47 卷 6 期
2007 年 12 月 4 日

微 生 物 学 报
Acta Microbiologica Sinica

47(6):956 ~ 962
4 December 2007

TPR repeats and ELTR pattern: length variation as a function evolution mechanism

YANG Ke-qian*

(*State Key Laboratory of Microbial Resources, Institute of Microbiology, Chinese Academy of Sciences, Beijing* 100101, *China*)

Abstract: TPR repeat was originally defined as a 34 amino acid structural repeat (TPR-34). Equal length tandem repeats (ELTR) was proposed to represent the ancestral repeat pattern. Length polymorphism of TPR repeats was analyzed using PATTINPROT, two new versions of TPR repeat of 40 and 42 amino acids were identified. These 'long' TPRs endow new functional capacities to the resulting proteins. A strong correlation between varied lengths and new functions supports the hypothesis that length variation is an underlying mechanism for the function evolution of repeat containing proteins.

Keywords: TPR, repeat length, tandem duplication, evolution, pattern

CLC number: Q939, Q78 **Document code**: A **Article ID**: 0001-6209(2007)06-0956-07

1 INTRODUCTION

Tandem arrays of nucleotide sequence repeats found on genomes, are classified according to the length of the repeat unit as satellite (several thousand bp), minisatellites (10-1000 bp) and microsatellites (1 – 10 bp). If these repeats occur in protein-coding sequences, will result in amino acid sequence repeats (protein repeats). Protein repeats also vary considerably in length from one to a few amino acid repetitions to large repetitions containing multiple domains. Here we consider a repeat that constitutes a single secondary structural unit in protein (structural repeat, SR, for review, see Andrade *et al*[1]). SR repeat typically occurs tandemly in protein and assembles into a super structure with each repeat unit as the secondary structural element. Therefore it differs conceptually from motif and domain that can exist alone and function independently. The importance of SRs is reflected by their presence in many proteins and the structural or functional roles they mediate[1-3]. To detect and analyze repeats, the task is not simply to detect homology and their arrangements in a protein, but also to detect the "repeat pattern", and to understand the evolution of these patterns, the pattern in SRs is defined by length.

To analyze SR evolution, let's simplify repeat pattern in a biological background. Assume all SRs originated at a genetic level by DNA duplication (not necessarily true), the simplest SR pattern is equal length tandem repeats (ELTR). The purpose of proposing such a hypothetical repeat pattern is many folds: ELTR is easy to visualize or understand; it is probably derived from the simplest (oldest) duplication mechanism; and most importantly, it sets a useful restraint (length) on reoccurring sequence pattern. Most importantly, pattern represents an evolutionary event (marker).

If repeats indeed originated as ELTRs, then the ELTR pattern represents the ancestors, and any deviations from the ELTRs represent descendents; degeneration of ELTR during evolution owing to point mutation, insertion and deletion is easy to understand, the reverse process to generate ELTRs from degenerate repeats is highly unlikely. By identifying ancestors and descendents, establishing their vertical evolutionary relationships and analyzing their distributions in extant organisms we could extract useful evolutionary information. Suppose mutational rate remained constant in different lineages of organisms and over time, the degeneracy of repeating pattern and repeat sequence could reflect the age of the repeat. However, above scenario only represents an over simplified scheme of repeat evolution, recombination between homologous repeats

* Corresponding author. Tel/Fax: 86-10-64807459; E-mail: yangkq@im.ac.cn
Received: 8 May 2007/Accepted: 15 August 2007/Revised: 19 September 2007

often results in repeats of mixed length, this is particularly true when repeats become older. To observe a preferred length is rare, and should be analyzed for its functional significance and evolutionary implications. In theory, there are no restrictions on repeat length, in the case of SRs, the only restriction is structural constraints, i.e. the repeats must be long enough to form stable secondary structures and must have conserved amino acids at conserved positions.

Currently, repeat search methods are homology based that are very flexible on the interval lengths between repeats[3-8], they are powerful to identify irregular and degenerate repeats but do not have restraint on the original pattern of the repeat.

