The Nature of Protein Domain Evolution: Shaping the Interaction Network

Christoph P. Bagowski*[,1], Wouter Bruins[2] and Aartjan J.W. te Velthuis*[,3,4]

[1]*German University Cairo, Faculty of Pharmacy and Biotechnology, New Cairo City, Egypt;* [2]*Institute of Biology, Leiden University, 2333 AL Leiden, The Netherlands;* [3]*Department of Medical Microbiology, Molecular Virology Laboratory, Leiden University Medical Center, Albinusdreef 2, 2333 ZA Leiden, The Netherlands;* [4]*Department of Bionanoscience, Delft University of Technology, Lorentzweg 1, 2628 CJ, Delft, The Netherlands*

Abstract: The proteomes that make up the collection of proteins in contemporary organisms evolved through recombination and duplication of a limited set of domains. These protein domains are essentially the main components of globular proteins and are the most principal level at which protein function and protein interactions can be understood. An important aspect of domain evolution is their atomic structure and biochemical function, which are both specified by the information in the amino acid sequence. Changes in this information may bring about new folds, functions and protein architectures. With the present and still increasing wealth of sequences and annotation data brought about by genomics, new evolutionary relationships are constantly being revealed, unknown structures modeled and phylogenies inferred. Such investigations not only help predict the function of newly discovered proteins, but also assist in mapping unforeseen pathways of evolution and reveal crucial, co-evolving inter- and intra-molecular interactions. In turn this will help us describe how protein domains shaped cellular interaction networks and the dynamics with which they are regulated in the cell. Additionally, these studies can be used for the design of new and optimized protein domains for therapy. In this review, we aim to describe the basic concepts of protein domain evolution and illustrate recent developments in molecular evolution that have provided valuable new insights in the field of comparative genomics and protein interaction networks.

Received on: May 08, 2010 - Revised on: June 04, 2010 - Accepted on: June 13, 2010

Keywords: Protein domain, PDZ domain, systems biology, superfamily, molecular evolution, interactome.

INTRODUCTION

The protein universe is the collection of proteins of all biological species that exist or have once existed on Earth [1]. Our sampling and understanding of it began over half a century ago, when the first peptide and protein sequences were determined by Sanger [2, 3] and, subsequently, the sequencing of RNA and DNA [4-6]. In the meantime, the genome projects of the last decade have uncovered an overwhelming amount of sequence data and researchers are now starting to address a series of fundamental questions that should shed light onto protein evolution processes [7-10]. For instance, how many gene encoding sequences are present in one genome? How many sequences are repetitive and are these sequences similar in the various organisms on Earth? Which genes were involved in the large scale genome duplications that we see in animals?

A comparison of sequences for evolutionary insight is best achieved by looking at the structural and functional (sub)units of proteins, the protein domains. By convention, domains are defined as conserved, functionally independent protein sequences, which bind or process ligands using a core structural motif [11-13]. Examples of domain modes of actions in signaling cascades for instance, are to connect different components into a larger complex or to bind signaling-molecules [14, 15]. Protein domains can usually fold independently, likely due to their relatively limited size, and are well known to behave as independent genetic elements within genomes [16, 17]. The sum of these features makes protein domains readily identifiable from raw nucleotide and amino acid sequences and many protein family resources (e.g., Superfamily and SMART [see Table **1**]) indeed fully rely on such sequence similarity and motif identifications [18, 19].

DOMAIN IDENTIFICATION, SEQUENCE ALIGNMENT AND PHYLOGENY

The algorithms that are used for domain identification are built around a set of simple assumptions that describe the process of evolution. In general, evolution is believed to form and mold genomes largely *via* three mechanisms, namely i) chemical changes through the incorporation of base analogs, the effects of radiation or random enzymatic errors by polymerases, ii) cellular repair processes that counter mutations, and iii) selection pressures that manifest themselves as the positive or negative influence that determines whether the mutation will be present in subsequent generations [20, 21]. By definition, each of these phenomena

*Address correspondence to these authors at the German University Cairo, Faculty of Pharmacy and Biotechnology, New Cairo City, Egypt; Tel: ++20-2-27590690; Fax: ++20-2-27590772;
E-mail: christoph.bagowski@guc.edu.eg

Department of Medical Microbiology, Molecular Virology Laboratory, Leiden University Medical Center, Albinusdreef 2, 2333 ZA Leiden, The Netherlands; Tel: ++31 71 526 1652; Fax: ++31 71 526 6761;
E-mail: a.j.w.te_velthuis@lumc.nl

1389-2029/10 $55.00+.00

Table 1. List of Public Resources and Databases Relevant to Domain Analysis

Resource	URL
Protein domain databases	
Pfam	http://www.sanger.co.uk/Pfam/
Prosite	http://www.expasy.org/prosite/
SMART	http://smart.embl-heidelberg.de
Superfamily	http://supfam.mrc-lmb.cam.ac.uk/SUPERFAMILY/hmm.html
Structural analysis	
CATH	http://www.cathdb.info/latest/index.html
SCOP	http://scop.mrc-lmb.cam.ac.uk/scop/
SSM	http://www.ebi.ac.uk/msd-srv/ssm/
Swiss-MODEL	http://swissmodel.expasy.org/
Alignment software	
BLAST	http://www.ncbi.nlm.nih.gov/blast/Blast.cgi
ClustalW	http://www.ebi.ac.uk/Tools/clustalw2/
Muscle	http://www.ebi.ac.uk/muscle/
Protein interaction	
HPRD	http://www.hprd.org
MINT	http://mint.bio.uniroma2.it/mint/Welcome.do
STRING	http://string.embl.de/
Phylogenetic analysis	
MrBayes (*Bayesian*)	http://mrbayes.csit.fsu.edu/
PhyML (*Max. Likelihood*)	http://atgc.lirmm.fr/phyml/
PHYLIP (*various*)	http://evolution.genetics.washington.edu/phylip.html
CAPS (*residue coevolution*)	http://bioinf.gen.tcd.ie/~faresm/page11/page11.html
Visualization	
Pymol (*structural*)	http://pymol.sourceforge.net/
NJplot (*phylogeny*)	http://pbil.univ-lyon1.fr/software/njplot.html
DiepView (*structural*)	http://spdbv.vital-it.ch/
TreeView (*phylogeny*)	http://taxonomy.zoology.gla.ac.uk/rod/treeview.html
Visant (*protein interaction*)	http://visant.bu.edu/
Sequence depositories	
Ensembl (*genome projects*)	http://www.ensembl.org
PDB (*structures*)	http://www.rcsb.org/pdb/home/home.do
NCBI	http://www.ncbi.nlm.nih.gov/sites/gquery?itool=toolbar
UniProt	http://www.expasy.uniprot.org/

has its own rate, while their combined effect gives a certain probability for the change of one defined amino acid (or nucleotide) to another within a specific time interval.

Although already informative in its own right, mutation data can be significantly different among species due to dissimilar metabolisms, generation times, population sizes, life-

styles, reproductive strategies, or the lack of apparent polymerase-dependent proofreading such as in positive-stranded RNA viruses [22-25]. Consequently, substitution rates need therefore be calculated to correctly compare two or more sequences and hunt uncharted genomes for comparable domains. Particularly this last strategy, using general rate matrices like BLOSOM and PAM, is an elegant example of how new protein functions can be discovered [26-30]. Fast algorithms for pair-wise alignments can be found in the Basic Local Alignment Search Tool (BLAST), whereas multiple sequence alignments (MSAs, Fig. **1A**) in which multiple sequences are compared simultaneously are commonly created with for example ClustalX and MUSCLE (see Table **1**) [31-34].

