

The Bricolage of Protein Internal Repeats?

G.Lavorgna, L. Patthy, E. Boncinelli

DIBIT, Istituto Scientifico H. S. Raffaele, Via Olgettina 60, 20132 Milano, Italy

Evolution has brought about the formation of organisms of increasing complexity. A major role in this process is thought to have been played by factors, like exon-shuffling and gene duplication, which have increased the inter-molecular duplications of the more sophisticated proteomes. For example, gene duplication has contributed to the origin and evolution of vertebrates, which appear to possess several copies of an ancestral set of genes. A single gene in flies is usually represented by three or four related paralogous genes in mammals and this spare genetic capacity has permitted new possibilities, allowing to acquire new biochemical functions and expression capabilities. More than two decades ago, when only a handful of eukaryotic genes had been cloned, Francois Jacob had already envisioned some of these basic mechanisms. In fact, he argued that evolution could work as a tinkerer, rather than an engineer, implying that evolutionary processes construct things with the materials at hand and the outcome bears the limitations, or constraints, imposed by those materials. Translated in molecular terms, the raw materials are the existing set of genes, which can be, in part or entirely, elaborated again and redeployed to a new function during evolution. According to Jacob's view of 'recyclement' of biological material, we systematically investigated the possibility that, beside the increase of inter-molecular duplications, also an increase of the intra-molecular duplications had accompanied the biological evolution of proteins. We decided to look for repeated protein modules, as opposed to short, low-complexity sequence repeats (i.e., runs of Qs, STSTSTSTS, etc), because, on several instances, modules of proteins have been shown to be the building blocks on which the function of many multidomain proteins is based. As a result, we found, with a few exceptions, that (i) there is a correlation between the complexity of functions controlled by the proteome of a given organism and its degree of internal repetitiveness. (ii) The above correlation is often observed both for inter-domain comparisons (for example, archaeal proteins are, on average, more internally repeated than bacterial ones) and intra-domain comparisons (for example, human proteins are more internally repeated than those belonging to *D. melanogaster*). (iii) In case of reductive evolution, in which the biological complexity of an organism is lower than the one of its ancestor, like in case of endosymbiont organisms, a parallel decrease of the internal repetitiveness of the proteins was also detected.