The classical tetratricopeptide repeat (TPR) was first identified in cell division cycle (CDC) proteins in yeast as a degenerate 34-amino acid repeat[9,10], with the name reiterating its 34 amino acids length. Sikorsky *et al*. (1990)[10] and other reviewers[11,12] proposed a consensus with a high frequency of sequential appearance of hydrophobic amino acids at positions：1-4(W/L/F)-7(L/I/M)-8(G/A/S)-11(Y/L/F)-20(A/S/E) -24(F/Y/L)-27(A/S/L)-32(P/K/E)-34；residues in parentheses are alternatives for a single position numbered from N-terminus, no positions are completed conserved.

A single TPR adopts a helix-turn-helix structure with the two α-helices packed together by inter-helix interactions. Typically, TPRs are tamdemly arrayed in proteins, and a cluster of three consecutive TPRs is the predominant arrangement[13]. The first solved TPR protein structure of protein phosphatase 5[14] demonstrated that multiple TPRs form a right-handed superhelix with a groove of large surface area available for ligand banding.

2 RESULTS

2.1 Identification of TPRs of Different Lengths

ELTR TPRs could be easily recognized among known TPR containing proteins and offer an ideal pool of proteins to analyze. The question we want to address is whether the lengths of TPR varied during evolution. Toward this end, we used the program PATTINPROT on NPS@[15], it is particularly useful to identify recurring patterns in proteins. The best feature of this program is the pattern search could be designed with a very strict length requirement, a query template sequence was formulated to search for repeating patterns at a fixed length in protein data base, therefore, any disruption of ELTR will be excluded by this search strategy. Using this program, we identified ELTR TPRs with different lengths (an evolutionary event).

The original query template of TPR was formulated using the consensus sequence reported in literatures[11,12], repeat length is 34 amino acids, repeating times are 3, different search stringencies were tested and 75% identity to query template was chosen as reliable hit identification. The presence of genuine TPR repeats was further confirmed by going through the annotations of each identified protein and performing individual BLAST searches, all the TPR proteins identified by PATTINPROT are also annotated TPR proteins by others. The template specifications are as follow：

X(3)-[WLY] - X(2)-[LIM]-[GAS]-X(2)-[YLF]-X(8)-[ASE]-X(3)-[FYL]-X(2)-[ASL]-X(4)-[PKE]-X(5)-[WLY]-X(2)-[LIM]-[GAS]-X(2)-[YLF]-X(8)-[ASE]-X(3)-[FYL]-X(2)-[ASL]-X(4)-[PKE]-X(5)-[WLY]-X(2)-[LIM]-[GAS]-X(2)-[YLF]-X(8)-[ASE]-X(3)-[FYL]-X(2)-[ASL]-X(4)-[PKE]-X(5), in which X represents arbitrary amino acids and numbers in parenthesis represents lengths. The repeat length of the template was changed by varying the number of amino acids between residue [PKE] and residue [WLY], the region between PKE and WLY was chosen as a length variation target because this region is not predicted to disturb the overall secondary structure of TPR, hence 5 amino acids between the two positions represent the original 34 amino acid length, 6-16 amino acids between the two positions represent length increase from 35 to 45 amino acids. Shorter length TPR repeats (31-33 amino acids) were also tested by reducing the number of amino acids between [PKE] and [WLY]. The complete search strategy tested ELTR TPR of 31-45 amino acids lengths in current data base. The results are summarized in Table 1. As expected, the numbers of ELTR repeat containing proteins identified vary greatly with length：most TPR proteins (913) identified contain 34 amino acid repeat (TPR-34), and the total numbers of TPR-34 repeating patterns in 913 proteins are 2192, which is 2 folds higher than the number of proteins, indicating a genuine ELTR event. The numbers of proteins with a specific TPR length identified and the occurrence of many more patterns than proteins are 2 key indicators of significant and real ELTR events. Candidate

ELTR TPRs were detected across a wide length range, however, repeats of 34, 40 and 42 amino acids stood out as the most prevalent types, the corresponding repeats were designated TPR-34, TPR-40 and TPR-42. However, ELTRs of other lengths (although not prevalent) could still be real. Some of these repeats were found to be results of degenerate TPR-34, and some, such as the TPR-45 proteins were only found among fungal genomes. It should be pointed out that although the numbers of TPR-40 and TPR-42 proteins identified are much fewer than that of TPR-34 (64 and 136 proteins versus 913 respectively), the results should be taken with a grain of salt. One should bear in mind that the longer length TPR (acquired novel functions as discussed later in the text) may obey a sequence conservation rule that differs from the classical TPR consensus derived from TPR-34. Hence the pattern searches conducted here will miss a large portion of TPR-40 and TPR-42 proteins. Many TPR proteins were probably excluded in an ELTR search due to the stringent pattern constraint. To thoroughly analyze these proteins and firmly assign them to a specific length group is of great value to functional study.