Close relatives, sharing an overall sequence identity above for example 50% and a set of functional properties, can also be grouped into families and subfamilies. In turn, these families share also evolutionary relationships with other domains and form together so-called domain superfamilies [18, 35]. Evolutionary distances between related domain sequences can easily be estimated from sequence alignments, provided that the correct rate assumptions are made. Subsequently, these can be used to compute the phylogenies of the domain that share an evolutionary history. These, often tree-like graphs (Fig. **1B**), depend heavily on rate variation models, such as molecular clocks or relaxed molecular clocks (e.g., Maximum Likelyhood and Bayesian estimation), which are calibrated with additional evidence

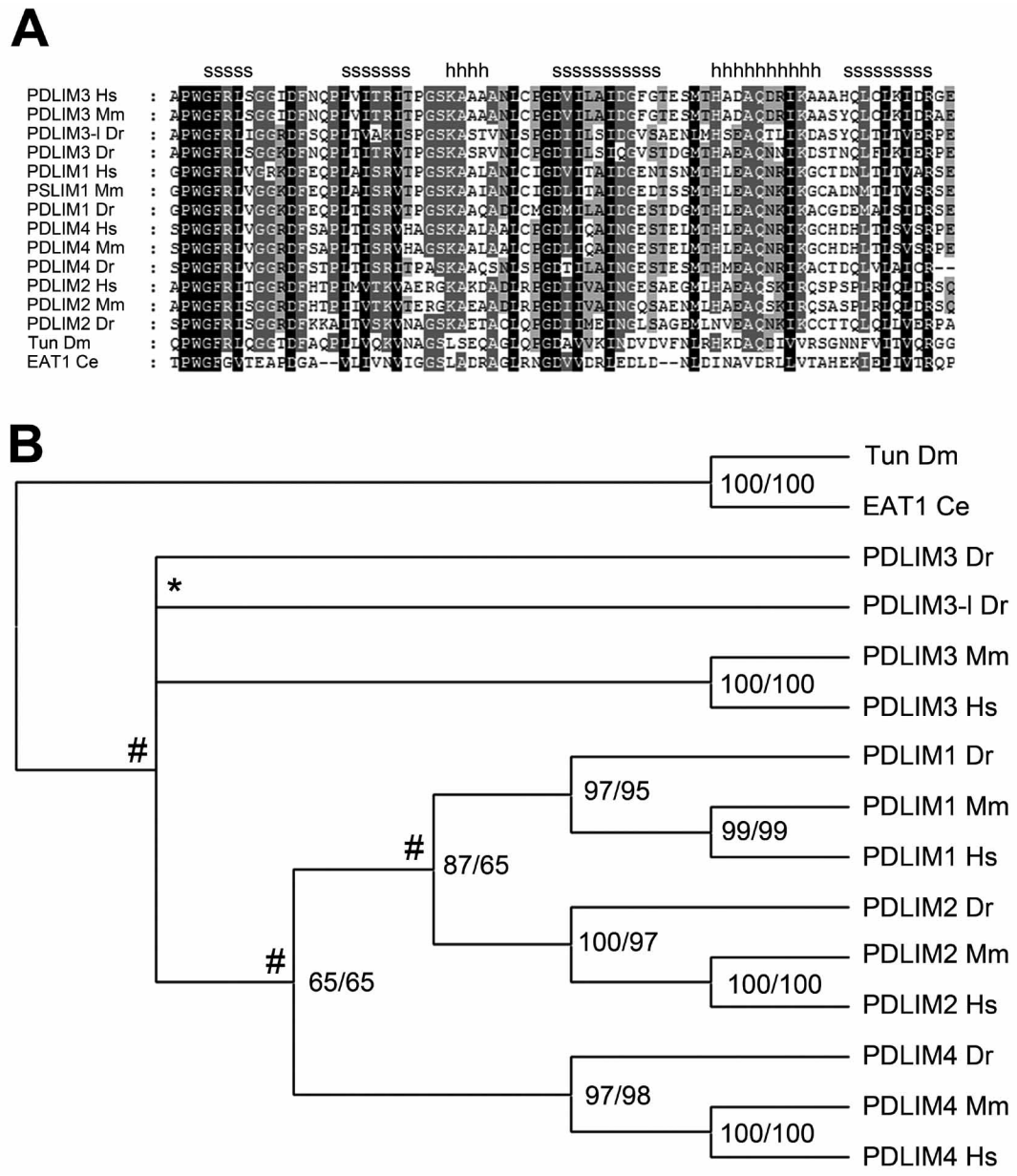

Fig. (1). Example of sequence alignment and phylogeny. (**A**) This figure shows an example alignment of the PDZ domain with different shadings representing the amount of conservation (100, 75 or 50%) at a particular position in the sequence. (**B**) This tree is the phylogenetic presentation of the alignment in Fig. (**1A**). It was computed using Bayesian estimation and presents the best-supported topology for the alignment. Numbers indicate % support by the two methods used, while # indicates gene duplication events in the common ancestor and * marks a species-specific duplication event. For computational details, please see [42].

such as fossils and may therefore also provide valuable information on aspects like divergence times and ancestral sequences [36-38]. Commonly used phylogenetic analysis strategies are listed in Table **1**.

A limitation of all inferred phylogenetic data is that it is directly dependent on the alignment and less so on the programs used to build the phylogenetic tree [39]. One of the shortcomings of automated alignments may thus derive from the fact that they commonly employ a scoring and penalty procedure to find the best possible alignment, since these parameters vary from species to species [22, 23], as mentioned above. Careful inspection of alignments is therefore advisable, even though software has been developed that combines the alignment procedure and phylogenetic analysis iteratively in one single program [40].

DOMAIN DIVERSIFICATION

Although sequence and phylogenetic analysis provide a relatively straightforward way for looking at domain divergence, comparison of solved protein structures has shown that protein tertiary organizations are much more conserved (>50%) than their primary sequence (>5%) [41]. For this reason, protein structures and their models provide significantly more insight into the relations of protein domains and how domain families diverged [16]. For example, the inactive guanylate kinase (GK) domain present in the MAGUK family was shown to originate from an active form of the GK domain residing in Ca2+ channel beta-subunits (CACNBs) through both sequence and structural comparison [42]. Furthermore, identification of functionally or structurally related amino acid sites in a fold sheds light on the complex, co-evolutionary dynamics that took place during selection [43].

As described above, the evolution of a protein domain is generally the result of a combination of a series of random mutations and a selection constraint imposed on function, i.e., the interaction with a ligand. The interaction between protein and ligand can be imagined as disturbances of the protein's energy landscape, which in turn bring about specific, three-dimensional changes in the protein structure [44, 45]. Binding energies however, need not be smoothly distributed over the protein's binding pocket as a limited number of amino acids may account for most of the free-energy change that occurs upon binding [45-47]. In these cases, new binding specificities (including loss of binding) may therefore arise through mutations at these hot spots. An example is a recent study of the PDZ domain in which it was shown that only a selected set of residues, and in particular the first residue of α-helix 2 (αB1), directly confers binding to a set of C-terminal peptides [48].

The folding of a domain is essentially based on a complex network of sequential inter-molecular interactions in time [49]. This has of course significant implications for domain integrity, particularly if one assumes that the core of a protein domain is and has to be largely structurally conserved. Indeed, even single mutations that arise in this area may easily derail the folding process, either because their free energy contribution influences residues in the direct vicinity or disturbs connections higher up in the intermolecular network [49]. It is therefore hypothesized that protein evolution took place at the periphery of the protein domain core, and that gradual changes *via* point mutations, insertions and deletions in surface loops brought about the evolutionary distance we see among proteins to date [21, 50-52].

