Matrix Biology Vol. 15/1996, pp. 311–312

László Patthy has written yet another very interesting review on the topic of modular proteins and their evolution. Whereas most reviews (for two other recent starting points see also Doolittle, 1995, and Bork et al., 1996) concentrate on the description of the phenomenon of module shuffling, László Patthy demonstrates the involvement of exon shuffling for a number of extracellular proteins. As these proteins only occur in multicellular organisms, the introduction of exon shuffling shortly before or at the beginning of the metazoan radiation is likely.

At the behest of the editors, I would like to complement the facts on extracellular modules by discussing some aspects of the evolution of cytoplasmic modular proteins based on our own work. A substantial fraction of intracellular regulatory proteins also contain modules (Bork and Koonin, 1996), but their intron positions and phasing is much less conserved, even in those cytoplasmic proteins that have apparently more recently evolved such as the protein tyrosine kinases, which are not present in yeast as judged by screening of its completely sequenced genome. Were the introns of the genes for these proteins eroded so that exon shuffling is not visible anymore? Or are these modules the subject of another shuffling mechanism? The latter is likely for those cytoplasmic modules that are already found in numerous distinct yeast proteins; consider Sh3 for a prominent example and the DEP domain for a recent addition (Ponting and Bork, 1996). These domains often coexist with more "modern" modules (that probably have spread after the metazoan radiation) in animal proteins.

Are there some other genetic mechanisms that compete with exon shuffling? A considerable fraction of large cytoplasmic proteins contains numerous successive repeats, sometimes associated with single modules. There are various types of those successive repeats (TPR, WD40, HEAT, ARM, spectrin, ankyrin, LRR, etc. to give some examples) with different underlying structures, such as stacked helices, beta-propellers, coiled coil elements, etc.; even 30 repeats in a row are not unlikely. These proteins are more likely to have evolved by slippage mechanisms, so that the scenario of the evolution of multidomain proteins becomes more complex.

Has exon shuffling been evolved with the extracellular modules? Where do they come from? A growing fraction of the extracellular modules without cysteine bridges is also found in cytoplasmic proteins, such as fibronectin-type III (Fn3), immunoglobulin (Ig)-related domains, and for a recent addition, the MATH domain (Uren and Vaux, 1996). It remains to be elucidated for each case whether extracellular modules have been added recently to cytoplasmic proteins or whether the cytoplasmic domains represent the ancestors of the more "modern" extracellular modules. For Ig domains the latter is likely, as the Ig fold has been found in numerous proteins and species; the introduction of cysteines not only stabilizes this fold but also allows a faster mutation rate of the surface residues (see Bork et al., 1996 and refs. therein). Of the more than 60 well-characterized extracellular modules (e.g. Bork et al., 1996 and refs. therein), a large fraction contains disulfide bridges that cannot survive in the reducing cytoplasmic environment. They most likely have ancestors without disulfide bridges which might be cellular modules or enzymes, such as an ATPase as in the case of von Willebrand factor type A domain, that have probably evolved without exon shuffling.

Are modules spread by exon shuffling able to continue to spread without introns? Several "modern" Fn3 domains that are undoubtedly the subject of exon shuffling have highly similar homologues in prokaryotes, possibly transmitted via horizontal gene transfer (Little et al., 1994). It has been shown unambiguously that their spreading in different extracellular, carbohydrate-degrading glycohydrolases of soil bacteria did not follow the evolution of the harboring enzymes (Little et al., 1994). Fn3 domains are by far not the only binding domains that coexist with catalytic units of distinct prokaryotic glycohydrolases; many of these "moving" domains have been classified (Gilkes et al., 1991). Recombination via DNA hairpins that encode the proline-rich linkers between such binding domains (Wu et al., 1990) is only one possible mechanism to produce their spreading.

In summary, exon shuffling is certainly a major player in the evolution of matrix proteins as has been nicely summarized in the minireview by László Patthy. As he points out, this type of genetic mechanism is not the only one leading to domain rearrangements. There are multiple indicators pointing to only a minor role of exon shuffling in the evolution of numerous modular cytoplasmic and nuclear modular proteins. Proteins with three or more domains are not rare in prokaryotes (as can be seen in the first completely sequenced bacterial genomes) and include even housekeeping enzymes such as DNA polymerases (Koonin and Bork, 1996). Is exon shuffling a mechanism specific to "modern" extracellular proteins, resulting from an enormous pressure during the expansion of the invertebrates in a relatively short period of time? Probably not, but certainty about it will probably not be reached in the near future as our understanding of genetic mechanisms are just emerging with the comparative analysis of completely sequenced genomes.

Peer Bork
EMBL, Meyerhofstr. 1, 69012 Heidelberg, and
Max-Delbrück-Center for Molecular Medicine,
Berlin-Buch, Germany

References

Bork, P. and Koonin, E. V.: Protein sequence motifs. *Curr. Opin. Struct. Biol. 6:* 366–376, 1996.

Bork, P., Downing, K., Kieffer, B. and Campbell, I. D.: Structure and distribution of extracellular protein modules. *Quart. Rev. Biophys. 29:* 119–167, 1996.

Doolittle, R. F.: The multiplicity of domains in proteins. *Ann. Rev. Biochem. 64:* 287–314, 1995.

Gilkes, N. R., Henrissat, B., Kilburn, D. G., Miller, Jr., R. C. and Warrent, R. A. J.: Domains in microbial 1,4-glycanases: Sequence conservation, function and enzyme families. *Microbiol. Rev. 55:* 303–315, 1991.

Koonin, E. V. and Bork, P.: Ancient duplication of DNA polymerase inferred from analysis of complete bacterial genomes. *Trends Biochem. 21:* 128–129, 1996.

Little, E., Bork, P. and Doolittle, R. F.: Tracing the spread of fibronectin type III domains in bacterial glycohydrolases. *J. Mol. Evol. 39:* 631–643, 1994.

Ponting, C. P. and Bork, P.: Pleckstrin's repeat performance: A novel domain in G-proteins signaling. *Trends Biochem. Sci. 21:* 245–246, 1996.

Uren, A. G. and Vaux, D. L.: TRAF proteins and mephrins share a conserved domain. *Trends Biochem. Sci. 21:* 128–129, 1996.

Wu, L. F., Tomich, J. M., Saier, Jr., M. H.: Structure and evolution of a multidomain multiphosphoryl transfer proteins. Nucleotide sequence of the fruB (HI) gene in *Rhodobacter capsulatus. J. Mol. Biol. 213:* 687–703, 1990.