Table 1 Distribution of ELTR TPR-containing proteins across a 31-45 amino acids repeat length range generated using PATTINPROT

Search Names	Randomized probability	Similarity/%	Database	Proteins identified	Patterns identified	Proteins in database (12 Dec 2005)
3tpr31	1.911e-18	75	Nr_Prot	0	0	2238571
3tpr32	1.906e-18	75	Nr_Prot	15	15	2238571
3tpr33	1.902e-18	75	Nr_Prot	14	15	2238571
3tpr34	1.897e-18	75	Nr_Prot	913	2192	2238571
3tpr35	1.893e-18	75	Nr_Prot	19	23	2238571
3tpr36	1.889e-18	75	Nr_Prot	2	2	2238571
3tpr37	1.884e-18	75	Nr_Prot	56	58	2238571
3tpr38	1.880e-18	75	Nr_Prot	9	11	2238571
3tpr39	1.875e-18	75	Nr_Prot	13	13	2238571
3tpr40	1.871e-18	75	Nr_Prot	64	129	2238571
3tpr41	1.866e-18	75	Nr_Prot	28	37	2238571
3tpr42	1.862e-18	75	Nr_Prot	136	339	2238571
3tpr43	1.858e-18	75	Nr_Prot	3	3	2238571
3tpr44	1.853e-18	75	Nr_Prot	2	2	2238571
3tpr45	1.849e-18	75	Nr_Prot	13	49	2238571

2.2 Functions of TPR-40 Proteins

TPR-40s were identified in both prokaryotic and eukaryotic proteins. TPR-40 proteins were further examined by sequence comparison and repeat alignments, to verify the repeat length. The functions of many TPR-40 proteins (particularly those of eukaryotic origins) have been characterized. TPR-40s were identified in human LGN (Fig. 1A) and rat AGS3 proteins. LGN was first identified as a $G_{\alpha i}$-binding protein in a yeast two-hybrid screen[16], whereas AGS3 was discovered in a functional screen for receptor-independent activators of G-proteins. 17 Both LGN and AGS3 were predicted to share a common architecture: an N-terminal domain contains 7 tandem-arrayed TPRs followed by a linker region that connects the TPR domain to a series of 4 G-protein regulatory (GPR) motifs[16-18]. The TPR domains in LGN and AGS3 were shown to interact with unidentified protein partners and the interaction directs these proteins to specific intracellular locations[19-21]. A homologue of LGN/AGS3 in *Drosophila melanogaster*, Pins (partner of Inscuteable), is a key component of a protein complex involved in regulation of asymmetric cell division[22,23]. Pins binds the protein Inscuteable via its TPR domain and G-protein via the C-terminal GPR motifs. These proteins form a complex with a number of other proteins to regulate asymmetric cell division in *Drosophila*[22,23]. Another type of TPR-40 protein is rapsyn, a peripheral membrane protein of the skeletal muscle synapse that is essential for the formation of highly organized structure of the vertebrate neuromuscular junction. It colocalizes with

nicotinic acetylcholine receptors (nAChR) in the postsynaptic membrane from the earliest stages of innervation[24]. In rapsyn(-/-) mutant mice, which die at early birth, nAChRs are not clustered[25]. For review of neuromuscular synapses formation and rapsyn function, see Burden (1998)[26]. The structure of rapsyn is not known, but its primary structure was predicted to contain 8[27] or 7[28] TPR repeats. The TPRs in rapsyn (Fig. 1B) is required for self-association of rapsyns[28], a phenomena not found with TPR-34, the structural bases for rapsyn self-association are of keen interest. Mutation in TPR domain of rapsyn caused severe use-dependent muscle fatigue in Zebrafish[29].