However, distant sites also contribute to the thermodynamics of catalytic residues. This is achieved through a mechanism called energetic coupling, which is shaped by a continuous pathway of van der Waals interactions that ultimately influences residues at the binding site with similar efficiency as the thermodynamic hotspots [53, 54]. Indeed in such cases, evolutionary constraints are not placed on merely one amino acid in the binding pocket, but on two or more residues that can be shown to be statistically coupled in MSAs [54, 55]. In addition to contributions to binding, these principles also explain why the core of a domain structure will remain largely conserved, while at functionally related places residues can (rapidly) co-evolve with an overall neutral effect [56]. Of course, these aspects of co-evolution are also of practical consequence for structure prediction and rational drug design [43].

DOMAIN DUPLICATION

Through selective mutation, protein domains have been the tools of evolution to create an enormous and diverse assembly of proteins from likely an initially relatively limited set of domains. The combined data in GenBank and other databases now covers over 200.000 species with at least 50 complete genomes and this greatly facilitates genome comparisons [57-59]. Following such extensive comparisons, currently > 1700 domain superfamilies are recognized in the recent release of the Structural Classification of Proteins (SCOP) [60] and it has become clear that many proteins consist of more than one domain [17, 61, 62]. Indeed, it has been estimated that at least 70% of the domains is duplicated in prokaryotes, whereas this number may even be higher in eukaryotes, likely reaching up to 90% [35].

There are various mechanisms through which protein domain or whole proteins may have been duplicated. On the largest scale, whole genome duplication such as those seen in the vertebrate genomes duplicated whole gene families, including postsynaptic proteins, hormone receptors and muscle proteins, and thereby dramatically increased the domain content and expanded networks [42, 63, 64]. On the other end of the scale, domains and proteins have been duplicated through genetic mechanisms like exon-shuffling, retrotranspositions, recombination and horizontal gene transfer [65-67]. Since the genetic forces, like exon-shuffling and genome duplication vary among species, the total number of domains and the types of domains present fluctuate per genome. Interestingly, comparative analyses of genomes have shown that the number of unique domains encoded in organisms is generally proportional to its genome size [60, 68]. Within genomes, the number of domains per gene, the so-called modularity, is related to genome size *via* a power-law, which is essentially the relation between the frequency f and an occurrence x raised by a scaling constant k (i.e., $f(x) \sim x^k$) [69, 70]. A similar correlation is found when the multi-domain architecture is compared to the number of cell types that is present in an organism, i.e., the organism complexity or when the number of domains in a abundant superfamily is plotted against genome size (Fig. **2**) [71, 72].

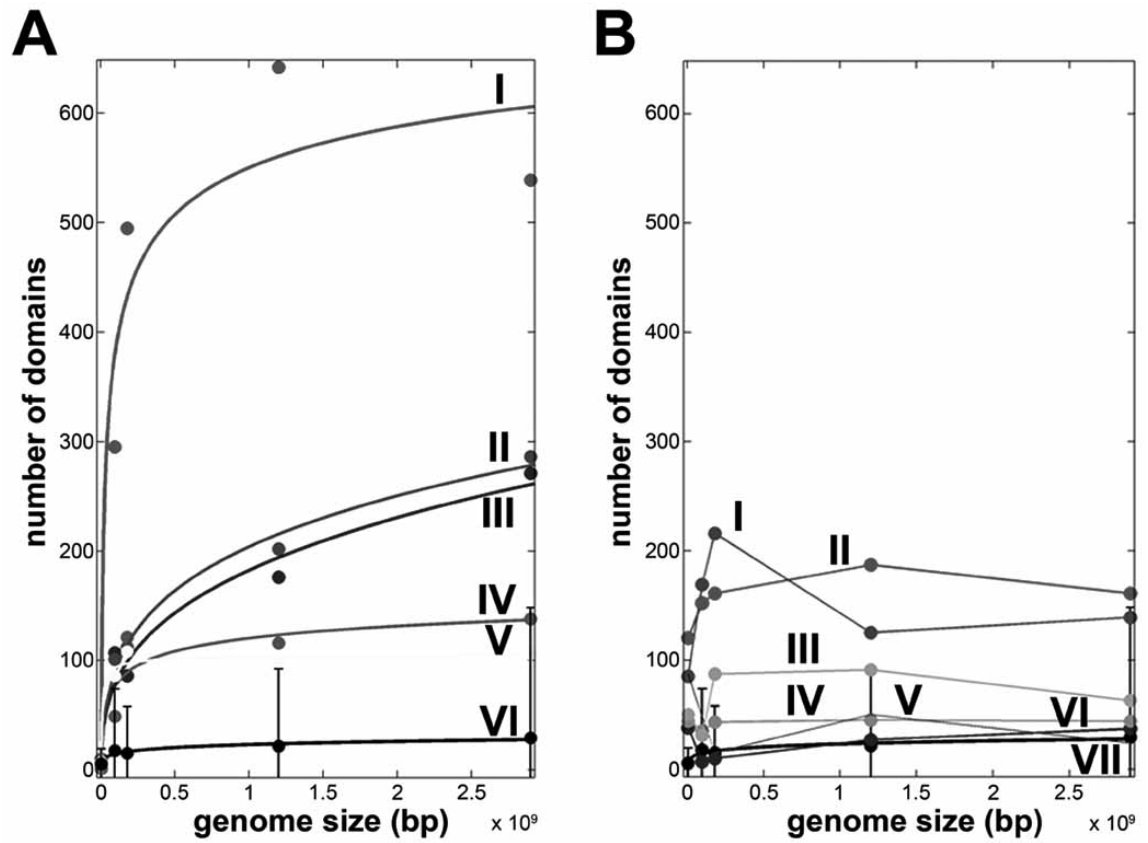

Fig. (2). Selection on superfamily domain size. (**A**) Increase in superfamily domain size fitted to a power-law for kinase-like domains (I), Ankyrin-repeats (II), PDZ-like (III), voltage-gated potassium channels (IV), the catalytic domain of metalloproteases (V) and the average increase in superfamily size (VI). R^2 value for each fit was at least 0.9. (**B**) Neutral or decreasing family sizes can be found for the MFS general substrate transporters (I), NAD(P)-binding Rossmann folds (II), Ribonucleases H (III), PLP-dependent transferases (IV), periplasmic binding proteins type II (V), ATPase domains of HSP90/topoisomerase II/histidine kinase-like folds (VI) and the average increase in superfamily size (VII) as in 2A.

DOMAIN SELECTION

Given the amount of domain duplication and apparent selection for specific multi-domain encoding genes in, for example, vertebrates, it may come as little surprise that not all domains have had the same tendency to recombine and distribute themselves over the genomes [68, 73]. In fact, some are highly abundant and can be found in many different multi-domain architectures, whereas others are abundant yet confined to a small sample of architectures or not abundant at all [68, 70]. Is there any significant correlation between the propensity to distribute and the functional roles domains have in cellular pathways?

Some of the most abundant domains can be found in association with cellular signaling cascades and have been shown to accumulate non-linearly in relation to the overall number of domains encoded or the genome size [70]. Additionally, the on-set of the exponential expansion of the number of abundant and highly recombining domains has been linked to the appearance of multicellularity [70]. A reoccurring theme among these abundant domains is the function of protein-protein interaction and it appears that particularly these, usually globular domains, have been particularly selected for in more complex organisms [70]. This positive relation is underlined by the association of these abundant domains with disease such as cancer and gene essentiality as the highly interacting proteins that they are part of have central places in cascades and need to orchestrate a high number of molecular connections [74, 75]. Their shape and coding regions, which usually lie within the boundaries of one or two exons, make them ideally suited for such a selection, since domains are most frequently gained through insertions at the N- or C-terminus and through exon shuffling [76-78].

