African Journal of Biotechnology Vol. 10(77), pp. 17700-1775, 5 December, 2011
Available online at http://www.academicjournals.org/AJB
DOI: 10.5897/AJB11.2730
ISSN 1684–5315 © 2011 Academic Journals

Full Length Research Paper

cDNA, genomic sequence cloning and analysis of the ribosomal protein L37A gene (RPL37A) from the giant panda (*Ailuropoda melanoleuca*) and its over-expression

Chunlian Wu, Jinghui Sun and Wanru Hou*

Key Laboratory of Southwest China Wildlife Resources Conservation, School of Life Sciences, China West Normal University, 1# Shida Road, Nanchong, Sichuan Province 637009, China.

Accepted 15 November, 2011

Ribosomal protein L37A (RPL37A) is a component of 60S large ribosomal subunit encoded by the *RPL37A* gene, which belongs to the family of ribosomal L37AE proteins, located in the cytoplasm. The complementary deoxyribonucleic acid (cDNA) and the genomic sequence of *RPL37A* were cloned successfully from giant panda using reverse transcription- polymerase chain reaction (RT-PCR) and touch-down PCR technology, respectively. Both sequences were analyzed preliminarily. The cDNA of *RPL37A* was 298 bp in length, containing an open-reading frame (ORF) of 279 bp encoding 92 amino acids with an estimated protein molecular weight of 10.23 kDa, and a theoretical isoelectric point (pI) of 11.11. The length of genomic sequence was 2587 bp, possessing four exons and three introns. Alignment analysis indicated that the nucleotide sequence of the coding sequence showed a high homology with previously reported L37A sequences for *Homo sapiens, Pongo abelii*, *Mus musculus* and *Bos taurus*. The amino acid sequence encoded by the *RPL37A* gene of giant panda shared a high homology with the four animals above. Topology prediction showed that there were one N-glycosylation site, three protein kinase C phosphorylation site and three N-myristoylation sites in the RPL37A protein of giant panda. The *RPL37A* gene could be readily expressed in *Escherichia coli* because it was fused with the N-terminally His-tagged protein which gave rise to accumulation of an expected 14-kD polypeptide, in good agreement with the predicted molecular weight. The over-expression product obtained could be purified for studies of its function. The cDNA of *RPL37A* was cloned successfully for the first time from the giant panda in this study. The result provide scientific material to enrich and improve the *RPL37A* gene database.

Key word: Giant panda, ribosomal protein L37A (RPL37A), cDNA cloning, sequence analysis, over-expression

INTRODUCTION

Ribosomes, the organelles that catalyze protein synthesis, consist of a small 40S subunit and a large 60S subunit. The organelles are the components of cells that synthesize proteins from all amino acids (Andrei, 2011). Ribosomal proteins, which are the components of the ribosomes, exhibits various secondary functions in DNA repair, apoptosis, drug resistance and proliferation (Wang et al., 2006). With the continuous advancement of techniques, the researchers are gradually revealing the physiological function of ribosomal proteins which play a major role in human disease and its development (Yang and Liu, 2005). The ribosomal protein L37A (RPL37A), encoded by a *RPL37A* gene, is a component of the 60S subunit (Barnard et al., 1994). This protein belongs to the L37AE family of ribosomal proteins which are located in the cytoplasm (Susana and Manuel, 2010). The protein contains a C4-type zinc finger-like domain (Barnard et al.,

*Correspondence author. E-mail: hwr168@yahoo.com.cn. Tel/Fax: +86-0817-2568653.

1994). As it is typical for genes encoding ribosomal proteins, there are multiple processed pseudogenes of this gene dispersed through the genome (Schwartz et al., 2004; Stelzl et al., 2005; Ewing et al., 2007).

The giant panda *Ailuropoda melanoleuca* is the most dazzling star of the animal kingdom, and known as "National treasure of China". Giant panda belongs to national level of endangered animal, with a very high ecological, scientific, economic, cultural and aesthetic value. So far, most studies have focused on macrogeography, however, molecular biology research has become a central issues at present (Du et al., 2007; Hou et al., 2008, 2009a, b; Jennie et al., 1992; Qiao et al., 2009; Sun et al., 2011; Wu et al., 2010; Zhang et al., 2009).