The TPR-40s in prokaryotic TPR proteins share high degree of similarity between adjacent repeats and the 40-amino acid length of the repeats was easily ascertained. For example, the repeats in a hypothetical protein (ZP_00106722) from the cyanobacterium *Nostoc punctiforme*,

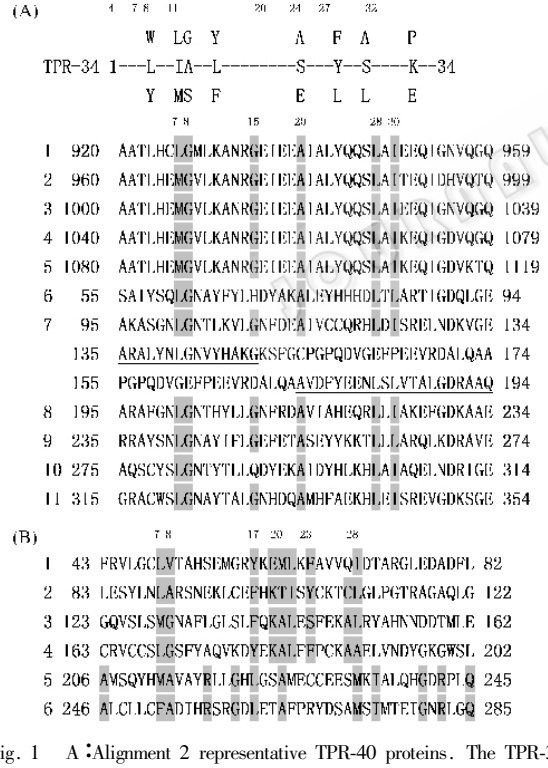

Fig. 1 A: Alignment 2 representative TPR-40 proteins. The TPR-34 consensus is given above the alignment and the conserved positions are numbered on top. Repeats: 1-5 were selected from a hypothetical protein (ZP_00106722) of *Nostoc punctiforme*; 6-11 were selected from human LGN (P81274). A disrupted TPR-40 between residues 135-194 was included in the alignment. The regions of the disrupted repeat related to TPR were underlined. The conserved positions are numbered on top of the alignment. B: Alignment of TPR-40s in mouse rapsyn (NP_033049). Repeats: 1-4 are similar in sequence; 5 and 6 are similar. Conserved positions are numbered on top.

are tandemly arrayed and share greater than 90% amino acid identity between repeats (Fig. 1A). This extraordinary degree of identity implies these TPRs are recently duplicated in this organism. TPR-40 proteins from prokaryotes with preliminary functional characterization, include AfsR from *Streptomyces*[30,31] and Rap proteins[32] from *Bacillus*. These latter TPRs share much lower sequence identities between adjacent repeats, but the 40 amino acids length could be ascertained after careful alignments, suggesting that these TPRs are much older than those just duplicated in cyanobacteria. The recently characterized prokaryotic TPR protein, NlpI[33] from *E. coli*, is a TPR-34 protein.

2.3 Functions of TPR-42 Proteins

Similarly, TPR-42s were identified in proteins of both prokaryotic and eukaryotic origins. In eukaryotes, TPR-42 was mainly identified in kinesin light chain (KLC, Gindhart and Goldstein, 1996[34]), a key determinant of intracellular trafficking. It interacts with kinesin heavy chain (KHC) to form the native heterotetrameric microtubule motor-kinesin[35,36]. Proposed functions of KLC include coupling of cargo to KHC and/or modulation of KHC ATPase activity[36]. The binding partners of KLC include calmodulin[37], Sunday driver[38] and amyloid precursor protein[39]; suggesting that KLCs regulate motor functions by integrating information from diverse signaling pathways. Kinesin-dependent axonal transport is mediated by interaction between KLC and a membrane-associated Sunday driver protein via KLC TPR domain. KLC is also shown to bind amyloid precursor protein through its TPR domain, again implicating KLC in axonal transport of APP, which plays a major role in the development of Alzheimer's disease. The structure of KLC is not known, the diverse binding capacities of this protein predict an interesting structure-function relationship to be discovered. The TPR-42s from prokaryotes share a much higher degree of sequence identities between adjacent repeats (data not shown), and repeats of 42 amino acid length could be easily recognized. And they share apparent sequence similarity to KLC. Unfortunately, most of these proteins were annotated TPR containing hypothetical proteins, their functions are essentially unknown. Again some prokaryotic TPR-42 proteins such as those in streptomycetes are less conserved at sequence

level between repeats when compared to those from cyanobacteria, suggesting an older origin. Examples of TPR-42 repeats in a putative KLC[40] (AAB87735) from the cyanobacterium *Plectonema boryanum* and eukaryotic sources are shown in (Fig. 2).