From a mutational point of view, protein-protein interaction domains are different from other domains as well and this appears to be particularly true for the group of small, relatively promiscuous domains like SH3 and PDZ. These domains are promiscuous in the sense that they both tend to physically interact with a large number of ligands [79, 80] and are prone to move through the genome to recombine with many other domains. It has been found that particularly these domains evolve more slowly than non-promiscuous domains [70]. This likely stems from the fact that they are required to participate in many different interactions, which makes selection pressures more stringent and the appearance of the branches on phylogenetic trees relatively short and more difficult to assess when co-evolutionary data in terms of other domains in the same gene family or expression patterns is limited [42, 63]. Non-promiscuous domains on the other hand can quite easily evade the selection pressure by

obtaining compensatory mutations either within themselves or their specific binding partner [70].

The overall phenomenon that the number of protein domains and their modularity increases as the genome expands has not been linked to a conclusive biological explanation yet. A rationale for the increase in interactions and functional subunits, however, may derive from the paradoxical absence of correlation between the number of genes encoded and organism complexity, the so-called G-value paradox [81]. There is indeed evidence that domains involved in the same functional pathway tend to converge in a single protein sequence, which would make pathways more controllable and reliable without the need for supplementary genes [73]. Additionally, the number of different arrangements found in higher eukaryotes is, given the vast scale of unique domains present, relatively limited. This in turn implies that evolutionary constraints have played an important role in selecting the right domain combinations and the right order from N- to C-terminus in multi-domain proteins [13, 82]. In fact, the ordering and co-occurrence of domains was demonstrated to hold enough evolutionary information to construct a tree of life similar to those based on canonical sequence data [70]. Furthermore, the increased use of alternative splicing and exon skipping in higher eukaryotes likely supplied a novel way of proteome diversification by restricting gene duplication and stimulating the formation of multi-domain proteins [83, 84]. In plants, however, the latter notion is not supported since both mono- and dicots show limited alternative splicing and a more extensive polyploidy [85-87].

THE EVOLUTION OF DOMAIN INTERACTION NETWORKS

It is clear that some of the above characteristics are underappreciated in the phylogenetic analysis of linear amino acid sequences. Moreover, the effects of evolution extend even further than these aspects and entail transcriptional and translational regulation, intramolecular domain-domain interactions, gene modifications and post-translational protein modifications [88-96]. New methods are thus being developed to take into account that when sequences evolve, their close and distant functional relationships evolve in parallel. Correlations of mutations have already been found between residues of different proteins [97, 98] and compensating mutational changes at an interaction interface were shown to recover the instability of a complex [99]. These observations are evidence for the current evolutionary models for the protein-protein interaction (PPIs) networks that are being constructed through large-scale screens [100-102]. In these, a gene duplication or domain duplication (depending on the resolution of the network) implies the addition of a node, while the deletion of a gene or domain reduces the amount of links in the network (Fig. **3**). In the next step, extensive network rewiring may take place, driven by the effect of node addition or node loss in the network (i.e., the duplicability or essentiality of a domain/protein) and mutations in the domain-interaction interface [67, 74, 103-105].

Beyond mutations at the domain and protein level, regulation of protein expression provides another vital mechanism through which protein networks can evolve. Microarray studies are now well under way to map genome-wide expression levels of related and non-related genes under a variety of conditions [91, 94-96]. For example, transcriptional comparisons have investigated aging [106] and pathogenicity [107]. Unfortunately, given the highly variable nature of gene expression and the fact that different species may respond different to external stimuli, such comparisons can only be performed under strictly controlled research conditions. To date most studies have therefore focused on the embryogenesis, metamorphosis, sex-dependency and mutation rates of subspecies [94, 108-111]. Other studies have revealed valuable information on promoter types and duplication events [91-94].

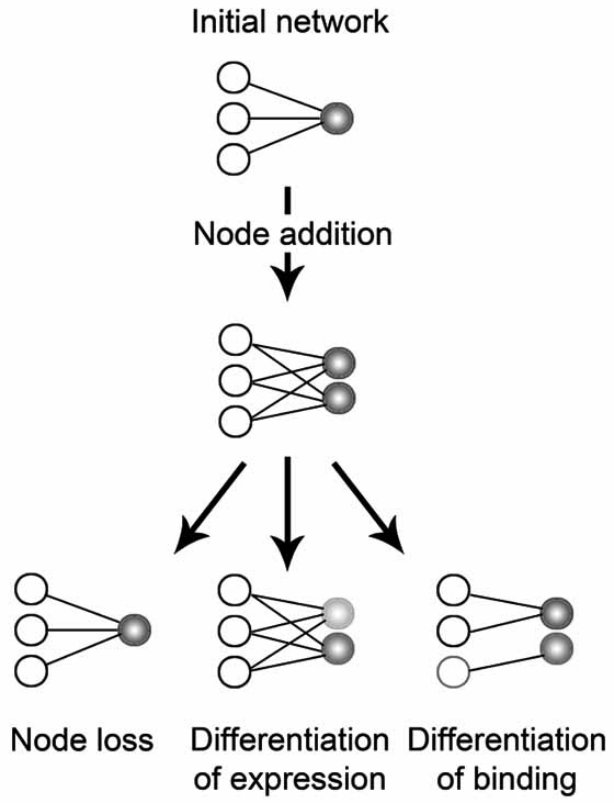

Fig. (3). Evolutionary models for protein-protein interactions. The evolution of protein networks is tightly coupled to the addition or deletion of nodes. Additionally, events that introduce mutations in binding interfaces of proteins may result in the addition or loss of links in the network. Node addition may take place through *e.g.*, domain duplication or horizontal gene transfer, while rewiring of the network is mediated by point mutations, alternative splice variants and changes in gene expression patterns.

To overcome the limitations mentioned in the previous paragraph, the analysis of co-expression data has been developed to supplement the direct comparison of individual gene expression changes [95]. In this procedure, a co-expression analysis of gene pairs within each species precedes the cross-comparison of the different organisms in the study. This approach thus primarily focuses on the similarity and differences of the orthologous genes within network, and is therefore ideally suited for the study of protein domain evolution and has already revealed that species-specific parts

of an expression network resulted *via* a merge of conserved and newly evolved modules [95, 112, 113].

CONCLUDING REMARKS

Finding evolutionary relationships protein domains is mostly based on orthology and thus commonly performed on best sequence matches. Identifying these and categorizing them depends largely on multiple sequence alignments and this will in most cases give good indications for function, fold and ultimately evolution. However, this approach usually discards apparent ambiguities that arise from species-specific variations (e.g., due to population size, metabolism or species-specific domain duplications or losses) and may therefore introduce significant biases [114]. Biases may also derive from the method of alignment, the rate variation model used to infer the phylogeny, and the sample size used to build the alignment [39, 40, 115]. Care should therefore be taken to not regard orthology as a one-to-one relationship, but as a family of homologous relations [91], to select for appropriate analysis methods [39, 115] and extend comparative data to protein interactions and expression profiles [91]. Indeed, as our wealth of biological information expands, our systems perspective will improve and provide us with an opportunity to reveal protein domain evolution at the level network organization and dynamics. Large-scale expression studies are beginning to show us evolutionary correlations between gene expression levels and timings [94, 106, 107, 112, 116], while others demonstrate spatial differences between paralogs or (partial) overlap between interaction partners [117-120]. Indeed, when we are able to map the spatio-temporal aspects of inter- and intra-molecular interactions we will begin to fully understand the versatile power of evolution that shaped the protein universe and life on Earth [118].

FUNDING

AV is supported by The Netherlands Organization for Scientific Research (NWO) through Toptalent grant 021.001.037.

