Research Journal of Biological Sciences 1 (1-4): 12-15, 2006

CTLH: A Novel Domain with A Typical "U" Shape Architecture

[1]Weiqi Zeng, [1]Yuequn Wang, Wuzhou Yuan, Yun Deng, Yongqing Li,
Chuanbing Zhu, Mingyao Liu and Xiushan Wu
The Center For Heart Development, College of Life Sciences, Hunan Normal University,
Changsha, 410081, Hunan, Peoples' Republic of China

Abstract: We described here a novel domain, CTLH, which can be found in proteins of most already-researched species currently. The domain, consisted of 58 amino acid residues, is not often-conserved in sequence (but has 5 regularly-arranged leucines that are highly conserved). However, its 3D architecture conserves greatly with a typical ´U´ shape (alpha helix+beta turn+alpha helix). Although the function of CTLH domain remains unknown, related studies aiming at this domain suggest it participates in the regulation of certain signaling pathways like MAPK, and is involved in some living activities of cells such as protein degradation and mitosis.

Key words: CTLH domain, 3D architecture, signaling pathways, living activities

INTRODUCTION

Gβ-like proteins are a big family of proteins that included several WD40 repeat domains. In our recent study, some of Gβ-like proteins have been identified to participate in cell signaling pathways. WDR26 protein is a 58 kDa, made of 514 amino acids. Overexpression of WDR26 protein suppresses transcriptional activities of SER and Elk1 in Cos-7 cells[1]. Besides WDR26, there are more Gβ-like proteins that act as regulatory factors in signaling pathway , such as hPIP1 and its homologue in yeast skb15[2,3]. Almost Gβ-like proteins share a WD region, containing various numbers of WD40 repeats, which is different from the 7 WD40 repeats of Gβ subunits. Some studies reveal WD40 repeats region is important for Gβ-like proteins, but is insufficient for their functions, suggesting other conserve domains are crucial in most Gβ-like proteins.

In a study for information to understand the molecular mechanism by which WDR26 affects MAPK signaling pathway, we further investigated the domain architecture of this protein and identified a novel domain, which is named C-terminal to LisH (Lissencephaly type-1-like homology) (CTLH). We found that CTLH domain is closely associated with WD40 region in lots of CTLH-containing proteins.

Characterization of the CTLH domain: The WDR26 protein (genebank identifier (gi) 55743152) contains 5 WD40 repeat domains (residues 204-494). Using Hydrophobic Cluster Analysis[4], we identified a distinct globular domain at the N-terminal region (residues 9-84). This domain is made of predicted alpha-helical sequence, and was disturbed by a Random coil (residues 40-47). As a result, this domain was curved into a ´U´ shape. On both sides of the CTLH domain£¬a couple of cysteines (residues 6 and 110) form a disulfide bridge that is helpful to maintain the ´U´ shape (Fig. 1).

PSI-BLAST search[5] of the non-redundant database at the National Center for Biological Information using this domain as query led to the identification of statistically significant similarities with other proteins. Using CTLH domain as query, we searched ExPASy Proteomics Server (Expert Protein Analysis System, http://www.expasy.org/)[6] and obtained 170 proteins containing the CTLH domain. These proteins can be sorted to 11 groups by their architectures (Fig. 3). In most CTLH domain-containing proteins (141%170), this domain was located on LISH (Lissencephaly type-1-like homology) domain´s C-terminus, and was named CTLH (C-terminal to LisH) domain. Besides LISH domain, WD40 repeat domains are common structures at CTLH protein´s C-terminus (49%170). Other domain, such as Zinc finger RING-type domain, exists infrequently at CTLH protein´s C-terminus (3%170).

Structure model of CTLH: After comparing most CTLH domain-containing proteins, we concluded a homologue sequence of CTLH domain according to the occurring frequency of each amino acid on its corresponding places (´TYKRYQLIHDSILQQELKEVLSWCSEHRAILKKNNST LELEVRLQRFIELIKSKKLCQ). This domain abounds with leucine and isoleucine, both of which take up to approximately 1/3 of the whole CTLH domain. 5 out of the

Corresponding Author: Xiushan Wu, The Center for Heart Development College of Life Sciences Hunan Normal University Changsha, Hunan, 410081 P. R. China