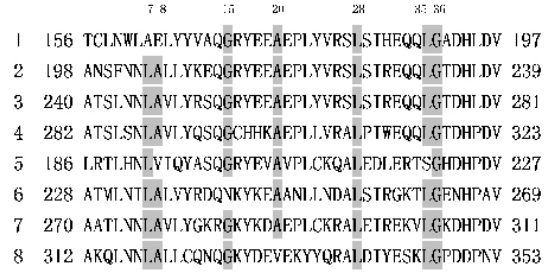

Fig. 2 Alignment of TPR-42s. Repeats: 1-4 are selected from a putative KLC (AAB87735) of cyanobacterium *P. boryanum*; 5-8 are selected from KLC (P46824) of *Drosophila melanogaster*. The conserved positions are numbered on top of the alignment.

2.4 Structure and Function of TPR Proteins

TPR derived structures mediate protein-proteins interactions.

The interaction surface is provided by a concave amphiphilic channel on the overall superhelix formed by multiple TPRs. Since the report of PP5 (a TPR-34 protein) structure, those of several other TPR proteins, alone or in complex with a peptide or protein ligand have been solved[41-48]. All these proteins are members of TPR-34 proteins or derivatives of TPR-34s. To date, a structure of a TPR-40 or TPR-42 proteins is still not known, their structures are expected to form a superhelix with a bigger concave surface to interact with bigger ligands. Sequence alignments of representative TPR-40 and TPR-42 proteins reveal consensuses that differ from that of TPR-34 and from each other. Since the consensus is predicted to vary between proteins of different functions, a general profile is not given in this paper. Nevertheless, conservations at key positions such as (G and A at position 8, and A at positions 20 and 27) suggest these repeats are related, and they share a general design principle. The functions of TPR-40 and TPR-42 proteins were studied in higher eukaryotes (animal kingdom), it is interesting to determine whether they are also present in lower eukaryotes, such as in yeast; and what functions they confer in lower eukaryotes. In prokaryotes, the TPR-40 proteins proved to be regulatory proteins, in the case of Rap proteins in *Bacillus*, it is known that their functions depend on their ability to interact with other proteins, but not known whether these interactions are mediated by the TPR domain. It is clear from studies of eukaryotic TPR-40 and TPR-42 proteins that these TPR domains mediate protein-protein interactions and their binding specificities are apparently different from those of TPR-34 proteins.

3 CONCLUSION REMARKS

In this article, length was proposed to be a defining parameter for repeat classification and evolution. Based on careful analyses, I showed that the length of TPR varied during evolution, and certain lengths (such as 34, 40 and 42 amino acids) are more prevalent than the others. The functions of TPR-34 proteins have been reviewed many times[11-13, 49], TPR-40 and TPR-42 proteins were considered TPR-34 proteins. Apparently, despite length divergence, TPR proteins still function by mediating protein-protein interactions, although the interaction partners and pathways have changed. Since the basic structural elements of TPR-34 are still retained in TPR-40 and TPR-42, proteins containing longer TPR should at least expand TPR-34's functional capacities. Why certain lengths were favored over the others? In his recent paper, Vermeij (2006)[50] suggests "the principle of physics and economics imply that many derived functional states are achieved many times in many clades because they impart substantial, widely applicable advantages to their bearers" and later "some configurations stabilize and self-organize more readily than others". The prevalence of TPR repeats on genomes of diverse organisms seems to support these predictions, some TPR repeats, particularly those in 40 and 42 length categories, clearly emerged very recently and independently in specific lineages (they share high degree of sequence identity between adjacent repeats and less disruption of ELTR pattern). Although it is not clear what "substantial or widely applicable advantages" they impart to the bearer, a plausible speculation is the expansion of TPR proteins laid the foundation for functional sophistication by providing more platforms to perform multiple task functions. They provide the hubs for protein-protein interactions[51].