This study was conducted using reverse transcription-polymerase chain reaction (RT-PCR) technique to amplify the complementary deoxyribonucleic acid (cDNA) of *RPL37A* gene from the total RNA and touchdown-PCR technique to amplify the genomic sequence of the *RPL37A* from total DNA from the skeleton muscle of giant panda, and then the sequence characteristics of the protein encoded by the cDNA were analyzed and compared with those of human and other mammalian species reported. We also over-expressed it in *Escerichia coli* using pET28a plasmids. To sum up, this study did not only enrich and supplement the information about *RPL37A* gene from the giant panda genomic library, but also provided the scientific basis for constructing the evolutionary tree. In addition, ribosomal proteins were used to treat cancer. Therefore, the study of mammalian ribosomal proteins has theoretic and applied importance.

MATERIALS AND METHODS

Skeletal muscle was collected from a dead giant panda at the Wolong Conservation Center, Sichuan, China. The collected skeletal muscle was frozen in liquid nitrogen and then used for deoxyribonucleic acid (DNA) and ribonucleic acid (RNA) isolation.

DNA and RNA isolation

The genomic DNA of the giant panda was isolated from muscle tissue (Li et al., 2010). The DNA obtained was dissolved in sterile water and kept at -20°C. Total RNAs were isolated from about 400 mg of muscle tissue using the total tissue/cell RNA extraction kits (Waton Inc.,Shanghai, China), according to the manufacturer's instructions. Total RNAs were dissolved in diethypyrocarbonate (DEPC) water, and kept at -70°C. DNA and RNA sample quality was checked using Experion (Bio-Rad, USA) and the quantification was performed spectrophotometrically.

Primers design, RT-PCR and cloning of cDNA sequence

The PCR primers were designed by Primer Premier 5.0, based on conserved sequences of the mRNA sequence of *RPL37A* from *Homo sapiens* (NM_000998), *Pongo abelii* (NM_001132041), *Mus musculus* (NM_009084) and *Bos Taurus* (NM_001035008). The specific primers of cDNA sequence are as follows: Pd- *RPL37A*-F: 5'-GCACTGGGTCTGCGGCGAC-3'; Pd- *RPL37A*-R: 5'-CTACTGGTCTTTCAACTCC-3'.

Total RNAs were synthesized into the first-stranded cDNAs using a reverse transcription kit with Oligo dT and the primers according to the manufacturer's instructions (Promega, USA) because only primers binding to Oligo (dT) can get full-length reverse transcriptase cDNA. The 20 µl of first-strand cDNA synthesis reaction system included 1 µg of total RNAs, 5 mM of $MgCl_2$, 1 mM of dNTPs, 0.5 µg of Oligo dT15, 10 U/µl of Rnase inhibitor, and 15 U of AMV reverse transcriptase. This reaction was incubated at 42°C for 60 min. The first-strand cDNA synthesized was used as a template. The total reaction volume for DNA amplification was 25 µl. Reaction mixtures contained 1.5 mM of magnesium chloride ($MgCl_2$), 200 µM each of dATP, dGTP, dCTP and dTTP (Promega, USA), 0.3 µM of each primer and 5.0 units of Taq plus DNA polymerase (Sangon Co., Shanghai, China). DNA amplification was performed using a MJ Research thermocycler, Model PTC-200 (Watertown, MA) with a program of 4 min at 94°C, followed by 30 cycles of 1 min at 94°C, 0.5 min at 45°C and 1.5 min at 72°C, and then ended with the final extension for 10 min at 72°C.