REFERENCES

[1] Ladunga, I. Phylogenetic continuum indicates "galaxies" in the protein universe: perliminary results on the natural group structures of proteins. *J. Mol. Evol.*, **1992**, *34*, 358-375.

[2] Bailey, K.; Sanger, F. The chemistry of amino acids and proteins. *Annu. Rev. Biochem.*, **1951**, *20*, 103-130.

[3] Sanger, F. Some peptides from insulin. *Nature*, **1948**, *162*, 491.

[4] Adams, J.M.; Jeppesen, P.G.; Sanger, F.; Barrell, B.G. Nucleotide sequence from the coat protein cistron of R17 bacteriophage RNA. *Nature*, **1969**, *223*, 1009-1014.

[5] Sanger, F.; Donelson, J.E.; Coulson, A.R.; Kössel, H.; Fisher, D. Use of DNA polymerase I primed by a synthetic oligonucleotide to determine a nucleotide sequenc of phage fl DNA. *Proc. Natl. Acad. Sci. USA*, **1973**, *70*, 1209-1213.

[6] Sanger, F.; Nicklen, S.; Coulson, A.R. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA*, **1977**, *74*, 5463-5467.

[7] Adams, M.D.; Celniker, S.E.; Holt, R.A.; Evans, C.A.; Gocayne, J.D.; Amanatides, P.G.; Scherer, S.E.; Li, P.W.; Hoskins, R.A.; Galle, R.F. The Genome Sequence of Drosophila melanogaster. *Science*, **2000**, *287*, 2185.

[8] Crosby, M.A.; Goodman, J.L.; Strelets, V.B.; Zhang, P.L.; Gelbart, W.M. FlyBase: genomes by the dozen. *Nucleic Acids Res.*, **2007**, *35*, D486-D491.

[9] Waterston, R.H.; Lindblad-Toh, K.; Birney, E.; Rogers, J.; Abril, J.F.; Agarwal, P.; Agarwala, R.; Ainscough, R.; Alexandersson, M.; An, P. Initial sequencing and comparative analysis of the mouse genome. *Nature*, **2002**, *420*, 520-562.

[10] Weinstock, G.M.; Robinson, G.E.; Gibbs, R.A.; Worley, K.C.; Evans, J.D.; Maleszka, R.; Robertson, H.M.; Weaver, D.B.; Beye, M.; Bork, P. Insights into social insects from the genome of the honeybee Apis mellifera. *Nature*, **2006**, *443*, 931-949.

[11] Castagnoli, L.; Costantini, A.; Dall'Armi, C.; Gonfloni, S.; Montecchi-Palazzi, L.; Panni, S.; Paoluzi, S.; Santonico, E.; Cesareni, G. Selectivity and promiscuity in the interaction network mediated by protein recognition modules. *FEBS Lett.*, **2004**, *567*, 74-79.

[12] Kuriyan, J.; Cowburn, D. Modular peptide recognition domains in eukaryotic signaling. *Annu. Rev. Biophys. Biomol. Struct.*, **1997**, *26*, 259-288.

[13] Doolittle, W.F. The multiplicity of domains in proteins. *Annu. Rev. Biochem.*, **1995**, *64*, 287-314.

[14] Hofmann, K. The modular nature of apoptotic signaling proteins. *Cell. Mol. Life Sci.*, **1999**, *55*, 1113-1128.

[15] Anantharaman, V.; Koonin, E.V.; Aravind, L. Regulatory potential, phyletic distribution and evolution of ancient, intracellular small-molecule-binding domains. *J. Mol. Biol.*, **2001**, *307*, 1271-1292.

[16] Orengo, C.A.; Thornton, J.M. Protein families and their evolution: a structural perspective. *Annu. Rev. Biochem.*, **2005**, *74*, 867-900.

[17] Han, J.; Batey, S.; Nickson, A.A.; Teichmann, S.A.; Clarke, J. The folding and evolution of multidomain proteins. *Nat. Rev. Mol. Cell Biol.*, **2007**, *8*, 319-330.

[18] Wilson, D.; Madera, M.; Vogel, C.; Chothia, C.; Gough, J. The SUPERFAMILY database in 2007: families and functions. *Nucleic Acids Res.*, **2007**, *35*, D308-313.

[19] Ponting, C.P.; Schultz, J.; Milpetz, F.; Bork, P. SMART: identification and annotation of domains from signalling and extracellular protein sequences. *Nuceic Acids Res.*, **1999**, *27*, 229-32.

[20] Ureta-Vidal, A.; Ettwiller, L.; Birney, E. Comparative genomics: genome-wide analysis in metazoan eukaryotes. *Nat. Rev. Genet.*, **2003**, *4*, 251-262.

[21] Bin Qian, R.A.G. Distribution of indel lengths. *Proteins Struct. Funct. Genet.*, **2001**, *45*, 102-104.

[22] Rosenberg, M.S.; Kumar, S. Heterogeneity of nucleotide frequencies among evolutionary lineages and phylogenetic inference. *Mol. Biol. Evol.*, **2003**, *20*, 610-621.

[23] Vingron, M.; Waterman, M.S. Sequence alignment and penalty choice. Review of concepts, case studies and implications. *J. Mol. Biol.*, **1994**, *235*, 1-12.

[24] Bromham, L. Who do species vary in their rate of molecular evolution. *Biol. Lett.*, **2009**, *5*, 401-404.

[25] Eckerle, L.C.; Becker, M.M.; Halpin, R.A.; Li, K.; Venter, E.; Lu, X.; Scherbakova, S.; Graham, R.L.; Baric, R.S.; Stockwell, T.B.; Spiro, D.J.; Denison, M.R. Infidelity of SARS-CoV Nsp14-exonuclease mutant virus replication is revealed by complete genome sequencing. *PLoS Path.*, **2010**, *6*, e1000896.

[26] Galperin, M.Y.; Koonin, E.V. Who's your neighbor? New computational approaches for functional genomics. *Nat. Biotech.*, **2000**, *18*, 609-613.

[27] Eisenberg, D.; Marcotte, E.M.; Xenarios, I.; Yeates, T.O. Protein function in the post-genomic era. *Nature*, **2000**, *405*, 823-826.

[28] Attwood, T.K. The role of pattern databases in sequence analysis. *Briefings in Bioinformatics*, **2000**, *1*, 45-49.

[29] Ashburner, M.; Ball, C.A.; Blake, J.A.; Botstein, D.; Butler, H.; Cherry, J.M.; Davis, A.P.; Dolinski, K.; Dwight, S.S.; Eppig, J.T.; Harris, M.A.; Hill, D.P.; Issel-Tarver, L.; Kasarskis, A.; Lewis, S.; Matese, J.C.; Richardson, J.E.; Ringwald, M.; Rubin, G.M.; Sherlock, G. Gene Ontology: tool for the unification of biology. *Nat. Genet.*, **2000**, *25*, 25-29.

[30] Snijder, E.J.; Bredenbeek, P.J.; Dobbe, J.C.; Thiel, V.; Ziebuhr, J.; Poon, L.L.; Guan, Y.; Rozanov, M.; Spaan, W.J.; Gorbalenya, A.E. Unique and conserved features of genome and proteome of SARS-coronavirus, an early split-off from the coronavirus group 2 lineage. *J. Mol. Biol.*, **2003**, *331*, 991-1004.

[31] Felsenstein, J. PHYLIP version 3.63. *Deptartment of Genetics, University of Washington, Seattle*, **2004**.

[32] Altschul, S.F.; Madden, T.L.; Schaffer, A.A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D.J. Gapped BLAST and PSI-BLAST: a new

generation of protein database search programs. *Nucleic Acids Res.*, **1997**, *25*, 3389-3402.

