A putative homolog of U2AF⁶⁵ in *S.cerevisiae*

Ewan Birney, Sanjay Kumar and Adrian R.Krainer*

Cold Spring Harbor Laboratory, PO Box 100, Cold Spring Harbor, NY 11724-2208, USA

Submitted June 22, 1992

While searching for conserved motifs in protein factors involved in pre-mRNA splicing we retrieved an open reading frame, YCL11c from the PIR database (accession S19338) derived from *Saccharomyces cerevisiae* chromosome III (1). The authors noted its similarity to *Xenopus* poly A binding protein (PABP). YCL11c also shows similarity to a variety of other RNA binding proteins (data not shown), all containing at least one RNA recognition motif (RRM), a region of around eighty amino acids, which includes two sub-motifs, RNP-2 and RNP-1 (2). YCL11c contains two good fits to the RRM consensus, and a third, more degenerate RRM at the C terminus. The N terminus is arginine rich (26% in the first 52 aa), with four RS dipeptides. Human U2AF⁶⁵, an essential non snRNP protein splicing factor (3), also has three RRM, the third being more degenerate, and an N-terminal domain rich in arginine (34% in the first 63 aa) with nine RS dipeptides. On the basis of overall architecture, we propose that the YCL11c gene product may be a yeast homolog of U2AF⁶⁵, rather than of PABP, which contains four RRM and no arginine-rich region, and for which a *S.cerevisiae* homolog is already known (4). The proteins have 22% identity, and some of the most conserved stretches lie outside the above domains (Figure 1). Using the Protomat suite of programs (5) we extracted consensus motifs for RNP-1 and RNP-2 from a set of 54 putative RRM sequences (2), in order to define the C-terminal RRM. We propose a different RNP-2 position from that previously reported (3) for the third RRM of U2AF⁶⁵, although there is no strong match to the RNP-2 consensus in this RRM for either protein. The YCL11c ORF is intronless, as are most *S.cerevisiae* genes.

U2AF⁶⁵ was originally purified as a protein factor required for the efficient binding of U2 snRNP to the branchpoint region (6). Purified U2AF⁶⁵ binds to the polypyrimidine tract located

between the 3' splice site and the branchpoint. This binding requires one or more RRM, whereas the RS domain appears to be an essential effector domain (3). Most introns in *S.cerevisiae* lack a well defined polypyrimidine tract, but contain a virtually invariant UACUAAC branchpoint sequence a short distance upstream of the 3' splice site (7). Whether the YCL11c gene product is essential, and promotes U2 snRNP binding to the branchpoint in yeast is unknown. Interestingly, human U2 snRNA can functionally replace the longer yeast U2 *in vivo* (8). If YCL11c is indeed a functional homolog of U2AF⁶⁵, the similarities in the splicing machinery of yeast and metazoans would extend to non-snRNP factors, in addition to the known similarities in snRNAs and snRNP polypeptides (7, 9).

ACKNOWLEDGEMENTS

Database searches were performed at the NCBI using the BLAST network service. E.B. is funded by Baring Brothers. This work was funded by NIH grants GM42699 and LM04971.

REFERENCES

1. Oliver, S.G. *et al.* (1992) *Nature* **357**, 38–46.
2. Kenan, D.J., Query, C.C. and Keene, J.D. (1991) *Trends Biochem. Sci.* **16**, 214–220.
3. Zamore, P.D., Patton, J.G. and Green, M.R. (1992) *Nature* **355**, 609–614.
4. Sachs, A.B., Bond, M.W. and Kornberg, R.D. (1986) *Cell* **45**, 827–835.
5. Henikoff, S. and Henikoff, J.G. (1991) *Nucleic Acids Res.* **19**, 6565–6572.
6. Ruskin, B., Zamore, P.D. and Green, M.R. (1988) *Cell* **52**, 207–219.
7. Ruby, S.W. and Abelson, J. (1991) *Trends Genet.* **7**, 79–85.
8. Shuster, E.O. and Guthrie, C. (1990) *Nature* **345**, 270–273.
9. Smith, V. and Barrell, B.G. (1991) *EMBO J.* **10**, 2627–2634.
10. Devereux, J., Haeberli, P. and Smithies, O. (1984) *Nucleic Acids Res.* **12**, 387–395.

Figure 1. Alignment of YCL11c ORF protein (top line) and hU2AF⁶⁵ (bottom line) using the GCG program Bestfit, with a gap weight 3.5 and gap extension 0.1 (10). One additional gap in the arginine-rich domain was introduced manually. Black boxes represent identical matches; stippled boxes represent conservative changes (V=L=I, F=Y=W, R=K, Q=N, D=E, S=T). RRM are shown by horizontal lines; RNP-1 and RNP-2 motifs are boxed. The broken line for the C-terminal RRM indicates a weak match to the consensus, particularly for RNP-2 of hU2AF⁶⁵.

* To whom correspondence should be addressed