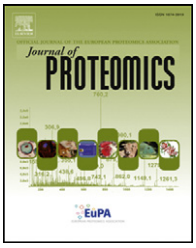
JOURNAL OF PROTEOMICS 81 (2013) 112–125

Available online at www.sciencedirect.com

SciVerse ScienceDirect

www.elsevier.com/locate/jprot



Transcriptional divergence plays a role in the rewiring of protein interaction networks after gene duplication☆

*Isabelle Gagnon-Arsenault, François-Christophe Marois Blanchet, Samuel Rochette, Guillaume Diss, Alexandre K. Dubé, Christian R. Landry**

Département de Biologie, Institut de Biologie Intégrative et des Systèmes, PROTEO, 1030 Avenue de la Médecine, Université Laval, Québec, Québec, Canada G1V 0A6

ARTICLE INFO

Available online 12 October 2012

Keywords:
Protein interaction network
Evolution
Gene duplication
Transcriptional regulation
Network rewiring

ABSTRACT

Gene duplication plays a key role in the evolution of protein–protein interaction (PPI) networks. After a gene duplication event, paralogous proteins may diverge through the gain and loss of PPIs. This divergence can be explained by two non-exclusive mechanisms. First, mutations may accumulate in the coding sequences of the paralogs and affect their protein sequences, which can modify, for instance, their binding interfaces and thus their interaction specificity. Second, mutations may accumulate in the non-coding region of the genes and affect their regulatory sequences. The resulting changes in expression profiles can lead to paralogous proteins being differentially expressed and occurring in the cell with different sets of potential interaction partners. These changes could also alter splicing regulation and lead to the inclusion or exclusion of alternative exons. The evolutionary role of these regulatory mechanisms remains largely unexplored. We use bioinformatics analyses of existing PPI data and proteome-wide PPI screening to show that the divergence of transcriptional regulation between paralogs plays a significant role in determining their PPI specificity. Because many gene duplication events are followed by rapid changes in transcriptional regulation, our results suggest that PPI networks may be rewired by gene duplication, without the need for protein to diverge in their binding specificities.

This article is part of a Special Issue entitled: From protein structures to clinical applications.

1. Introduction

All cellular processes depend on the architecture of protein–protein interaction networks (PINs) [1,2]. The evolutionary expansion of protein repertoires encoded in eukaryotic genomes has contributed to the complexity of these networks. These expansions have taken place largely through the duplication of existing genes followed by their divergence. Typical eukaryotic genomes contain from 30 to 65% of duplicated genes [3], which makes gene duplication a major evolutionary determinant of the architecture of eukaryotic genomes, proteomes and cellular networks. The evolution of PINs by gene duplication is thought to have contributed to many of their structural [4] and functional features, such as modularity and redundancy [5–7]. Understanding what molecular changes contribute to the divergence of protein–protein interactions (PPIs) between paralogous proteins after gene duplication is therefore a major question in evolutionary systems biology.

☆ This article is part of a Special Issue entitled: From protein structures to clinical applications.

* *Corresponding author*. Tel.: +1 418 656 3954; fax: +1 418 656 7176.
E-mail address: clandry@post.harvard.edu (C.R. Landry).

1874-3919/$ – see front matter
http://dx.doi.org/10.1016/j.jprot.2012.09.038

After gene duplication, the two paralogous proteins are identical and thus are expected to have the same interaction partners. If the two gene copies are preserved, they will eventually diverge by accumulating mutations, which will result in the rewiring of the architecture of ancestral networks by the addition of new interactions or by the elimination of pre-existing ones (Fig. 1) [8–10]. Investigating the molecular mechanisms by which PPIs change over time between paralogs is of prime importance because it will illuminate our understanding of how the intricate networks of interactions found in complex eukaryotes such as humans have evolved from simpler ones. Furthermore, a better description of the molecular underpinning of the divergence of PPIs among paralogous proteins would allow to establish relationships between mutations and changes in PPIs, which is a key for a better understanding of genotype–phenotype maps and the role of PPIs in human diseases [1].