[33] Thompson, J.D.; Gibson, T.J.; Plewniak, F.; Jeanmougin, F.; Higgins, D.G. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.*, **1997**, *25*, 4876-82.

[34] Pearson, W.R. Comparison of methods for searching protein sequence databases. *Protein Sci.*, **1995**, *4*, 1145-1160.

[35] Apic, G.; Gough, J.; Teichmann, S.A. An insight into domain combinations. *Bioinformatics*, **2001**, *17*, S83-89.

[36] Felsenstein, J. Evolutionary trees from DNA sequences: a maximum likelihood approach. *J. Mol. Evol.*, **1981**, *17*, 368-376.

[37] Huelsenbeck, J.P.; Ronquist, F. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*, **2001**, *17*, 754-755.

[38] Springer, M.S. Mammalian evolution and biomedicine: new views from phylogeny. *Biol. Rev.*, **2007**, *82*, 375-392.

[39] Kumar, S.; Filipski, A. Multiple sequence alignment: In pursuit of homologous DNA positions. *Genome Res.*, **2007**, *17*, 127-135.

[40] Lunter, G.; Miklos, I.; Drummond, A.; Jensen, J.; Hein, J. Bayesian coestimation of phylogeny and sequence alignment. *BMC Bioinformatics*, **2005**, *6*, 83.

[41] Chothia, C.; Lesk, A.M. The relation between the divergence of sequence and structure in proteins. *EMBO J.*, **1986**, *5*, 823-826.

[42] te Velthuis, A.; Admiraal, J.; Bagowski, C. Molecular evolution of the MAGUK family in metazoan genomes. *BMC Evol. Biol.*, **2007**, *7*, 129.

[43] Codoner, F.M.; Fares, M.A. Why should we care about molecular coevolution. *Evol. Bioinform. Online*, **2008**, *4*, 29-38.

[44] Freire, E. The propagation of binding interactions to remote sites in proteins: analysis of the binding of the monoclonal antibody D1.3 to lysozyme. *Proc. Natl. Acad. Sci. USA*, **1999**, *96*, 10118-10122.

[45] Luque, I.; Freire, E. Structural stability of binding sites: consequences for binding affinity and allosteric effects. *Proteins*, **2000**, *(Suppl 4)*, 63-71.

[46] Hidalgo, P.; MacKinnon, R. Revealing the architecture of a K+ channel pore through mutant cycles with a peptide inhibitor. *Science*, **1995**, *268*, 307-310.

[47] Atwell, S.; Ultsch, M.; De Vos, A.M.; Wells, J.A. Structural plasticity in a remodeled protein-protein interface. *Science*, **1997**, *278*, 1125-1128.

[48] Tonikian, R.; Sazinsky, S.; Currell, B.; Yeh, J.; Reva, B.; Held, H.; Appleton, B.; Evangelista, M.; Wu, Y.; Xin, X.; Chan, A.; Seshagiri, S.; Lasky, L.; Sander, C.; Boone, C.; Bader, G.; Sidhu, S. A specificity map for the PDZ domain family. *PLoS Biol.*, **2008**, *6*, e239.

[49] Luque, I.; Leavitt, S.; Freire, E. The linkage between protein folding and functional cooperativity: two sides of the same coin? *Annu. Rev. Biophys. Biomol. Struct.*, **2002**, *31*, 235-256.

[50] Benner, S.A.; Cohen, M.A.; Gonnet, G.H. Empirical and structural models for insertions and deletions in the divergent evolution of proteins. *J. Mol. Biol.*, **1993**, *20*, 1065-1082.

[51] Pascarella, S.; Argos, P. Analysis of insertions/deletions in protein structures. *J. Mol. Biol.*, **1992**, *224*, 461-471.

[52] Panchenko, A.; Madej, T. Structural similarity of loops in protein families: toward the understanding of protein evolution. *BMC Evol. Biol.*, **2005**, *5*, 10.

[53] Todd, M.J.; Freire, E. The effect of inhibitor binding on the structural stability and cooperativity of the HIV-1 protease. *Proteins*, **1999**, *36*, 147-156.

[54] Lockless, S.W.; Ranganathan, R. Evolutionary conserved pathways of energetic connectivity in protein families. *Science*, **1999**, *286*, 295-299.

[55] Neher, E. How frequent are correlated changes in families of protein sequences? *Proc. Natl. Acad. Sci. USA*, **1994**, *91*, 98-102.

[56] Fitch, W.M.; Markowitz, E. An improved method for determining codon variability in a gene and its application to the rate of fixation of mutations in evolution. *Biochem. Genet.*, **1970**, *4*, 579-593.

[57] Thomas, J.W.; Touchman, J.W.; Blakesley, R.W.; Bouffard, G.G.; Beckstrom-Sternberg, S.M.; Margulies, E.H.; Blanchette, M.; Siepel, A.C.; Thomas, P.J.; McDowell, J.C.; Maskeri, B.; Hansen, N.F.; Schwartz, M.S.; Weber, R.J.; Kent, W.J.; Karolchik, D.; Bruen, T.C.; Bevan, R.; Cutler, D.J.; Schwartz, S.; Elnitski, L.; Idol, J.R.; Prasad, A.B.; Lee-Lin, S.Q.; Maduro, V.V.B.; Summers, T.J.; Portnoy, M.E.; Dietrich, N.L.; Akhter, N.; Ayele, K.; Benjamin, B.; Cariaga, K.; Brinkley, C.P.; Brooks, S.Y.; Granite, S.; Guan, X.; Gupta, J.; Haghighi, P.; Ho, S.L.; Huang, M.C.; Karlins, E.; Laric, P.L.; Legaspi, R.; Lim, M.J.; Maduro, Q.L.; Masiello, C.A.; Mastrian, S.D.; McCloskey, J.C.; Pearson, R.; Stantripop, S.; Tiongson, E.E.; Tran, J.T.; Tsurgeon, C.; Vogt, J.L.; Walker, M.A.; Wetherby, K.D.; Wiggins, L.S.; Young, A.C.; Zhang, L.H.; Osoegawa, K.; Zhu, B.; Zhao, B.; Shu, C.L.; De Jong, P.J.; Lawrence, C.E.; Smit, A.F.; Chakravarti, A.; Haussler, D.; Green, P.; Miller, W.; Green, E.D. Comparative analyses of multi-species sequences from targeted genomic regions. *Nature*, **2003**, *424*, 788-793.

[58] Premzl, M.; Gready, J.E.; Jermiin, L.S.; Simonic, T.; Marshall Graves, J.A. Evolution of vertebrate genes related to prion and shadoo proteins--clues from comparative genomic analysis. *Mol. Biol. Evol.*, **2004**, *21*, 2210-2231.

[59] Siepel, A.; Bejerano, G.; Pedersen, J.S.; Hinrichs, A.S.; Hou, M.; Rosenbloom, K.; Clawson, H.; Spieth, J.; Hillier, L.W.; Richards, S. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.*, **2005**, *15*, 1034-1050.

[60] Andreeva, A.; Howorth, D.; Chandonia, J.-M.; Brenner, S.E.; Hubbard, T.J.P.; Chothia, C.; Murzin, A.G. Data growth and its impact on the SCOP database: new developments. *Nucleic Acids Res.*, **2008**, *36*, D419-425.

[61] Wolf, Y.I.; Grishin, N.V.; Koonin, E.V. Estimating the number of protein folds and families from complete genome data. *J. Mol. Biol.*, **2000**, *299*, 897-904.