After amplification, PCR products were separated by electrophoresis in 1.5% agarose gel with 1×TAE [Tris-acetate-(Ethylene diamine tetra-acetic acid (EDTA))] buffer, stained with ethidium bromide (EB) and visualized under ultraviolet (UV) light. The expected fragments of PCR products were harvested and purified from gel using a DNA harvesting kit (Promega, USA), and then ligated into a pET28a vector at 22°C for 12 h. The recombinant molecules were transformed into *E. coli* competent cells (JM109), and then spread on the LB-plate containing 50 µg/ml ampicillin, 200 mg/ml isopropyl-beta-Dthiogalactopyranoside (IPTG), and 20 mg/ml X-gal (5-bromo-4-chloro-3-indolyl β-D-galactopyranoside). Plasmid DNA was isolated and digested by PstI and ScaII to verify the insert size. Plasmid DNA was sequenced by Huada Zhongsheng Scientific Corporation (Beijing, China).

Cloning the genomic sequence of *RPL37A*

The PCR primers were the same as the *RPL37A*-F and *RPL37A*-R presented above. The genomic sequence of the *RPL37A* gene was amplified using Touchdown-PCR with the following conditions: 94°C for 30 s, 62°C for 45 s and 72°C for 4 min in the first cycle and the annealing temperature decreased by 1°C per cycle; after 20 cycles, the conditions changed to 94°C for 30 s, 52°C for 45 s, 72°C for 4 min for another 15 cycles. The amplified fragment was also purified, ligated into the clone vector and tansformed into the *E. coli* cells. Finally, the recombinant fragment was sequenced by Sangon (Shanghai, China).

Construction of the expression vector and overexpression of recombinant RPL19

The PCR fragment corresponding to the RPL37A polypeptide was amplified from the *RPL37A* cDNA clone with the forward primer: 5'-GACGGATCC ATGGCTAAAC GCA -3' (BamH1) and the reverse primer: 5'-ATTCTCGAG TTACGCCAAGCTT-3'(XhoI), respectively. The PCR was performed at 94°C for 3 min; 30 cycles of 30 s at 94°C, 45 s at 53°C and 1 min at 72°C; 10 min at 72°C. The amplified PCR product was cut and ligated into corresponding site of the pET28a vector (Stratagen, USA). The resulting construct was transformed into *E. coli* BL21 (DE3) strain (Novagen, Germany) and used for the gene expression induction by adding IPTG at an OD600 of 0.6, and cultured further for 4 h at 37°C, using the empty vector transformed BL21(DE3) as a control. The recombinant protein samples were induced after 0, 0.5, 1.5, 2, 2.5, 3 and 3.5 h and then separated by sodium dodecyl sulphate- polyacrylamide gel electrophoresis (SDS-PAGE) and stained with Commassie blue

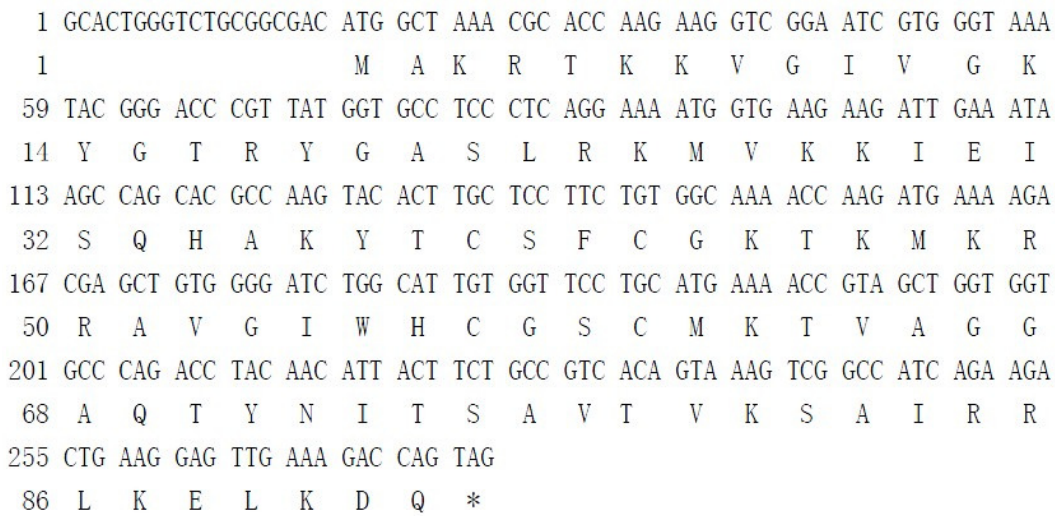

Figure 1. cDNA sequence of RPL37A gene and the deduced amino acid sequence from giant panda (*representing the stop codon).