TPR-34 proteins have the highest percentage of representation in database, the widest distribution among extant organisms, they represent the oldest form of TPRs. Both TPR-40 and TPR-42 proteins have lower absolute numbers in database, narrower distribution, both

indicators of later origin. A general impression is TPRs have more degenerate versions and functional variations in eukaryotes than in prokaryotes, does this suggests that TPR was first originated, selected and fixed in eukaryotes? In conclusion, repeat length variation exemplifies a simple function innovation strategy used by all 3 domains of life, although the strategy may have been used by the eukaryotes for a longer time. A broader inference is that length variation in any duplication event could be a fundamental genetic variation mechanism employed by nature for function and genome evolution.

ACKNOWLEDGEMENTS: This work is supported by the knowledge innovation fund from Chinese Academy of Sciences.

REFERENCES

[1] Andrade MA, Perez-Iratxeta C, Ponting CP. Protein repeats: structures, functions, and evolution. *J Struct Biol*, 2001, **134**: 117 – 131.

[2] Andrade MA, Petosa C, O'Donoghue SI, *et al*. Comparison of ARM and HEAT protein repeats. *J Mol Biol*, 2001, **309**: 1 – 18.

[3] Marcotte EM, Pellegrini M, Yeates TO, Eisenberg D. A census of protein repeats. *J Mol Biol*, 1999, **293**: 151 – 160.

[4] Andrade MA, Ponting CP, Gibson TJ, *et al*. Homology-based method for identification of protein repeats using statistical significance estimates. *J Mol Biol*, 2000, **298**: 521 – 537.

[5] Heger A, Holm L. Rapid automatic detection and alignment of repeats in protein sequences. *Proteins*, 2000, **41**: 224 – 237.

[6] Pellegrini M, Marcotte EM, Yeates TO. A fast algorithm for genome-wide analysis of proteins with repeated structures. *Proteins*, 1999, **35**: 440 – 446.

[7] Ponting CP. Proteins of the endoplasmic reticulum associated degradation pathway: domain detection and function prediction. *Biochem J*, 2000, **351**: 527 – 535.

[8] Ponting CP, Mott R, Bork P, *et al*. Novel protein domains and repeats in *Drosophila melanogaster*: insights into structure, function, and evolution. *Genome Res*, 2001, **11**: 1996 – 2008.

[9] Hirano T, Kinoshita N, Morikawa K, *et al*. Snap helix with knob and hole: essential repeats in *Schizosaccharomyces pombe* nuclear protein *nuc*2^+. Cell, 1990, **60**: 319 – 328.

[10] Sikorski RS, Boguski MS, Goebl M, *et al*. A repeating amino acid motif in CDC23 defines a family of proteins and new relationship among genes required for mitosis and RNA synthesis. *Cell*, 1990, **60**: 307 – 317.

[11] Lamb JR, Tugendreich S, Hieter P. Tetratrico peptide repeat interactions: to TPR or not to TPR? *Trends Biochem Sci*, 1995, **20**: 257 – 259.

[12] Blatch GL, Lassle M. The tetratricopeptide repeat: a structural motif mediating protein-protein interactions. *BioEssays*, 1999, **21**: 932 – 939.

[13] D'Andrea LD, Regan L. TPR proteins: the versatile helix. *Trends Biochem Sci*, 2003, **28**: 655 – 662.

[14] Das AK, Cohen PW, Barford D. The structure of the tetratricopeptide repeats of protein phosphatase 5: implications for TPR-mediated protein-protein interactions. *EMBO J*, 1998, **17**: 1192 – 1199.

[15] Combet C, Blanchet C, Geourjon C, *et al*. NPS@: network protein sequence analysis. *Trends Biochem Sci*, 2000, **25**: 147 – 150.

[16] Mochizuki N, Cho G, Wen B, *et al*. Identification and cDNA cloning of a novel human mosaic protein, LGN, based on interaction with $G_{\alpha i2}$. *Gene*, 1996, **181**: 39 – 43.

[17] Takesono A, Cismowski MJ, Ribas C, *et al*. Receptor-independent activators of heterotrimeric G-protein signaling pathways. *J Biol Chem*, 1999, **274**: 33202 – 33205.

[18] Bernard ML, Peterson YK, Chung P, *et al*. Selective interaction of AGS3 with G-proteins and the influence of AGS3 on the activation state of G-proteins. *J Biol Chem*, 2001, **276**: 1585 – 1593.

[19] Peterson YK, Bernard ML, Ma H, *et al*. Stabilization of the GDP-bound conformation of $G_{i\alpha}$ by a peptide derived from the G-protein regulatory motif of AGS3. *J Biol Chem*, 2000, **275**: 33193 – 33196.

