In silico human genome search and classification of H-ferritin-like genes - (session: Comparative Genomics and Molecular Evolution)

Pasqualina D,Ursi*, Ermanna Rovida*, Paolo Arosio#,Isabella Zanella#

*Istituto di Tecnologie Biomediche, CNR, Via F.lli Cervi 93, 20090 Segrate (Mi), Italy
#Dipartimento Materno Infantile e Tecnologie Biomediche, University of Brescia, Viale Europa 11, 25123 Brescia, Italy

Similarity search analysis of EST database, with H-ferritin cDNA, has recently led to the identification of an intronless gene, encoding a protein similar to H-ferritin (FTH) but with a long N-terminal extension for mitochondrial export. The mature form of this mitochondrial ferritin is about 80% identical to cytosolic H ferritin.(Levi S, Corsi B, Bosisio M, Invernizzi R, Volz A, Sanford D, Arosio P, Drysdale J. A human mitochondrial ferritin encoded by an intronless gene. J Biol Chem 2001 Jul 6;276(27):24437-40).
This finding raises the possibility that other H-ferritin-like DNA sequences might encode functional genes. To this aim we performed systematic in silico studies on the human genome. BLAST analysis using full H-ferritin cDNA sequence identified 29 DNA fragments with a E-values < 0.0001, used a threshold. After sequence alignment, they could be separated into three blocks. The first includes 14 sequences, which overlap more than 70% of the full FTH-cDNA, which belong to the category of processed pseudogenes. They had > 88% identity to the query and show few gaps or substitutions that in most cases disabled the expression of a functional protein. However, two of them have potential ORFs encoding protein sequences of the same size and highly homologous to FTH (8 and 11 substitutions, respectively), and one has a potential ORF encoding a longer sequence with an N-terminal extension. None of these sequences were represented in EST database, and all showed polyA stretches at 3, and/or repeat flanking regions. This indicates that they represent non-functional pseudo genes. The second group of 7 components include sequences, which overlap 20-60% of FTH-cDNA with identity between 80 and 85%. All potential ORF carried disabling mutations and none of them was represented in EST database. We concluded that they represent non-functional pseudogenic fragments. The third group was composed by sequences which overlapped about 50% FTH-cDNA with 70-80% identity. It included the previously described mitochondrial ferritin MtF, on chromosome 5, and five sequences on chromosome X. One of them (FTHL17) was already described and found expressed in spermatogonia and encodes a peptide of 183 amino acids (Wang PJ, McCarrey JR, Yang F, Page DC. An abundance of X-linked genes expressed in spermatogonia. Nat Genet 2001 Apr;27(4):422-6). The other sequences are located in close proximity, one encodes for a peptide of 158 residues and the other ones are characterized by N-terminal extensions of 30-70 amino acids. Interestingly, in none of them the residues of the ferroxidase centre are fully conserved. They have high similarity to sequences present in EST database, lack polyA stretches and repeat flanking regions, and might be functional genes. Some of these DNAs have been cloned in expression vectors, and work is in progress to study structure and expression of the corresponding proteins.