

## The Structural Consequences after Protein Domain Duplication Events

Jun Lin\*, Leping Feng

Guilin Medical University, Guilin, 541004 and State Key Laboratory of Genetic Resources and Evolution,  
Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, 650223, China

\* Correspondence should be addressed to J.L. (E-mail: [linj@mspil.edu.cn](mailto:linj@mspil.edu.cn))

**Abstract:** Protein structures involving gene duplication and gene fusion events can be highly illuminated, and researchers have long sought fundamental explanations for evolutionary origins of duplicated protein structures. How a peptide containing two similar domains, we called it *Gemini* gene/protein, evolved. For examples, the structure of 1vk6 (NADH pyrophosphatase) consists of two structurally similar domains, and the N-terminal domain holds a rudiment Nudix fold, while the C-terminal domain has the canonical Nudix fold. In this paper, we show several examples of two similar domains in one single peptide. Our work focused on the protein structural consequences after the duplication-fusion events in gene innovation, which may help further understand a gene or protein domain's evolutionary fate after gene duplication event.

[Jun Lin, Leping Feng. **The Structural Consequences after Protein Domain Duplication Events.** *Life Sci J* 2013;10(3):610-614] (ISSN:1097-8135). <http://www.lifesciencesite.com>. 90

**Keywords:** domain duplication; duplication-fusion; Identical Domain Insertion (IDI); Circular Permutation (CP); origin of new genes

### 1. Introduction

More than forty years ago, Susumu Ohno stated the gene duplication was the most important factor underlying genome evolution (Ohno, 1970). The most common fate of gene duplication appears to be the simple silencing of one member of the pair (non-functionalization or pseudogenization). A second outcome may be that one copy acquires a novel function (neo-functionalization) with the other one retaining the original function. Another possibility is that the ancestral functions are split between the duplicated gene pairs (sub-functionalization). However, gene fusion and fission events are common during the evolutionary process, and these events have been observed in various organisms. Duplication events often result in protein domains which can be regulated separately. Fusion genes, also known as chimeric genes, are formed when the reading frames of two or more genes are joined by recombination events (Leonard et al., 2012). In the fields of protein structure analysis that one of the major mechanisms by which proteins evolve is through gene duplication and modification of the resulting duplicated gene products. Many proteins contain domains that have clearly arisen through duplication and fusion events. The duplication event could overwrite a stop codon, a fusion between two duplicates of the same gene can be created. Depending on the subsequent fate of the two copies, this can have little or no effect on the structure or result in such large changes that the relationship between the novel protein and its ancestor can be obscured completely (Leonard et al., 2012). The human kua-UEV (Long, 2000) and the double MutT/Nudix domain protein (Lin et al., 2008; Lin et

al., 2009) are examples. Gene fusion events, the recombination of domains at the protein level, played important roles in building the initial gene structure. Generally, a gene duplication-fusion event is a process that a gene duplication event is followed by a gene fusion event (Cousineau et al., 1997). Two enzymes in the pathway of histidine biosynthesis, HisA and HisF, from a hyperthermophile bacterium *Thermotoga maritima* possess an eightfold alpha/beta barrel structure known as the TIM (Triosephosphate IsoMerase) barrel. The structure evolved from a duplication-fusion event of the gene of a common half-barrel ancestor, followed by diversification of catalytic function; the ancestral barrel structure was probably a homodimer consisting of two identical half barrels (Lang et al., 2000).

Circular permutations are a frequent event in molecular evolution, and they have been observed in many protein families and superfamilies (Lindqvist et al., 1997). The evolutionary mechanisms of circular permutation in DNA methyltransferase genes (Jeltsch, 1999) and tyrosine phosphatase superfamilies (Huang, 2003) have been illustrated. However, circular permutation can perturb local tertiary structure, resulting in improved protein catalytic activity. Protein engineering has benefits of reorganizing the polypeptide chain of a protein by circular permutation (Yu et al., 2011).