[20] Pizzinat N, Takesono A, Lanier SM. Identification of a truncated form of the G-protein regulator AGS3 in heart that lacks the tetratricopeptide repeat domains. *J Biol Chem*, 2001, **276**: 16601 – 16610.

[21] Blumer JB, Chandler LJ, Lanier SM. Expression analysis and subcellular distribution of the two G-protein regulators AGS3 and LGN indicate distinct functionality. *J Biol Chem*, 2002, **277**: 15897 – 15903.

[22] Schaefer M, Shevchenko A, Shevchenko A, *et al*. A protein complex containing Inscuteable and the G_{α}-binding protein Pins orients asymmetric cell divisions in *Drosophila*. *Curr Biol*, 2000, **10**: 353 – 362.

[23] Yu F, Morin X, Cai Y, *et al*. Analysis of partner of inscuteable, a novel player of *Drosophila* asymmetric divisions, reveals two distinct steps in inscuteable apical localization. *Cell*, 2000, **100**: 399 – 409.

[24] Noakes PG, Phillips WD, Hanley TA, *et al*. 43K protein and acetylcholine receptors colocalize during the initial stages of neuromuscular synapse formation in *vivo*. *Dev Biol*, 1993, **155**: 275 – 280.

[25] Gautam M, Noakes PG, Mudd J, *et al*. Failure of postsynaptic specialization to develop at neuromuscular junctions of rapsyn-deficient mice. *Nature*, 1995, **377**: 232 – 236.

[26] Burden SJ. The formation of neuromuscular synapses. *Genes Dev*, 1998, **12**: 133 – 148.

[27] Ponting CP, Phillips C. Rapsyn's knobs and holes: eight tetratrico peptide repeats. *Biochem J*, 1996, **314**: 1053 – 1054.

[28] Ramarao MK, Bianchetta MJ, Lanken J, *et al*. Role of rapsyn tetratricopeptide repeat and coiled-coil domains in self-association and nicotinic acetylcholine receptor clustering. *J Biol Chem*, 2001, **276**: 7475 – 7483.

[29] Ono F, Shcherbatko A, Higashijima S, *et al*. The Zebrafish motility mutant twitch once reveals new roles for rapsyn in synaptic function. *J Neurosci*, 2002, **22**: 6491 – 6498.

[30] Horinouchi S, Kito M, Nishiyama M, *et al*. Primary structure of AfsR, a global regulatory protein for secondary metabolite formation in *Streptomyces coelicolor* A3(2). *Gene*, 1990, **95**: 49 –

[31] Horinouchi S, Suzuki H, Beppu T. Nucleotide sequence of *afsB*, a pleiotropic gene involved in secondary metabolism in *Streptomyces coelicolor* A3(2) and "*Streptomyces lividans*". *J Bacteriol*, 1986, **168**: 257 – 269.

[32] Core L, Perego M. TPR – mediated interaction of RapC with ComA inhibits response regulator-DNA binding for competence development in *Bacillus subtilis*. *Mol Microbiol*, 2003, **49**: 1509 – 1522.

[33] Wilson CG, Kajander T, Regan L. The crystal structure of NlpI. A prokaryotic tetratricopeptide repeat protein with a globular fold. *FEBS J*, 2005, **272**: 166 – 179.

[34] Gindhart JG, Goldstein LS. Tetratrico peptide repeats are present in the kinesin light chain. *Trends Biochem Sci*, 1996, **21**: 52 – 53.

[35] Cabeza-Arvelaiz Y, Shih LC, Hardman N, *et al*. Cloning and genetic characterization of the human kinesin light-chain (KLC) gene. *DNA Cell Biol*, 1993, **12**: 881 – 892.

[36] Gauger AK, Goldstein LS. The *Drosophila* kinesin light chian. Primary structure and interaction with kinesin heavy chain. *J Biol Chem*, 1993, **268**: 13657 – 13666.

[37] Matthies HJ, Miller RJ, Palfrey HC. Calmodulin binding to and cAMP-dependent phosphorylation of kinesin light chains modulate kinesin ATPase activity. *J Biol Chem*, 1993, **268**: 11176 – 11187.