[62] Koonin, E.V.; Fedorova, N.D.; Jackson, J.D.; Jacobs, A.R.; Krylov, D.M.; Makarova, K.S.; Mazumder, R.; Mekhedov, S.L.; Nikolskaya, A.N.; Rao, B.S. A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes. *Genome Biol.*, **2004**, *5*, R7.

[63] te Velthuis, A.J.W.; Isogai, T.; Gerrits, L.; Bagowski, C.P. Insights into the molecular evolution of the PDZ-LIM family and indentification of a novel conserved protein motif. *PLoS ONE*, **2007**, *2*, e189.

[64] Markov, G.V.; Tavares, R.; Dauphin-Villemant, C.; Demeneix, B.A.; Baker, M.E.; Laudet, V. Independent elaboration of steroid hormone signaling pathways in metazoans. *Proc. Natl. Acad. Sci. USA*, **2009**, *106*, 11913-11918.

[65] Lercher, M.J.; Pal, C. Integration of horizontally transferred genes into regulatory interaction networks takes many million years. *Mol. Biol. Evol.*, **2008**, 25, 559-67.

[66] Gogarten, J.P.; Doolittle, W.F.; Lawrence, J.G. Prokaryotic evolution in light of gene transfer. *Mol. Biol. Evol.*, **2002**, *19*, 2226-2238.

[67] Wagner, A. How the global structure of protein interaction networks evolves. *Proc. Biol. Sci.*, **2003**, *270*, 280-284.

[68] Koonin, E.V.; Aravind, L.; Kondrashov, A.S. The impact of comparative genomics on our understanding of evolution. *Cell*, **2000**, *101*, 573-576.

[69] Cohen-Gihon, I.; Lancet, D.; Yanai, I. Modular genes with metazoan-specific domains have increased tissue specificity. *Trends Genet.*, **2005**, *21*, 210-213.

[70] Basu, M.K.; Carmel, L.; Rogozin, I.B.; Koonin, E.V. Evolution of protein domain promiscuity in eukaryotes. *Genome Res.*, **2008**, *18*, 449-461.

[71] Koonin, E.V.; Wolf, Y.I.; Karev, G.P. The structure of the protein universe and genome evolution. *Nature*, **2002**, *420*, 218-223.

[72] Tordai, H.; Nagy, A.; Farkas, K.; Bányai, L.; Patthy, L. Modules, multidomain proteins and organismic complexity. *FEBS J.*, **2005**, *272*, 5064-5078.

[73] Marcotte, E.M.; Pellegrini, M.; Ng, H.L.; Rice, D.W.; Yeates, T.O.; Eisenberg, D. Detecting protein function and protein-protein interaction from genome sequences. *Science*, **1999**, *285*, 751-753.

[74] Jeong, H.; Mason, S.P.; Barabasi, A.L.; Oitvai, Z.N. Lethality and centrality in protein networks. *Nature*, **2001**, *411*, 41-42.

[75] Hahn, M.W.; Kern, A.D. Comparative genomics of centrality and essentiality in three eukaryotic protein-interaction networks. *Mol. Biol. Evol.*, **2005**, *22*, 803-806.

[76] Wiener, J.; Beaussart, F.; Bornberg-Bauer, E. Domain deletions and substitutions in the modular protein evolution. *FEBS J.*, **2006**, *273*, 2037-2047.

[77] Patthy, L. Genome evolution and the evolution of exon-shuffling-a review. *Gene*, **1999**, *238*, 103-114.

[78] Liu, M.; Walch, H.; Wu, S.; Grigoriev, A. Significant expansion of exon-bordering protein domains during animal proteome evolution. *Nucleic Acids Res.*, **2005**, *33*, 95-105.

[79] Basdevant, N.; Weinstein, H.; Ceruso, M. Thermodynamic basis for promiscuity and selectivity in protein-protein interactions: PDZ domains, a case study. *J. Am. Chem. Soc.* **2006**, *128*, 12766-12777.

[80] Agrawal, V.; Kishan, K.V. Promiscuous binding nature of SH3 domains to their target proteins. *Protein Pept. Lett.*, **2002**, *9*, 185-193.

[81] Betran, E.; Long, M. Expansion of genome coding regions by acquisition of new genes. *Genetica*, **2002**, *115*, 65-80.

[82] Bashton, M.; Chothia, C. The geometry of domain combination in proteins. *J. Mol. Biol.*, **2002**, *315*, 927-939.

[83] Kim, E.; Magen, A.; Ast, G. Different levels of alternative splicing among eukaryotes. *Nucleic Acids Res.*, **2007**, *35*, 125-131.

[84] Ast, G. How did alternative splicing evolve? *Nat. Rev. Genet.*, **2004**, *7*, 773-781.

[85] Kopelman, N.M.; Lancet, D.; Yanai, I. Alternative splicing and gene duplication are inversely correlated evolutionary mechanisms. *Nat. Genet.*, **2005**, *37*, 588-589.

[86] Adams, K.L.; Wendel, J.F. Polyploidy and genome evolution in plants. *Curr. Opin. Plant. Biol.*, **2005**, *8*, 135-141.

[87] Severing, E.I.; van Dijk, A.D.; Stiekema, W.J.; van Ham, R.C. Comparative analysis indicates that alternative splicing in plants has a limited role in functional expansion of the proteome. *BMC Genomics*, **2009**, *10*, 154.

[88] Tavares, G.A.; Panepucci, E.H.; Brunger, A.T. Structural characterization of the intramolecular interaction between the SH3 and guanylate kinase domains of PSD-95. *Mol. Cell*, **2001**, *8*, 1313-1325.

[89] McGee, A.W.; Bredt, D.S. Identification of an Intramolecular Interaction between the SH3 and Guanylate Kinase Domains of PSD-95. *J. Biol. Chem.*, **1999**, *274*, 17431-17436.

[90] Krojer, T.; Pangerl, K.; Kurt, J.; Sawa, J.; Stingl, C.; Mechtler, K.; Huber, R.; Ehrman, M.; Clausen, T. Interplay of PDZ and protease domain of DegP ensures efficient elimination of misfolded proteins. *Proc. Natl. Acad. Sci. USA*, **2009**, *105*, 7702-7707.

[91] Tirosh, I.; Bilu, Y.; Barkai, N. Comparative biology: beyond sequence analysis. *Curr. Opin. Biotechnol.*, **2007**, *18*, 371-377.

[92] Tirosh, I.; Weinberger, A.; Carmi, M.; Barkai, N. A genetic signature of interspecies variations in gene expression. *Nat. Genet.*, **2006**, *38*, 830-834.

[93] Landry, C.R.; Oh, J.; Hartl, D.L.; Cavalieri, D. Genome-wide scan reveals that genetic variation for transcriptional plasticity in yeast is biased towards multi-copy and dispensable genes. *Gene*, **2006**, *366*, 343-351.

[94] Hooper, S.D.; Boue, S.; Krause, R.; Jensen, L.J.; Mason, C.E.; Ghanim, M.; White, K.P.; Furlong, E.E.M.; Bork, P. Identification of tightly regulated groups of genes during Drosophila melanogaster embryogenesis. *Mol. Syst. Biol.*, **2007**, *3*, 72.

[95] Stuart, J.M.; Segal, E.; Koller, D.; Kim, S.K. A gene-coexpression network for global discovery of conserved genetic modules. *Science*, **2003**, *302*, 294-255.

[96] Bergmann, S.; Ihmels, J.; Barkai, N. Similarities and Differences in Genome-Wide Expression Data of Six Organisms. *PLoS Biol.*, **2004**, *2*, e9.

[97] Burger, L.; van Nimwegen, E. Accurate prediction of protein-protein interactions from sequence alignments using a Bayesian method. *Mol. Syst. Biol.*, **2008**, *4*, 165.

