C G M S C D S D H D T I L V V D Y V F D T D A M 180  
744 ATTGCAATGATCTAACGGAGGACGATTACTACCTCTGAAAGAACTCTAAATGGCCACGCTTACGGACATTTAGAACACTTACGAAAGAAACTTTCGCGCTC  
I D Y T N E E D D I L P E E K L O Y A T E R A Q A L E T E E R I L T I O E K R I L L P V 225  
879 CATCACAAACGATGGCTACTACAGGAACTGAAATTTGGGGATTTGGGTGAAACCGGGATGGGGATGGGGGTTGGGTGATGGGTTAGGGTTAGGGGGAG  
H Q K D E L L Q E I K K Q V L I V A P E V A F V R E G S G K T T O L P Q Y L V E D G F T D Q G K 270  
1014 TTACAGATGCCATAACAAACAGCTGGTGTGGCGGCCACTCTGTTGTCTGCCAGAACTGAGATGAAATGATCTGGTCTGGGGAACTGGCTCTGAGGGGAA  
L O I T O P R E V A T S V A F V R E G S G K T T O L P Q Y L V E D G F T D Q G K 315  
1149 AAAACGGTTCATAAGATATGACGGCAGCTGGTGTAACTAGAGAAATTCTTACAGATTCCAATAAACGAAATTATCTGGTGTGTAGGATGAACTGAGCTTACGCAAGAGATACTGGTGTAGGT  
K T W L I Y T M T D G M L L R E C H S L S K M R S C I M D C H E R T L A T D I L G 360  
1284 TTGGCTGAAGATAATTGGCCGAGGGGCCACTTTAAATGGCTTACCATCACGCAACATGAACGCAAAATTGAGCTTTGGTCTGGGAGGAGATACTCTGTA  
L L K D I P O R P T L L K L L I S A T M N K F S E F F D M N C P T F P G G R R Y F V 405  
1419 GATATCCTACATTAATCAGCCGGGGCAAACTATACATACGGCTGTAACTAACACTTTCGCTTAATCAGTACTACGCCAAAGTTCACCGGGGGGATATACTGGCTTTCAGCCGCAAGGAAATTGAAAGA  
D I H Y T F C E A T T I I H A A I T T I F Q I D T E P L A N D I L V F L T Q G E E E I E R 450  
1554 ACAACAAACCAAACTGGGAGGAGATCATGGTCAAAATTAGTGTGACAAACAAATGATTAACTCTTACATCCAAATTTCACGCCAACTGGGAAATTCTGG  
T K T K L E E I M S K L G S R T E Q N I I T P I Y A A N L P Q E Q L K I F Q P T E F D K T P N C R 495  
1689 AAAGTGTGTGGCAACATAATTGGGCCAACTTCAATGATGATTGTGATTGATGTCATGGTGTGTGAAAGGAAATTCTGGCTACACTGGCAACAGATACTGGTGTAGCT  
E V V L A T T H A T A E T S L T I D G I R Y V D P F V R E G V S T G M T Q L L T V P 540  
1824 TTGTCAAGACTCTAGTTGATGGCAGGCGCCGGCTAGGCGCTGGTGTGGCCGGGGAAAGTGTGGTGTGGCAAGTGTGGCTACACCTGGCTAAGGGAGATACTGGTGA  
C S R A V S D O R A G R A G R E C P G K E R F I T V R W S Y L H E L E P K F E I T T 585  
1959 ATCTCTCTAAATCAGGTTATTATTATTATTACCTGCTGGCTACTGACCTAAATAAGTGTGCGGATAGTCCTGGGAACTTACCGAGCTTTCGGGAGGATACTGGCTTAAAT  
N L S N T V L L L D T D L I F K R L K S D M K P F L R E S L R Y E D I Y L G A L N 630  
2094 AGCAAAAGGGCAACACTACCTGGGTGGTTAAATGATGTCAGGTTCTGTAACCTGTCACGGGGGGCTACCGGAGCAGAACTTCAGCTGTGACACGTCAGGTGATGAGGTTGGGCAACCTGG  
S K D R H T G C C T G C F I S G F P H V I V O L G P T G Y O T M G R S S G G L N V V S V H 675  
2229 TGATGCTCATGAAACACTTCTATTATGGCTAACGAGGCTGGGCGGCGGCGGCTGGGCGGCTGGGCGGCTGGGCGGCTGGGCGGCTGGGCGGCTGGGCGGCTGGGCGGCTGGGCGGCTGGGCGGCTGG  
S M Z H T C P D D R G D A S C D T S C R D E S H D Y L E I F N O R M R N S K F S 720  
2364 AGACCTGGCTGGGAGGAGCTGCTACGATGTCATGGCTGGGCGGCTGGGCGGCTGGGCGGCTGGGCGGCTGGGCGGCTGGGCGGCTGGGCGGCTGGGCGGCTGGGCGGCTGG  
S C O D H E K I T R V R N I R M O L F C S E K V G L V E K P F E K I F T C D L V I P K E E W L 765  
2499 ATGCTGGGTCTACATAAATGCTGAAATGCTGGGTCTACAGCCGGGTTCCTGGTGAATATTGCTACATTTGCTACATGGCTAGCTGGTCTACAGCTGGCTACGGTTAGGTGTCTCAT  
H I G Y I N A R T R C F I S G F P H V I V O L G P T G Y O T M G R S S G G L N V V S V H 810  
2634 CCTCTGGCTACATGTCCTGGTAACTCATAGAAAGGCGGGAGGACCAACGAAATGATGGTGTCTGGCTACGTGAGGGAAATTTCAGGGCTGCTGTAGTTCCTCTGGGAGAATGTC  
P T S I L F V N H K E K A Q R P S K V Y V L Y Q O L L N L T S K E F I R D C L V I P K E E W L 855  
2769 ATAGACATGTTCCCACAAATCTTAAATGATGAACTGAGATGATACTAGATTTAGCTGATGGATGGTGTGTGATGTCATGGCTACGGCTCATCTCATGATGGCTACGG  
I D D V P G I F K D L I D D E T H R G R \* 876  
2904 TGGTAGCTGGCTATGTTATATATATATAGCTGAACTGAGATGATACTAGATTTAGCTGTTGCTGAACTTGCGCTAAAGGATCTGCAAGCAGCTCCTCCTAAATTGCGTGTGTGCTGG  
3174 ATCCCTGCTCTCTGCTGACTCTTACTAAATACTCTGTTATCTGCTCTTCAACAGATGCTAACTGAACTCTGGTCTGCTCTGGTCTGCTCTGGTCTGCTCTGGTCTGCTCTGG

Figure 1. The DNA and deduced protein sequences of the yeast *PRP2* gene. Plasmid pJD207/RN2–1 was obtained from J.Woolford, and the 3.2 kbp EcoRI/BamHI fragment containing the *PRP2* gene was subcloned into pUC18. DNA sequence was determined by the dideoxy method using synthetic oligonucleotides as primers. The number of nucleotide is given on the left and the number of amino acid on the right. Motifs found in a family of RNA-dependent ATPase are underlined (7, 9). The cysteine and histidine residues in the putative zinc-finger are boxed.