R250 (Bio-rad, USA).

Data analysis

The sequence was analyzed by GenScan software. Homology research of the giant panda *RPL37A* compared with the gene sequences of other species were performed using Blast 2.1 (http://blast.ncbi.nlm.nih.gov/Blast.cgi). Open-reading frame (ORF) of the DNA sequence was searched using ORF finder software (http:/www.ncbi.nlm.nih.gov/gorf/orfig.cgi). Protein structure of the RPL37A sequence cloned was deduced using PredictProtein software (http:// cubic. Bioc. Columbia. edu/predictprotein/). Multiple sequence alignment was performed by DNAMAN 6.0 (Lynnon Biosoft, USA). The prediction of protein functional sites and biochemical characteristics were done using the software ExPASy Proteomics Server (http://swissmodel.expasy.org/).

RESULTS

Analysis of the *RPL37A* cDNA from giant panda

A cDNA fragment of about 300 bp was amplified from giant panda with primers *RPL37A*-F and *RPL37A*-R above. The length of the cloned cDNA was 298 bp. Blast research showed that the cloned cDNA sequence shared a high homology with the *RPL37A* gene from other four mammals reported, including *H. sapiens, P. abelii, M. musculus, B. taurus*. Based on the high identity, we concluded that the isolated cDNA is the cDNA encoding giant panda RPL37A protein. The RPL37A gene sequence has been submitted to Genbank (accession number: HQ318091). An ORF of 279 bp encoding 92 amino acids was found in the cDNA sequence (Figure 1). The initiation codon of *RPL37A* is ATG, and terminator codon is TAG. The average levels of the bases sequence is: A, 30.5%; C, 22.2%; G, 27.2% and T, 20.1%.

Analysis of the genomic sequence of *RPL37A* from giant panda

DNA fragment of about 2500 bp was amplified with primers *RPL37A*-F and *RPL37A*-R. The length of the cloned DNA fragment was 2587 bp. The comparison between the cDNA sequence and the genomic sequence of *RPL37A* amplified from giant panda was performed by software Lasergene (Wu et al., 2010). The result indicate that the cDNA sequence was in full accordance with four short fragments in the DNA fragment, which showed that the DNA fragment amplified was the genomic sequence of the *RPL37A* gene from giant panda. The genomic sequence of the *RPL37A* gene has been submitted to Genbank (accession number: HQ318092). Comparison of the nucleotide sequences of the genomic and cDNA sequences indicated that the genomic sequence of RPL37A possess four exons and three introns, which is also supported by restriction mapping of the genomic and cDNA sequences.

Prediction and analysis of functional sites on RPL37A protein of giant panda

Primary structure analysis revealed that the molecular weight of the putative RPL37A protein of giant panda was 10.23 kDa with a theoretical pI of 11.11, containing 24 positively charged amino acid residues (Arg, Lys and His), three negatively charged amino acid residues (Asp and Glu) and 65 uncharged residues. Among them, lysine (Lys) was highest and proline (Pro) was the lowest. Topology prediction showed that there were one N-glycosylation site, three protein kinase C phosphorylation sites, and three N-myristoylation sites in the RPL19 protein of giant panda (Figure 2).

Figure 2. Functional sites of RPL37A protein among five mammalian species. ▭, N-glycosylation site; ▭, protein kinase C phosphorylation site; ▬, N-myristoylation site; Pd, giant panda; Ho, *H. sapiens*; Po, *P. abelii*; Bo, *B. taurus*; Mu, *M. musculus*.

Table 1. Secondary structure of RPL37A protein among five mammalian species.