[38] Bowman AB, Kamal A, Ritchings BW, *et al*. Kinesin-dependent axonal transport is mediated by the Sunday driver (SYD) protein. *Cell*, 2000, **103**: 583 – 594.

[39] Kamal A, Stokin GB, Yang Z, *et al*. Axonal transport of amyloid precursor protein is mediated by direct binding to the kinesin light chain subunit of kinesin – I. *Neuron*, 2000, **28**: 449 – 459.

[40] Celerin M, Gilpin AA, Dossantos G, *et al*. Kinesin light chain in a eubacterium. *DNA Cell Biol*, 1997, **16**: 787 – 795.

[41] Abe Y, Shodai T, Muto T, *et al*. Structural basis of presequence recognition by the mitochondrial protein import receptor Tom20. *Cell*, 2000, **100**: 551 – 560.

[42] Gatto GJJr, Geisbrecht, BV, Gould SJ, *et al*. Peroxisomal targeting signal-1 recognition by the TPR domains of human PEX5. Nat Struct Biol 2000; 7: 1091 – 1095. Erratum in: Nat Struct Biol, 2002, 9: 788.

[43] Grizot S, Fieschi F, Dagher MC, *et al*. The active N-terminal region of $p67^{phox}$. Structure at 1.8 ? resolution and biochemical characterizations of the A128V mutant implicated in chronic granulomatous disease. *J Biol Chem*, 2001, **276**: 21627 – 21631.

[44] Kumar A, Roach C, Hirsh IS, *et al*. An unexpected extended conformation for the third TPR motif of the peroxin PEX5 from *Trypanosoma brucei*. *J Mol Biol*, 2001, **307**: 271 – 282.

[45] Lapouge K, Smith SJ, Walker PA, *et al*. Structure of the TPR domain of $p67^{phox}$ in complex with Rac GTP. *Mol Cell*, 2000, **6**: 899 – 907.

[46] Scheufler C, Brinker A, Bourenkov G, *et al*. Structure of TPR domain-peptide complexes: critical elements in the assembly of the Hsp70-Hsp90 multichaperone machine. *Cell*, 2000, **101**: 199 – 210.

[47] Sinars CR, Cheung-Flynn J, Rimerman RA, *et al*. Structure of the large FK506-binding protein FKBP51, an Hsp90-binding protein and a component of steroid receptor complexes. *Proc Natl Acad Sci USA*, 2003, **100**: 868 – 873.

[48] Taylor P, Dornan J, Carrello A, *et al*. Two structures of cyclophilin 40: folding and fidelity in the TPR domains. *Structure (Camb)*, 2001, **9**: 431 – 438.

[49] Goebl M, Yanagida M. The TPR snap helix: a novel protein repeat motif from mitosis to transcription. *Trends Biochem Sci*, 1991, **16**: 173 – 177.

[50] Vermeij GJ. Historical contingency and the purported uniqueness of evolutionary innovations. *Proc Natl Acad Sci USA*, 2006, **103**: 1804 – 1809.

[51] Ekman D, Light S, Bjorklund AK, *et al*. What properties characterize the hub proteins of the protein-protein interaction network of *Saccharomyces cerevisiae*? *Genome Biol*, 2006, 7: R45.

TPR 重复和 ELTR 排型:重复的长度变化是一种功能进化的机制

杨克迁

(中国科学院微生物研究所 国家资源前期开发重点实验室 北京 100101)

摘 要:TPR 重复最初定义为一个 34 个氨基酸的蛋白结构重复。本文用 PATTINPROT 软件对基因库中 TPR 重复长度多样性进行了分析,发现 40 和 42 个氨基酸 TPR 重复大量存在。含 40 和 42 个氨基酸 TPR 重复序列蛋白的功能分析说明,这些长的 TPR 重复可以赋予蛋白新的功能。根据以上分析,提出重复序列的长度变化可能是功能进化的一种发生机制。

关键词:TPR;重复长度;串联重复;进化;排型

中图分类号:Q939,Q78 **文献标识码**:A **文章编号**:0001-6209(2007)06-0962-07

作者简介 杨克迁(1964 –),男,青岛人,研究员,博士,主要从事微生物生理研究。Tel/Fax:86-10-64807459;E-mail:yangkq@im.ac.cn

收稿日期:2007-05-08 **接受日期**:2007-08-15 **修回日期**:2007-09-16