In this paper, we show five examples of two similar domains in one single peptide form different protein families. Our work focused on the protein structural consequences after the duplication-fusion events, which may help to further understand a gene or protein domain's evolutionary fate after gene

duplication events. We describe the effects that these events can have on protein structures and study the examples of each type of event that have been observed. Finally, we synthesize a general evolutionary model of the protein sequences with two similar structures.

## 2. Material and Methods

The structural data are from the Protein Data Bank (PDB), and the PDB codes of three-dimensional structures are 1jsg, 1vk6, 2ch4, 1wv3 and 1vgl. All structural classification of structures that obtained from the PDB database was done by SCOP and CATH database (Lewis et al., 2013). The structures were visualized by Raswin program and structure superimpose was performed by VMD software. To obtain an alignment, we linked that some segment to its C-terminal and some large insertion regions in sequence have been deleted. The results of structure superposition and structure-based alignments obtained from different programs are checked by graphics. The sequences of from PDB are taken as the seeds with which to search the Uniprot by PSI-Blast (Altschul et al., 2009). E-values  $\leq 0.005$  are used in the database searches. The sequences hit by seeds are analyzed and aligned by the ClustalW program (Thompson et al., 2012). The sequence is divided into four putative fragments, subdomain I, II, III, and IV, based on its structures. All four fragments are aligned by ClustalW, and they are linked in different orders. These fragments with the different orders are aligned with the different sequences.

## 3. Results

### 3.1 Common domain duplication and fusion of proteins of oncogene products and Nudix fold

The crystal structure 1jsg of protein p14-TCL1 (Hoh et al., 1998), an oncogene product from Homo sapiens involved in T-cell prolymphocytic leukemia, consists of two structurally similar domains. The crystal structure of 1jsg N-terminal domain (1jsg chain a: 12-57) and C-terminal domain (1jsg chain a: 70-114) proved that the two domains have similar structures and possess a half beta barrel. Moreover, the two domains' sequences were aligned based on structures showed that the region overlapped with 39 equal residues between two structures, having a RMSD (Root Mean Square Deviation) value equal to 1.85 angstroms. The high sequence identity (87%) revealed significant similarity between two structures. Two similar structures are in one single molecule and form a beta barrel. Because of sequences' similarity between two domains, it is more likely that the two domains are homologues from an ancient common ancestor. This molecule was caused by an ancient gene duplication-fusion event. In this scenario, the fusion event occurred after gene duplication event forming the beta barrel. The structure of 1vk6 (NADH

pyrophosphatase) consists of two structurally similar domains, and the N-terminal domain holds a rudiment Nudix fold, while the C-terminal domain has the canonical Nudix fold. The two individual domains underwent different fates and the molecule was result from an ancient gene duplication-fusion event. In this scenario, the fusion event occurred after gene duplication event forming the double Nudix domain protein (Figure 1).

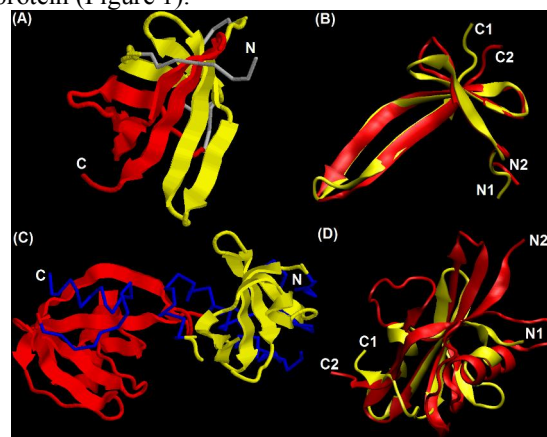


Figure 1 Two examples for protein domain duplication-fusion events (A) Two halves beta sheet form a barrel; yellow and red: 12-57 and 70-114 residues of 1jsg chain A (B) Structure superimpose of two halves' domain (C) Two Nudix domains in one chain yellow and red: 8-68 and 126-229 residues of 1vk6 chain A (D) Structure superimpose of two Nudix domain. N, N1, and N2: N-terminal; C, C1, and C2: C-terminal.