Species	Sequence length	Amino acid site (%)		
		Helix	Strand	Coil
A. melanoleuca	92	27.17	38.04	34.78
H. sapiens	92	29.35	33.70	36.96
B. taurus	92	29.35	33.70	36.96
M. musculus	92	29.35	33.70	36.96
P. abelii	92	29.35	33.70	36.96

Analysis of secondary structure of RPL37A protein from giant panda

The secondary structure prediction of RPL37A protein showed that 38.04% of the protein sequence was strand, 27.17% was helix and 34.78% was coil. The secondary structure of RPL37A protein from the giant panda was compared with RPL37A proteins reported for other four mammalian species (Table 1). It showed that there was very little difference in secondary structure of RPL37A protein among five mammalian species.

Over-expression of the *RPL37A* gene in *E. coli*

The *RPL37A* gene was over-expressed in *E. coli*, using pET28a plasmids carrying strong promoter and terminator sequences derived from phage T7. Therefore, the *RPL37A* gene was amplified individually by PCR and cloned into a pET28a plasmid, resulting in a gene fusion endocoding a protein bearing a His-tag extension at the N terminus. Protein expression was tested by SDS-PAGE analysis of protein extracts from recombinant *E. coli* BL21 strains (Figure 3). The results indicate that the protein RPL37A fused with the N-terminally His-tagged form gave rise to the accumulation of an expected 14 kDa polypeptide that formed inclusion bodies. Apparently, the recombinant protein was expressed after half an hour of induction and reached the highest level after 2.5 h. The expression product obtained could be used to purify the protein and to study its function further.

DISCUSSION

In this study, we cloned the genomic sequence and the cDNA clone encoding a RPL37A protein from giant panda. The genomic sequence of *RPL37A* was 2587 bp in size. Comparison of the nucleotide sequences of the genomic and cDNA sequences indicated that the genomic sequence of RPL37A possess four exons and three introns, which is also supported by restriction mapping of the genomic and cDNA sequences. As compared to other *RPL37A* sequences of some mammals incluing *H. sapiens, P. abelii, M. musculus* and *B. taurus*, there were same exons and introns. However, there were differences in length in the genomic sequence, introns, the 5′-untranslated sequence and the

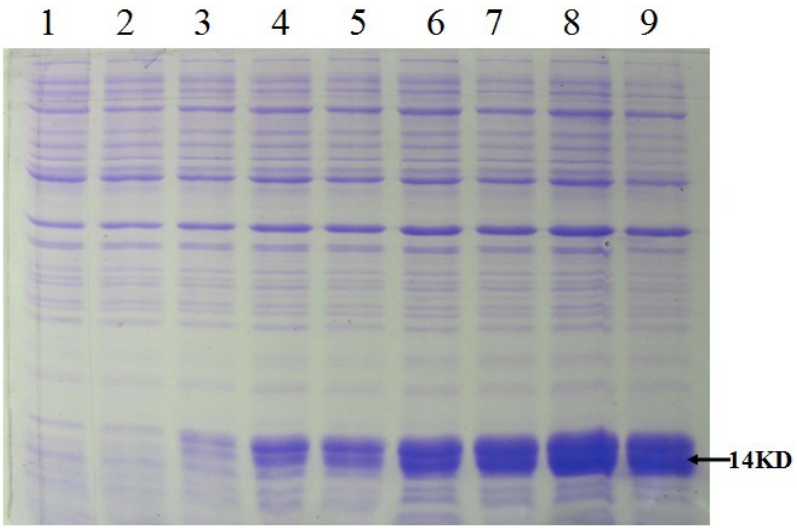

Figure 3. Protein over-expression in *E. coli*. Lane 1: control; lanes 2 to 9: 0, 0.5, 1, 1.5, 2, 2.5, 3 and 4 h of induction, respectively; arrow, molecular weight.

Table 2. *RPL37A* genomic sequence among five mammalian species.