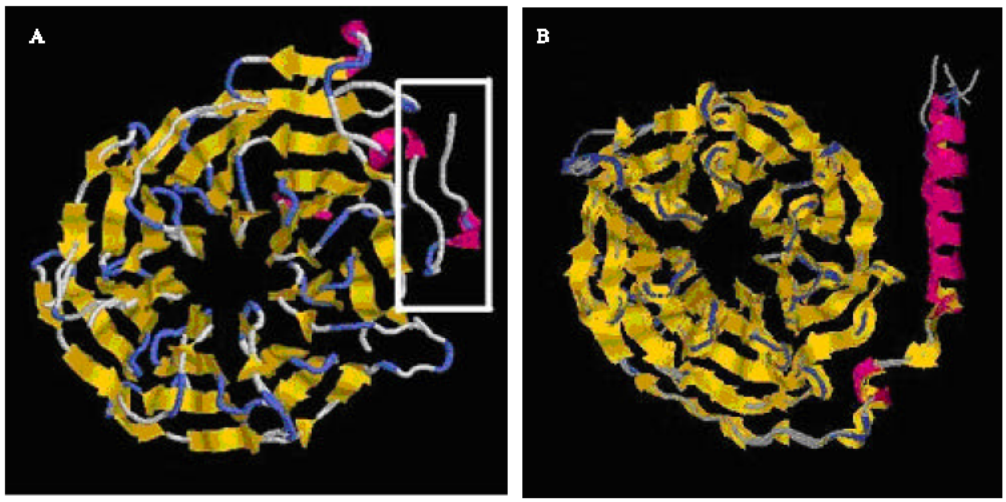

Fig. 1: 3D model of the WDR26 (A) and G-beta subunit (B). White rectangle marks the CTLH domain

total 10 leucines are conserved over an extent of 80% (residues 12, 21, 31, 42, 50), arranging in some order (the 5 leacines arranged at intervals of 8-11 residues) (Fig. 2). This illustrates the possibility that these leucines must play some crucial role in CTLH domain. Compared with leucine, the conservation of other amino acids are not obvious (generally<65%). However, in light of spacial architecture, the variation of different amino acids has not evidently affected the 3D structure of the domain. Among all the discovered CTLH-containing proteins, this domain has always existed in an extremely typical ´U´ shape (alpha helix + beta turn + alpha helix), indicating this special structure is indispensable for the domain to function properly[7].

We stuidy the protein structure database by using the protein sequence of WDR26 (http://www.expasy.org) and found several similarly structured protein 3D files (1erjA, 1erjB, 1erjC, 1tbgC, 1gg2B, 1tbgB, 1tbgD and 1omwB). Based upon these documents, with the protein structure simulating software (http://swissmodel.expasy.org//SWISS-MODEL.html), we obtained a theoretic 3D structure model and a Gβ-like protein model likewise[8,9]. Meanwhile, we found that both of the models have similar turbo-like configurations in the WD40 region, yet differs distinctly at the C-terminus (in CTLH domain) with a clear U shape structure.

Functional prediction for CTLH domain: CTLH domain is a fairly archaic domain by origin and the CTLH-containing protein has been founded in many different species from yeast to human. However, the CTLH domain seldom appears separately in a certain protein. Only 1 out of 11 groups of CTLH-containing proteins exists separately (Fig 3). Among the concomitant domains of CTLH, LisH and WD40 repeat domains are the most frequent domain structures found in the same protein. Otherwise, high-conserved leucine-rich repeats, which bind the substrate for ubiquitination, are crucial for CTLH functions[10].

The CTLH domain-containing proteins detected in yeast and animals has been demonstrated to participate in cellular activities, such as regulation of signaling pathway, protein degradation, cell proliferation and cell apoptosis[11-13]. Take yeast for instance: two of the novel CTLH-containing proteins (Gid1 and Gid2) are involved in ubiquitin-proteasome–dependent degradation and Gid2 is necessary for FBPase ubiquitination[14,15]. Another CTLH protein *Smu1* is a component of the mitogen-activated protein kinase (MAPK) pathways. *Smu1* of the *C. elegans* interacts with one or more additional factors to regulate the alternative splicing of unc-52 and other transcripts. Mutations in the Smu-1 gene of *C. elegans* were shown to suppress mutations in the genes Mec-8 and Unc-52. Mec-8 encodes a putative RNA binding protein that affects the accumulation of specific alternatively spliced mRNA isoforms produced by unc-52 and other genes[16]. In *Saccharomyces cerevisiae*, disruption of the gene *Smu1* resulted in a delayed mating response in a mating-type-specific manner and also in a severe reduction in disease production on maize[17,18]. In our recent studies, WDR26, a typical gene of another subfamily of CTLH proteins, is demonstrated to regulate the MAPK signaling pathway by repressing two down-stream transcriptors, ELK1 and SRE[1].

In summary, we have identified a novel protein domain, CTLH, that associated with the LisH and the WD40 repeat domains in the N-terminus of a large family of proteins. This domain forms a “U” shape structure,

Fig. 2: Multiple sequences alignment of representative CTLH domain. The alignment is constructed on the basis of the PSI-BLAST results by ClustalW method (http://searchlauncher.bcm.tmc.edu/). Protein identifiers (Accession number) are indicated on the left of the bar while consensus sequence is showed at the bottom. Color of the background shows consensus amino acids. The name of species is arranged as follows: CAH98507 *Plasmodium berghei,* CAH86738 *Plasmodium chabaudi,* EAA60138 *Aspergillus nidulans*, XP_331392 *Neurospora crassa*, XP_475629 *japonica cultivar*, CAB82702 *Arabidopsis thaliana*, AAH80474 *Xenopus tropicalis*, XP_423702 *Gallus gallus*, NP_955843 *Danio rerio*, CAF96979 *Tetraodon nigroviridis*, AAH06470 *Homo sapiens*, AAF72195 *Mus musculus*, XP_313610 *Anopheles gambiae*, XP_220393 *Rattus norvegicus*, XP_394073 *Apis mellifera*, AAO51804 *Dictyostelium discoideum*, CTLH domain, CAG80181 *Yarrowia lipolytica*, CAG59405 *Candida glabrata*