### 3.2 Identical Domain Insertion

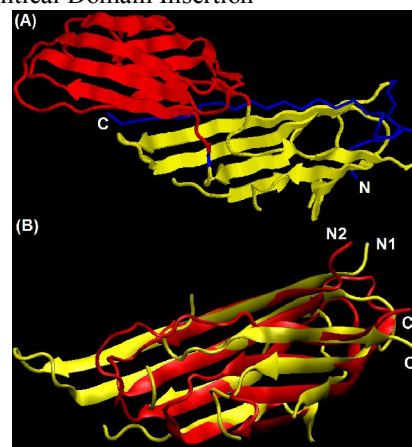


Figure 2 Identical domain insertion (A) two obvious similar structures in one chain; blue: linker regions (B) structure superimpose of two protein domains. red: the inserted domain (2vgl m: 284-393); yellow: the domain (2vgl m: 170-283, 392-409) was inserted by another identical one; N, N1, and N2: N-terminal; C, C1, and C2: C-terminal.

The structure of 2vgl (Collins et al., 2002) is an example of identical domain insertion (IDI). The IDI event (Figure 2) is one domain inserted by an identical domain in protein sequence and formed a dimer structure. The adaptor complexes medium subunit domain (1vgl) is similar to that found in cytochrome f and certain transcription factors and the domain found at the C-terminus of the mu subunit from various clathrin adaptors (AP1, AP2 and AP3). The C-terminal domain has an immunoglobulin-like beta-sandwich fold consisting of 9 strands in 2 sheets with a Greek key topology and Adaptor complexes medium subunit family also contains members which are coatomer subunits (Figure 3).

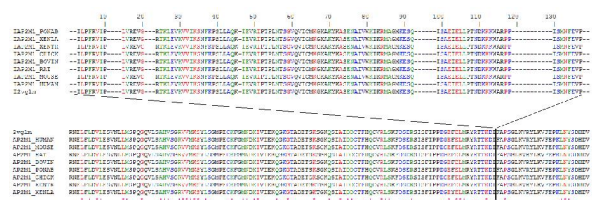


Figure 3 Sequence alignment of two identical domains of 2vgl homologous sequence

The mu subunit possesses a conserved N-terminal domain which may be the region of interaction with other AP proteins and a less conserved C-terminal domain (Figure 3). Tandem duplication, a mutational process for DNA molecules in which a short stretch of DNA is transformed into several adjacent copies, is not well understood, unequal recombination is widely thought as the key biological mechanism responsible for it (Fitch, 1977). However, the evolution process of IDI event remains unclear.

### 3.3 Circular permutation process in two-domain proteins

The structures of 2ch4 and 1wv3 are the protein structure involved in domain duplication and circular permutation events (Figure 4). Circular permutations are a frequent event in molecular evolution, and they have been observed in many protein families and superfamilies. The 1wv3 protein's N-terminal domain is original domain while C-terminal one is circularly permuted (Figure 4B).

The CheW proteins (PDB:2ch4 chain w) interact with the methyl accepting chemotaxis proteins (MCPs) and relay signals to CheY and are part of the chemotaxis signaling mechanism in bacteria. This protein affects flageller rotation containing a domain composed of two beta-sheet subdomains, each of which forms a five-stranded beta-barrel around the hydrophobic core. The CheW domain is an around 150-residue domain involved in the two-component

signaling systems regulating bacterial chemotaxis. The CheW protein contains two OB-fold domains, and the canonical domain inserted into the circularly permuted one (Figure 4A). Some of proteins with circular permutation have been confirmed to be the results of divergent evolution (Malone et al., 1995; Ponting et al., 1995; Lupas, 1996). However, because of sequences' similarity between two domains, it is more likely that the two domains are homologues from an ancient common ancestor. A possible evolution scenario is the full ancestral gene of the protein that had contained in transcript unit with three or four genes, such as I-II-III-IV-I-II-III-IV-I-II-III-IV, can be formed by two-round gene duplication and fusion events. The protein molecule can be obtained by reading this transcript unit in a reading frame with bold-type numbered subdomains (IV-I-II-III-IV-I-II-III). Thus, the CheW protein family may have resulted from an ancestral gene by the duplications, fusions, and circular permutations. The circular permutation in the CheW protein is similar to the circular permutation in DNA methyltransferases described by Jeltsch (Jeltsch, 1999). The evolutionary mechanism in the CheW protein can be used to explain how the two similar circularly permuted domains in a single peptide can be formed by circular permutations during molecular evolution.