Species	Size (bp)	Number of exon	Join sites in the CDS	GenBank accession number
A. melanoleuca	2587	4	8 to 10, 395 to 523, 1145 to 1227, 2524 to 2587	HQ_318092
H. sapiens	2671	4	79 to 81, 474 to 602, 1153 to 1235, 2545 to 2608	NC_000002
P. abeli	2626	4	31 to 33, 433 to 561, 1111 to 1193, 2505 to 2568	NC_012593
M. musculus	2554	4	62 to 64, 413 to 541, 883 to 965, 2243 to 2306	NC_000067
B. tauru	2623	4	31 to 33, 398 to 526, 1091 to 1173, 2207 to 2270	NC_007300

CDS, Coding sequence.

Table 3. Comparison of RPL37A from giant panda with other four species.

Items	Species			
	H. sapiens	***P. abelii***	***B. taurus***	***M. musculus***
CDS similarity (%)	90.32	90.32	92.83	89.96
AA similarity (%)	98.91	98.91	98.91	98.91
Molecular weight (kD)	10.27	10.27	10.27	10.27
pI	11.11	11.11	11.11	11.11

CDS, Coding sequence.

3′-untranslated sequence (Table 2). The variations in lengths of the introns determined the lengths of the *RPL37A* gene.

The alignment analysis of the cDNA sequence and the deduced amino acid sequence of *RPL37A* gene between giant panda and other mammals reported including *H. sapiens, P. abelii, M. musculus* and *B. taurus*, showed that giant panda shared high homology in the nucleotide sequence with the four mammals above, and the highest homology with *B. taurus*. Meanwhile, it shared the highest homology in amino acid sequences with *H. sapiens, P. abelii, M. musculus* and *B. taurus* (Table 3). In comparison of *RPL37A* genetic coding sequence between giant panda and four mammals, some variable sites were found. Among these variable sites, some are degeneration sites and others are single variable sites. Further analysis indicates that those variable sites were caused by transformation or transition of bases, which did not result in changes in the amino acid sequences encoded. This belonged to synonymous mutations, which did not cause changes in functional sites of the giant panda *RPL37A* gene (Li et al., 2010). Eventually,

functional sites in RPL37A protein of the five mammalians had the same species, positions and numbers. These results indicate the coding sequence of *RPL37A* and the deduced amino acid sequence were highly conserved. This striking pattern of evolutionary conservation is reasonable, as ribosomal protein genes are a group of highly conserved housekeeping genes (Wool, 1979).

According to the comparison, the results show that the molecular weight of the putative protein among the five mammalians was the same and that the theoretical pI was exactly identical (Table 3). Secondary structure analysis show that although the amino acid sequence had different structure, this did not cause changes in their functional genes, and the corresponding functional sites was not changed. In this sense, RPL37A from giant panda shared high homology with other four mammals reported.

The obtained *RPL37A* gene was expressed efficiently in a prokaryotic organism such as *E. coli* using pET28a plasmids, and the gained fusion protein was in accordance with the expected 14 kDa polypeptide. These results suggest that the protein was successfully expressed in the bacterial vector and it was just the protein encoded by the *RPL37A* from giant panda. The expression product obtained could be used for further purification and to study its function.

In summary, the complete coding sequence of *RPL37A* gene has been cloned using RT-PCR technology successfully. The characteristics of genomic sequence and cDNA clones encoding ribosomal proteins would be beneficial in the study of ribosomal biogenesis and would allow the elucidation of structure, organization and regulation of genes encoding ribosomal proteins in eukaryote. These data will enrich and supplement the information about *RPL37A* gene in Genebank databases. In addition, it will contribute to the protection of gene resources and the discussion of the biological diversity of this endangered species.

ACKNOWLEDGMENTS

The authors acknowledge Key Chinese National Natural Science Foundation (30570275), Application Foundation Project in Sichuan Province (2009JY0061), Startup Project of China West Normal University (06B026) and Youth Fund Project of Educational Committee of Sichuan Province (07ZB089).

REFERENCES

Andrei AK (2011). Structural aspects of translation termination on the ribosome. RNA- a publication of the RNA Society. 8(17): 1409-1421.

Barnard GF, Staniunas RJ, Puder M, Steele Jr. GD, Chen LB (1994). Human ribosomal protein L37 has motifs predicting serine/threonine phosphorylation and a zinc-finger domain. Biochim. Biophys. Acta. 1218(3): 425-428.