The molecular mechanisms that contribute to the divergence of PPIs after gene duplication can be grouped in two broad categories, those that involve a divergence of the coding region of the gene (C-evolution) and those that involve the non-coding regions and thus transcriptional regulation (R-evolution) (*e.g.* [11–14]). In the first case, the coding sequence may accumulate mutations that modify the sequence of the protein and its localization, binding interfaces, structure or posttranslational modifications, all of which may influence its ability to interact with other proteins. To a large extent, the evolution of the coding sequence can also affect protein abundance by modifying, for instance, the mRNA structure, translation efficiency and protein degradation. These sequence properties have recently been shown to have an impact similar to that of mRNA abundance on the abundance of proteins [15]. The divergence in PPIs is, in this case, an intrinsic property of the coding sequence of the gene. Studies have shown that few amino acid changes may be sufficient to alter the interaction specificity of a protein [16–19] such that C-evolution is likely a common mechanism by which PPIs diverge among paralogs. Accordingly, studies examining the rate and molecular mechanisms by which PINs evolve have mostly concentrated on these types of changes, particularly on changes in protein sequences (*e.g.* [10,20]). In the R-evolution scenario, differences in the genomic environment (non-coding regions) of the gene may lead to differences in transcriptional regulation. Assuming that they were duplicated as well, the *cis* regulatory region of a gene (upstream or downstream of the CDS and introns) may accumulate mutations and diverge, resulting in modifications of its transcriptional regulation [21,22] (Fig. 1). In addition to the flanking *cis* regulatory sequences, other elements of the genomic environment like chromatin structure and the position of the gene in the nuclear space could also diverge and affect transcriptional regulation [23,24]. As a consequence, increases in mRNA and protein abundance, for instance, can result in new bindings with interaction partners that have lower affinity. Similarly, changes in timing of expression could simply bring together proteins that were otherwise not expressed in the same cell type or stage of the cell cycle. The ability to interact may exist before the changes in regulation occur but these changes are needed to bring the interacting proteins together in the same cell and at the same time. Expression timing during cell cycle, for example, has been shown to play a role in the interaction specificity of the cyclins, the regulatory subunits of the kinases regulating cell cycle progression [11–14]. Changes in the alternative splicing of genes could also alter the inclusion or exclusion of protein interaction domains and modify the interaction partners of a protein [25].

In many of the cases mentioned above, changes in transcriptional regulation could occur quickly between paralogs because partial gene duplications are frequent and may only involve the

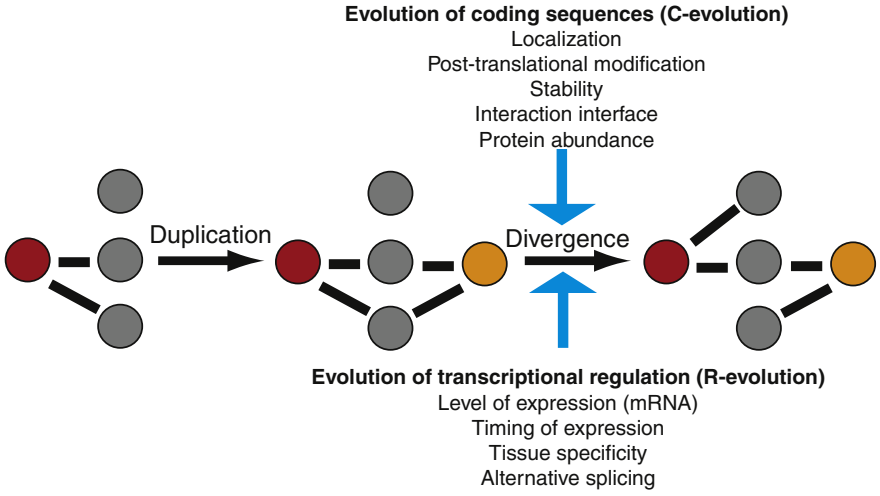


Fig. 1 – Evolution of protein–protein interactions after gene duplication. Shortly after gene duplication, two paralogous proteins are identical and thus have identical patterns of PPIs. If the new gene copy is not lost, mutations will accumulate and contribute to the divergence of PPIs between the two paralogs, which will gain and lose PPI partners. These mutations will have different effects whether they are in the coding or non-coding regions of the genes. While it has been clearly demonstrated that changes in the amino acid sequences of proteins can affect their ability to interact with other proteins, it is less clear whether the evolution of the non-coding sequences of a gene can play a similar role.