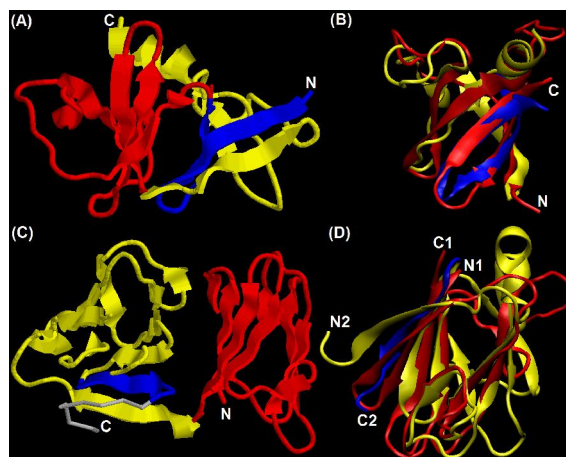


Figure 4 Examples of protein domain duplication & circular permutation (A) two OB-fold structures in one peptide of 2ch4 (B) the superimpose of the two OB-fold structures. red: canonical OB-fold (2ch4 w: 28-96); and circularly permuted OB-fold (2ch4 w: 9-27(blue), 2ch4 w: 97-147(yellow)) (C) EssC N-terminal domain-like protein of crystal structure 1wv3 (D) structural alignment of canonical domain (1wv3 chain a: 1-78(red)) and circularly permuted topology (1wv3 chain a: 79-168(yellow) and 169-177(blue)). N, N1, and N2: N-terminal; C, C1, and C2: C-terminal



The importance of gene duplication in evolution has long been recognized, and three evolutionary consequences of duplicated genes have been described. Furthermore, the evolutionary mechanism of circular permutation in DNA methyltransferase genes has been illustrated (Jeltsch, 1999). The author's work differ from former work is that our paper focus on the phenomena of two similar protein domains in a single peptide. Three type two-domain containing peptides were listed, including common domain duplication-fusion, identical domain duplication and circular permuted two-domain proteins (Figure 5).

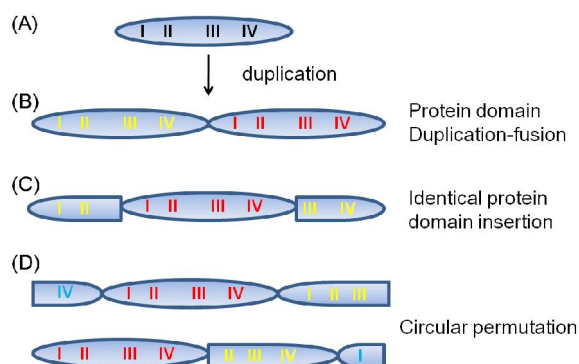


Figure 5 Three structural consequences after protein domain duplication events (A) protein domain ancestor (B)(C)(D) different results of domain architectures after domain duplication. I, II, III, and IV: putative subdomains of the protein domains

#### Acknowledgements:

The project was supported by the Fund to Jun Lin from the State Key Laboratory of Genetics Resources and Evolution of China (Grant No. GREKF10-11), National Science Fund of China (Grant No. 31100963) and partly supported by Program for Innovative Research Team of Guilin Medical University (PIRTGMU). We thank the staff of the State Key Laboratory of Genetics Resources and Evolution of China, CAS for valuable comments.