Du YJ, Luo XY, Hao YZ, Zhang T, Hou WR (2007). Cloning and Overexpression of Acidic Ribosomal Phosphoprotein P1 Gene (*RPLP1*) from the Giant Panda. Inter. J. Biol. Sci. 3: 428-433.

Ewing RM, Chu P, Elisma F, Li H, O'Connor L, Topaloglou T, Figeys D (2007). Large-scale Mapping of Human Protein-Protein Interactions by Mass Spectrometry. Mol. Syst. Biol. 3: p. 89.

Hou YL, Hou WR, Ren ZL, Hao YZ, Zhang T (2008). cDNA, Genomic Sequence and Overexpression of Crystalline alpha-B Gene (CRYAB) of the Giant Panda. Int. J. Bio. Sci. 4: 415-421.

Hou YL, Du YJ, Hou WR, Zhou CQ, Hao YZ, Zhang T (2009a). Cloning and Sequence analysis of translocase of inner mitochondrial membrane 10 homolog (yeast) gene (TIMM10) from the giant panda. J. Cell Anim. Biol. 3(1): 9-14.

Hou YL, Hou WR, Ren ZL, Hao YZ, Zhang T (2009b). cDNA Cloning and Overexpression of Ribosomal Protein S19 Gene (*RPS19*) from the Giant Panda. DNA. Cell. Biol. 28(1): 41-47.

Jennie PM, Alison M, Rong HL (1992). Activins, Inhibins, and Follistatins: Further Thoughts on a Growing Family of Regulator. Biol. Med. 201: 1-15.

Li RQ, Fan W, Tian G, Zhu HM (2010). The sequence and de novo assembly of the giant panda genome. Nature, 463: 311-317.

Qiao Q, Ma QY, Zhu H, Chen DK (2009). Study on the Cloning, in vitro Expression and Bioactivity of *Ailuropoda Melanoleuca* Qinling Subspecies Interleukin-2. Chin. J. Cell. Mol. Imunol. 25: 479-482.

Schwartz EI, Intine RV, Maraia RJ (2004). CK2 is responsible for phosphorylation of human La protein serine-366 and can modulate rpL37 5'-terminal oligopyrimidine mRNA metabolism. Mol. Cell Biol. 24(21): 9580-9591.

Stelzl U, Worm U, Lalowski M, Haenig C, Brembeck FH (2005). A human protein-protein interaction network: a resource for annotating the proteome. Cell, 122(6): 957-968.

Susana L, Manuel S (2010). Depletion of ribosomal protein L37 occurs in response to DNA damage and activates p53 through the L11/MDM2 pathway. Cell cycle, 9(19): 405-412.

Sun B, Hou YL, Hou WR, Su XL, Li J, Wu GF, Song Y (2011). cDNA, Genomic Sequence Cloning and Over-expression of Ribosomal Protein L15 gene (*RPL15*) from the Giant Panda. Afr. J. Agric. Res. 6(9): 2108-2114

Wang H, Zhao LN, Li KZ (2006). Overexpression of Ribosomal Protein L15 is Associated with Cell Proliferation in Gastric Cancer. BMC Cancer, 6: 91-92.

Wool IG (1979). The structure and function of eukaryotic ribosomes. Annu. Rev. Biochem. 48: 719-754.

Wu GF, Hou YL, Hou WR, Song Y, Zhang T (2010). Giant Panda Ribosomal Protein S14: cDNA, Genomic Sequence Cloning, Sequence Analysis, and Overexpression. Genet. Mol. Res. 9: 2004-2015.

Yang F, Liu WP (2005). Study on the Relation of Ribosomal Protein Gene and Human Diseases. Chinese J. Clin. Exp. Pathol. 3: 354-356.

Zhang T, Hou WR, Hou YL, Hao YZ, Du YJ, Wu GF (2009). cDNA, Genomic Sequence Cloning and Overexpression of Ribosomal Protein S20 gene (*RPS20*) from the Giant Panda (*Ailuropoda melanoleuca*). Afr. J. Biotechnol. 8: 5627-5632.