Nucleotide sequence analysis of the A protein of the U1 small nuclear ribonucleoprotein particle: the murine protein contains a 5′ amino-terminal tag

Michele M.Bennett, Michelle A.Baron and Joe Craft*
Section of Rheumatology, Yale University School of Medicine, New Haven, CT 06510, USA

Received July 9, 1993; Revised and Accepted July 28, 1993 GenBank accession no. L15447

To analyze the evolutionary stability of the U1 snRNP A protein (U1A) of the U1 small nuclear ribonucleoprotein (U1 snRNP), its cDNA was cloned from a murine liver cDNA library and compared to the nucleotide sequence of human U1A (1). The nucleotide sequences of the murine and human clones were similar, with 87% identity for the putative coding regions and 3′ untranslated regions. A main area of divergence was the 5′ untranslated region which had 40% identity. Another difference was the presence of two potential sites for protein initiation in the mouse clone, with nucleotides ATG at positions 100–102 and 118–120; like the human cDNA, the *Xenopus laevis* U1A cDNA only contained one start site (2; Figure 1A). The 5′ ATG codon in the murine cDNA was preceded by an in-frame stop codon TGA, and like the more 3′ ATG, was enclosed in a sequence consistent with the consensus for translation initiation CCA/GCCATGG (3; Figure 1A). Based upon designation of the upstream codon as the initiation site, the mouse and human proteins were 96% identical at the deduced amino acid level, with a main difference being the addition of six amino acids at the amino-terminal in the former (Figure 1B). Human U1A also had an extra glycine and ten other amino acid differences. These changes gave mouse U1A a calculated molecular weight of 31.83 kDa versus 31.28 kDa for the human protein, although both had 12 charged residues with calculated isoelectric points of 10.62 and 10.64, respectively. Despite the amino acid differences, the two RNP binding domains (4) were conserved.

The assignment of the upstream ATG as the initiation codon in the mouse clone with the resulting difference in calculated molecular weight was confirmed by electrophoretic mobilities of the two proteins in SDS–PAGE. T3 polymerase driven RNA transcripts of the mouse cDNA after translation *in vitro* produced a 32-kDa protein that migrated slightly slower than the human protein produced in a like manner (data not shown). In addition, the endogenous U1A protein immunoprecipitated from murine Ehrlich ascites cells with an anti-U1 snRNP antiserum co-migrated with the murine protein produced *in vitro*, and slower than U1A from human HeLa cells. The A protein from purified murine U1 snRNPs (5) also migrated more slowly in SDS–PAGE than did purified human U1A (not shown). Amino acid sequence analysis revealed that the amino terminal of the purified murine protein was blocked.

Previous comparisons of the human and mouse U snRNP core proteins B and D revealed that they are identical at a deduced amino acid level except for one conservative substitution (an isoleucine for leucine at position 205 in human B; 6, 7). The 70 K and A proteins of the U1 snRNP also are conserved between mice and humans (96% for A as shown here, and 94% for 70 K [8]), but less than the identity, or virtual identity, of the core proteins, suggesting less evolutionary constraints on conservation of these proteins in comparison to the core proteins common to all the U snRNPs involved in pre-mRNA splicing.

ACKNOWLEDGEMENTS

We thank Jack Keene and Dan Kenan, Duke University School of Medicine, for the gift of the cDNA clone of human U1A. This work was supported in part by NIH grants AR40072 and AR07107, and by grants from the Connecticut Chapters of the Arthritis Foundation and the Lupus Foundation.

REFERENCES

1. Sillekens,P.T.G., Habets,W.J., Biejer,R.P. and van Venrooij,W.J. (1987) *EMBO J.* **6**, 3841–3848.
2. Scherly,D., Kambach,C., Boelens,W., van Venrooij,W.J. and Mattaj,I.W. (1991) *J. Mol. Biol.* **219**, 577–584.
3. Kozak,M. (1984) *Nucleic Acids Res.* **12**, 857–872.
4. Scherly,D., Boelens,W., van Venrooij,W.J., Dathan,N.A., Hamm,J. and Mattaj,I.W. (1989) *EMBO J.* **8**, 4163–4170.
5. Krainer,A.R. (1988) *Nucleic Acids Res.* **16**, 9415–9429.
6. Griffith,A,J., Schmauss,C. and Craft,J. (1992) *Gene* **114**, 195–201.
7. Mitsuda,T., Eisenberg,R.A. and Cohen,P.L. (1992) *J. Autoimmun.* **5**, 277–287.
8. Hornig,H., Fischer,U., Costas,M., Rauh,A. and Lührmann,R. (1989) *Eur. J. Biochem.* **182**, 45–50.

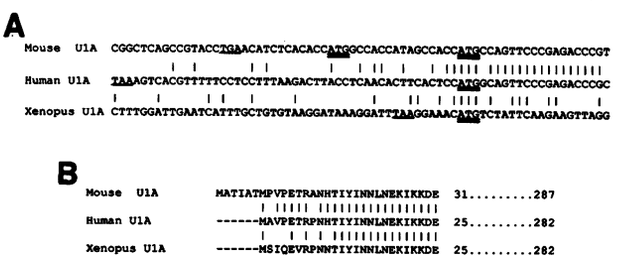

Figure 1. The nucleotide (Panel A) and deduced amino acid sequences (Panel B) encompassing the potential start sites of the human (1), *X.laevis* (2), and murine U1A proteins. Vertical lines represent identical bases, or amino acids, in comparison to the mouse sequence. Potential translation initiation codons (ATG) are shown by a double underline, while upstream in-frame stop codons are denoted by single lines.

* To whom correspondence should be addressed