[98] Pazos, F.; Helmer-Citterich, M.; Ausiello, G.; Valencia, A. Correlated mutations contain information about protein-protein interaction. *J. Mol. Biol.*, **1997**, *271*, 511-535.

[99] Mateu, M.G.; Fersht, A.R. Mutually compensatory mutations during evolution of the tetramerization domain of tumor suppressor p53 lead to impaired hetero-oligomerization. *Proc. Natl. Acad. Sci. USA*, **1999**, *96*, 3595-3599.

[100] Gavin, A.; Bosche, M.; Krause, R.; Grandi, P.; Marzioch, M.; Bauer, A.; Schultz, J.; Rick, J.M.; Michon, A.; Cruciat, C.; Remor, M.; Hofert, C.; Schelder, M.; Brajenovic, M.; Ruffner, H.; Merino, A.; Klein, K.; Hudak, M.; Dickson, D.; Rudi, T.; Gnau, V.; Bauch, A.; Bastuck, S.; Huhse, B.; Leutwein, C.; Heurtier, M.; Copley, R.R.; Edelmann, A.; Querfurth, E.; Rybin, V.; Drewes, G.; Raida, M.; Bouwmeester, T.; Bork, P.; Seraphin, B.; Kuster, B.; Neubauer, G.; Superti-Furga, G. Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature*, **2002**, *415*, 141-147.

[101] Ho, Y.; Gruhler, A.; Heilbut, A.; Bader, G.D.; Moore, L.; Adams, S.L.; Millar, A.; Taylor, P.; Bennett, K.; Boutilier, K.; Yang, L.; Wolting, C.; Donaldson, I.; Schandorff, S.; Shewnarane, J.; Vo, M.; Taggart, J.; Goudreault, M.; Muskat, B.; Alfarano, C.; Dewar, D.; Lin, Z.; Michalickova, K.; Willems, A.R.; Sassi, H.; Nielsen, P.A.; Rasmussen, K.J.; Andersen, J.R.; Johansen, L.E.; Hansen, L.H.; Jespersen, H.; Podtelejnikov, A.; Nielsen, E.; Crawford, J.; Poulsen, V.; Sørensen, B.D.; Matthiesen, J.; Hendrickson, R.C.; Gleeson, F.; Pawson, T.; Moran, M.F.; Durocher, D.; Mann, M.; Hogue, C.W.; Figeys, D.; Tyers, M. Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry. *Nature*, **2002**, *415*, 180-183.

[102] Stelzl, U.; Worm, U.; Lalowski, M.; Haenig, C.; Brembeck, F.H.; Goehler, H.; Stroedicke, M.; Zenkner, M.; Schoenherr, A.; Koeppen, S.; Timm, J.; Mintzlaff, S.; Abraham, C.; Bock, N.; Kietzmann, S.; Goedde, A.; Toksöz, E.; Droege, A.; Krobitsch, S.; Korn, B.; Birchmeier, W.; Lehrach, H.; Wanker, E.E. A human protein-protein interaction network: a resource for annotating the proteome. *Cell*, **2005**, *122*, 957-968.

[103] Prachumwat, A.; Li, W.H. Protein function, connectivity, and duplicability in yeast. *Mol. Biol. Evol.*, **2006**, *23*, 30-39.

[104] Wuchty, S. Evolution and topology in the yeast protein interaction network. *Genome Res.*, **2004**, *14*, 1310-1314.

[105] Fraser, H.B. Modularity and evolutionary constraint on proteins. *Nat. Genet.*, **2005**, *37*, 351-352.

[106] McCarroll, S.A.; Murphy, C.T.; Zou, S.; Pletcher, S.D.; Chin, C.-S.; Jan, Y.N.; Kenyon, C.; Bargmann, C.I.; Li, H. Comparing genomic expression patterns across species identifies shared transcriptional profile in aging. *Nat. Genet.*, **2004**, *36*, 197-204.

[107] Jeon, J.; Park, S.; Chi, M.; Choi, J.; Park, J.; Rho, H.; Kim, S.; Goh, J.; Yoo, S.; Choi, J.; Park, J.; Yi, M.; Yang, S.; Kwon, M.; Han, S.; Kim, B.R.; Khang, C.H.; Park, B.; Lim, S.; Jung, K.; Kong, S.; Karunakaran, M.; Oh, H.; Kim, H.; Kim, S.; Park, J.; Kang, S.; Choi, W.; Kang, S.; Lee, Y. Genome-wide functional analysis of pathogenicity genes in the rice blast fungus. *Nat. Genet.*, **2007**, *39*, 561-565.

[108] Rifkin, S.A.; Houle, D.; Kim, J.; White, K.P. A mutation accumulation assay reveals a broad capacity for rapid evolution of gene expression. *Nature*, **2005**, *438*, 220-223.

[109] Rifkin, S.A.; Kim, J.; White, K.P. Evolution of gene expression in the Drosophila melanogaster subgroup. *Nat. Genet.*, **2003**, *33*, 138-144.

[110] Ranz, J.M.; Castillo-Davis, C.I.; Meiklejohn, C.D.; Hartl, D.L. Sex-dependent gene expression and evolution of the Drosophila transcriptome. *Science*, **2003**, *300*, 1742-1745.

[111] White, K.P.; Rifkin, S.A.; Hurban, P.; Hogness, D.S. Microarray analysis of Drosophila development during metamorphosis. *Science*, **1999**, *286*, 2179-2184.

[112] Jordan, I.K.; Marino-Ramirez, L.; Wolf, Y.I.; Koonin, E.V. Conservation and coevolution in the scale-free human gene coexpression network. *Mol. Biol. Evol.*, **2004**, *21*, 2058-2070.

[113] Oldham, M.C.; Horvath, S.; Geschwind, D.H. Conservation and evolution of gene coexpression networks in human and chimpanzee brains. *Proc. Natl. Acad. Sci. USA*, **2006**, *103*, 17973-17978.

[114] Frazer, K.A.; Elnitski, L.; Church, D.M.; Dubchak, I.; Hardison, R.C. Cross-species sequence comparisons: a review of methods and available resources. *Genome Res.*, **2003**, *13*, 1-12.

[115] Blouin, C.; Butt, D.; Roger, A.J. Impact of taxon sampling on the estimation of rates of evolution at sites. *Mol. Biol. Evol.*, **2005**, *22*, 784-791.

[116] Torarinsson, E.; Yao, Z.; Wiklund, E.D.; Bramsen, J.B.; Hansen, C.; Kjems, J.; Tommerup, N.; Ruzzo, W.L.; Gorodkin, J. Comparative genomics beyond sequence-based alignments: RNA structures in the ENCODE regions. *Genome. Res.*, **2008**, *18*, 242-251.

[117] Ott, E.B.; Te Velthuis, A.J.W.; Bagowski, C.P. Comparative analysis of splice form-specific expression of LIM Kinases during zebrafish development. *Gene Expr. Patterns*, **2007**, *7*, 620-629.

[118] Bork, P.; Serrano, L. Towards cellular systems in 4D. *Cell*, **2005**, *121*, 507-509.

[119] Yadav, R.K.; Girke, T.; Pasala, S.; Xie, M.; Reddy, G. Gene expression map of the Arabidopsis shoot apical meristem stem cell niche. *Proc. Natl. Acad. Sci. USA*, **2009**, *106*, 4941-4946.

[120] Schmid, M.; Davison, T.S.; Henz, S.R.; Pape, U.J.; Demar, M.; Vingron, M.; Schölkopf, B.; Weigel, D.; Lohmann, J.U. A gene expression map of Arabidopsis thaliana development. *Nat. Genet.*, **2005**, *37*, 501-506.