

Structural and evolutionary analysis of eukaryotic mRNA untranslated regions

Pesole G.1,4, Gissi, C., Grillo G. 4, Licciulli F. 4, Larizza A.3 and Liuni S. 2,4

1 Dipartimento di Fisiologia e Biochimica Generali, Università di Milano, Milano, Italy

2 Centro di Studio sui Mitocondri e Metabolismo Energetico, C.N.R., Bari, Italy

3 Dipartimento di Biochimica e Biologia Molecolare, Università di Bari, Bari, Italy

4 Area di Ricerca del CNR, Bari, Italy

The 5' and 3' untranslated regions of eukaryotic mRNAs may play a crucial role in the regulation of gene expression controlling mRNA localization, stability and translation efficiency. In order to study the general structural and compositional features of these sequences we have previously developed UTRdb, a specialized database of 5' and 3' UTR sequences of eukaryotic mRNAs cleaned from redundancy (Pesole, Liuni et al. 2000).

Utrdb (release 10.0) contains 75,448 entries (26,145,985 nucleotides) which are also annotated for the presence of functional sequence patterns whose biological activity has been experimentally demonstrated. All these patterns have been collected in the UTRsite database where for each functional pattern, corresponding to a specific entry, the consensus structure is reported with a short description of its biological activity and the relevant bibliography. Furthermore, UTRdb entries have been annotated for the presence of repeated elements present in the Repbase database (Jurka 1998). A total of 5,818 functional elements and 54,975 repetitive elements are annotated in UTRdb. All Web resources we implemented for the retrieval and the analysis of UTR sequences are available at the UTR home page (Pesole and Liuni 1999b) we recently implemented. UTRdb entries can be retrieved through the SRS system where crosslinks to UTRsite as well as to the nucleotide or aminoacid primary database are also established. Through the Web facility UTRscan any input sequence can be searched for the presence of a functional pattern annotated in UTRsite and UTRfasta allows to assess sequence similarity between a query sequence and UTRdb entries. The analysis of complete UTR sequences contained in this database showed that 5'-UTR sequences, on the average 187 nucleotides long, were 1,2 to 4,3 times shorter than the corresponding 3'-UTR sequences in the various taxonomic groups considered. As far as the compositional properties were concerned, on average 5'-UTR sequences resulted in all cases GC richer than 3'-UTR sequences and significant correlation was found between the GC content of 5' and 3'-UTR sequences and the GC content of the third silent codon positions of the corresponding protein coding genes (Pesole, Liuni et al. 1997). Some structural features of 5'UTRs were investigated, such as presence of upstream ORFs and context of initiator ATG, which are known to affect the mRNA translation efficiency (Pesole, Bernardi et al. 1999). In order to assess the level of functional constraint of UTR sequences we have studied their evolutionary dynamics also in comparison with the corresponding coding regions. With suitable evolutionary models we have calculated the nucleotide substitution rate of 5'-UTR, 3'-UTR, synonymous and asynonymous positions by comparing complete human, murid (rat and mouse) and artiodactyl mRNAs, for which a suitable number of orthologous sequences was available.

Bibliography

Acknowledgements

This work was partially financed by EC grant ERB-BIO4-CT96-0030.

Jurka, J. (1998). "Repeats in genomic DNA: mining and meaning." *Curr. Op. Struct. Biol.* 8: 333-337.

Pesole, G., G. Bernardi, et al. (1999). "Isochore specificity of AUG initiator context of human genes." *FEBS Lett* 464(1-2): 60-2.

Pesole, G. and S. Liuni (1999b). "Internet resources for the functional analysis of 5' and 3' untranslated regions of eukaryotic mRNAs." *Trends Genet* 15(9): 378.

Pesole, G., S. Liuni, et al. (2000). "UTRdb and UTRsite: specialized databases of sequences and functional elements of 5' and 3' untranslated regions of eukaryotic mRNAs." *Nucleic Acids Res*

28(1): 193-196.

Pesole, G., S. Liuni, et al. (1997). "Structural and compositional features of untranslated regions of eukaryotic mRNAs." *Gene* 205(Dec 31): 95-102.