Fig. 3: Schematic representation of the domain architecture of representative CTLH proteins. The length scale is given in the right corner, using amino acid residue as unit. Abbreviations: LISH Lissencephaly type-1-like homology domain; CTLH c-terminal to LisH domain; WD_R~ Trp-Asp (WD) repeats domain; ZF_R~ Zinc finger RING-type domain

whose conservation is comparatively low in sequence yet high in architecture. The functional role of this novel domain is not clear although it may play a role in most of the Gβ-like proteins.

ACKNOWLEDGEMENTS

We are grateful to all members of the Center for Heart Development, College of Life Sciences in Hunan

Normal University for their excellent technical assistance and encouragement. This study was supported in part by the National Natural Science Foundation of China (No. 90508004, 30470867, 30270722, 30570934, 30571048, 30570265), PCSIRT of Education Ministry of China (IRT0445), National Basic Research Program of China £¨2005CB522505£©, and the Foundation of Hunan Province (No. 04FJ2006).

REFERENCES

1. Zhu, Y., *et al.*, 2004. WDR26: A novel Gbeta-like protein, suppresses MAPK signaling pathway. J. Cell Biochem., 15: 579-587.
2. Xia, C., *et al.*, 2001. Phosphorylation and regulation of G-protein-activated phospholipase C-beta 3 by cGMP-dependent protein kinases. J. Biol. Chem., 276: 19770-19777.
3. Xia, C., *et al.*, 2001. Regulation of the p21-activated kinase (PAK) by a human Gbeta-like WD-repeat Protein, hPIP1. Proc. Natl. Acad Sci. U.S.A., 22: 6174-6179.
4. Callebaut, I., *et al.*, 1997. Deciphering protein sequence information through hydrophobic cluster analysis (HCA): Current status and perspectives. Cell Mol. Life Sci., 53: 621-645.
5. Altschul, F., *et al.*, 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Res., 25: 3389-3402.
6. Gasteiger, E., *et al.*, 2003. ExPASy: The proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Res., 31: 3784-3788.
7. Cuff, A., *et al.*, 1998. JPred: A consensus secondary structure prediction server. Bioinformatics, 14: 892-893.
8. Peitsch, C., 1995. Protein modelling by E-Mail. Bio. Tech., 13: 658-660.
9. Peitsch, C., 1996. ProMod and Swiss-Model: Internet-based tools for automated comparative protein modelling. Biochem. Soc. Trans., 24: 274-279.
10. Zheng, N., *et al.*, 2002. Structure of the Cul1-Rbx1-Skp1-F boxSkp2 SCF ubiquitin ligase complex. Nature 18: 703-709.
11. Jones, J., R. Waterston and M.A. Marra, 2001. Changes in gene expression associated with developmental arrest and longevity in Caenorhabditis elegans. Genome Res., 11: 1346-1352.
12. Yang, N., *et al.*, 2003. Sequence analysis of a 282-kilobase region surrounding the citrus Tristeza virus resistance gene (Ctv) locus in Poncirus trifoliata L. Raf. Plant Physiol., 131: 482-492.
13. Smith, G., *et al.*, 2004. An ste20 homologue in Ustilago maydis plays a role in mating and pathogenicity. Eukaryot Cell, 3: 180-189.
14. Regelmann, J., *et al.*, 2003. Catabolite degradation of fructose-1, 6-bisphosphatase in the yeast Saccharomyces cerevisiae: A genome-wide screen identifies eight novel GID genes and indicates the existence of two degradation pathways. Mol. Biol. Cell, 14: 1652-1663.
15. Hammerle, M., *et al.*, 1998. Proteins of newly isolated mutants and the amino-terminal proline are essential for ubiquitin-proteasome-catalyzed catabolite degradation of fructose-1,6-bisphosphatase of Saccharomyces cerevisiae. J. Biol. Chem., 273: 25000-25005.
16. Spike, A., *et al.*, 2001. Analysis of smu-1, a gene that regulates the alternative splicing of unc-52 pre-mRNA in Caenorhabditis elegans. Mol. Cell Biol., 21: 4985-4995.
17. Schule, T., *et al.*, 2000. Ubc8p functions in catabolite degradation of fructose-1, 6-bisphosphatase in yeast. EMBO J., 19: 2161-2167.
18. David, M., *et al.*, 2004. Identification of Factors Regulating Poly(A) Tail Synthesis and Maturation. Mol. Cell Biol., 24: 4196-4206.