#### Corresponding Author:

Jun Lin Ph.D.  
Deputy Director  
Department of Biotechnology  
Guilin Medical University  
Guilin, Guangxi 541004, China  
E-mail: [linj@mspil.edu.cn](mailto:linj@mspil.edu.cn)

#### References

- Ohno S. Evolution by gene duplication. Springer-Verlag. 1970.
- Leonard G, Richards TA. Genome-scale comparative analysis of gene fusions, gene fissions, and the fungal tree of life. *Proc Natl Acad Sci USA*. 2012; 109(52):21402-7.
- Long M. A new function evolved from gene fusion. *Genome Res*. 2000; 10(11):1655-7.
- Lin J, Tian B, Hua Y. Structural and functional diversity of Nudix fold. *Protein Pept Lett*. 2008; 15(1):108-12.
- Lin J, Hu Y, Tian B, Hua Y. Evolution of double MutT/Nudix domain-containing proteins: similar domain architectures from independent gene duplication-fusion events. *J Genet Genomics*. 2009; 36(10):603-10.
- Cousineau B, Leclerc F, Cedergren R. On the origin of protein synthesis factors: a gene duplication/fusion model. *J Mol Evol*. 1997; 45(6): 661-70.
- Lang D, Thoma R, Henn-Sax M, Sterner R, Wilmanns M. Structural evidence for evolution of the beta/alpha barrel scaffold by gene duplication and fusion. *Science*. 2000; 289(5484):1546-50.
- Lindqvist Y, Schneider G. Circular permutations of natural protein sequences: structural evidence. *Curr Opin Struct Biol*. 1997; 7(3):422-7.
- Jeltsch A. Circular permutations in the molecular evolution of DNA methyltransferases. *J Mol Evol*. 1999; 49(1):161-4.
- Huang JF. Different protein tyrosine phosphatase superfamilies resulting from different gene reading frames. *Mol Biol Evol*. 2003; 20(5):815-20.
- Yu Y, Lutz S. Circular permutation: a different way to engineer enzyme structure and function. *Trends Biotechnol*. 2011; 29(1):18-25.
- Lewis TE, Sillitoe I, Andreeva A, Blundell TL, Buchan DW, Chothia C, Cuff A, Dana JM, Filippis I, Gough J, Hunter S, Jones DT, Kelley LA, Kleywegt GJ, Minneci F, Mitchell A, Murzin AG, Ochoa-Montano B, Rackham OJ, Smith J, Sternberg MJ, Velankar S, Yeats C, Orengo C. Genome3D: a UK collaborative project to annotate genomic sequences with predicted 3D structures based on SCOP and CATH domains. *Nucleic Acids Res*. 2013; 41 (Database issue):D499-507.
- Altschul SF, Gertz EM, Agarwala R, Schaffer AA, Yu YK. PSI-BLAST pseudocounts and the minimum description length principle. *Nucleic Acids Res*. 2009; 37(3):815-24.
- Thompson JD, Gibson TJ, Higgins DG. *Curr. Protoc. Bioinformatics*. 2002.

15. Hoh F, Yang YS, Guignard L, Padilla A, Stern MH, Lhoste JM, van Tilbeurgh H. Crystal structure of p14TCL1, an oncogene product involved in T-cell prolymphocytic leukemia, reveals a novel beta-barrel topology. *Structure*. 1998; 6(2):147-55.
16. Collins BM, McCoy AJ, Kent HM, Evans PR, Owen DJ. Molecular architecture and functional model of the endocytic AP2 complex. *Cell*. 2002; 109(4):523-35.
17. Fitch WM. Phylogenies constrained by the crossover process as illustrated by human hemoglobins and a thirteen-cycle, eleven-amino-acid repeat in human apolipoprotein A-I. *Genetics*. 1977; 86(3):623-44.
18. Malone T, Blumenthal RM, Cheng X. Structure-guided analysis reveals nine sequence motifs conserved among DNA amino-methyltransferases, and suggests a catalytic mechanism for these enzymes. *J Mol Biol*. 1995; 253(4):618-32.
19. Ponting CP, Russell RB. Swaposins: circular permutations within genes encoding saposin homologues. *Trends Biochem Sci*. 1995; 20(5):179-80.
20. Lupas A. A circular permutation event in the evolution of the SLH domain? *Mol Microbiol*. 1996; 20(4):897-8.

7/1/2013