Molecular Biodiversity of the ferritin protein family

Di Micco P[1], Morea V[2]

Motivation

The conservation of Biodiversity is a high priority of the international community. To preserve Biodiversity it is necessary to identify the mechanisms underlying its existence. Biodiversity can be studied at different levels of biological organization: molecules, cells, organisms and ecosystems. Molecular Biodiversity provides essential information to understand complex systems and can be described with the highest detail. We have undertaken a comparative analysis of Molecular Biodiversity of the available structures and sequences of members of an important protein family with well characterized function, namely the prokaryotic members of the ferritin family. The ferritin family includes ferritin (Fn), bacterioferritin (Bfn) and DNA binding proteins from starved cells (Dps). Recognition of these proteins is complicated by the fact that they are not distinguished by the major protein structure and sequence classification databases (e.g., SCOP, Superfamily, Pfam). The aim of the work is to assign Fn, Bfn and Dps domains to all completely sequenced bacterial and archaeal genomes, and to identify the existence of relationships between each occurring combination of Fn, Bfn and Dps genes and bacterial features such as pathogenicity and/or adaptation to specific ecological environments (e.g., acid pH, high salinity, temperature). The acquired knowledge of sequence-structure-function relationships for members of this family can be exploited for the rational design of protein variants endowed with novel properties for biotechnological and biomedical applications.

Methods

Most Molecular Biodiversity studies take into account DNA, RNA or protein sequences solely and, therefore, overlook a number of evolutionary relationships, since many genes have diverged beyond the ability of current sequence comparison methods to recognize their homology. In contrast, the three-dimensional structures of homologous proteins do not diverge as much, due to the constraints imposed by the maintenance of protein function. Therefore, to draw an exhaustive picture of the Molecular Biodiversity of the ferritin protein family, we both used information on protein structures and structure/function relationships reported in the literature and performed ad hoc structural analyses. In particular, we took advantage of information on the 'key-residues' involved in iron binding

[1] University of Rome "Sapienza", Department of Biochemical Sciences "A. Rossi Fanelli" [2] National Research Council, Institute of Molecular Biology and Pathology (CNR-IBPM)

and/or oxidation (a mechanism of cell protection shared by Fn, Bfn and Dps proteins) or in DNA binding and condensation (shared by Dps proteins only).

Results

Based on the available information on structure/function relationships, and in particular on the 'key-residues' involved in iron binding/oxidation, the large majority of the sequences belonging to the ferritin family in the Superfamily database was assigned to the Fn, Bfn or Dps protein domains. The observation that the residues in the vicinity of the iron binding/oxidation site can be classified into two groups is one of the novel outcomes of the analysis. The Superfamily database is a suitable starting point to obtain an exhaustive picture of the Biodiversity of the ferritin family, since it contains gene assignments to protein families for the largest available collection of completely sequenced prokaryote genomes. To establish whether the use of a single, though extensive, database suffices to ensure exhaustiveness, the analysis will be extended to the Pfam database, where protein sequences are classified into families based on different principles.

Contact e-mail

patrizio.dimicco@uniroma1.it