



# Giant panda ribosomal protein S14: cDNA, genomic sequence cloning, sequence analysis, and overexpression

G.-F. Wu, Y.-L. Hou, W.-R. Hou, Y. Song and T. Zhang

Key Laboratory of Southwest China Wildlife Resources Conservation  
(Ministry of Education), College of Life Science,  
China West Normal University, Nanchong, China

Corresponding author: W.-R. Hou  
E-mail: hwr168@yahoo.com.cn

Genet. Mol. Res. 9 (4): 2004-2015 (2010)  
Received June 3, 2010  
Accepted July 17, 2010  
Published October 13, 2010  
DOI 10.4238/vol9-4gmr899

**ABSTRACT.** RPS14 is a component of the 40S ribosomal subunit encoded by the *RPS14* gene and is required for its maturation. The cDNA and the genomic sequence of *RPS14* were cloned successfully from the giant panda (*Ailuropoda melanoleuca*) using RT-PCR technology and touchdown-PCR, respectively; they were both sequenced and analyzed. The length of the cloned cDNA fragment was 492 bp; it contained an open-reading frame of 456 bp, encoding 151 amino acids. The length of the genomic sequence is 3421 bp; it contains four exons and three introns. Alignment analysis indicates that the nucleotide sequence shares a high degree of homology with those of *Homo sapiens*, *Bos taurus*, *Mus musculus*, *Rattus norvegicus*, *Gallus gallus*, *Xenopus laevis*, and *Danio rerio* (93.64, 83.37, 92.54, 91.89, 87.28, 84.21, and 84.87%, respectively). Comparison of the deduced amino acid sequences of the giant panda with those of these other species revealed that the RPS14 of giant panda is highly homologous with those of *B. taurus*, *R. norvegicus* and *D. rerio* (85.99, 99.34 and 99.34%, respectively), and is 100% identical with the others. This degree of conservation of RPS14 suggests evolutionary selection. Topology prediction shows that there

are two N-glycosylation sites, three protein kinase C phosphorylation sites, two casein kinase II phosphorylation sites, four N-myristoylation sites, two amidation sites, and one ribosomal protein S11 signature in the RPS14 protein of the giant panda. The *RPS14* gene can be readily expressed in *Escherichia coli*. When it was fused with the N-terminally His-tagged protein, it gave rise to accumulation of an expected 22-kDa polypeptide, in good agreement with the predicted molecular weight. The expression product obtained can be purified for studies of its function.

**Key words:** cDNA cloning; *RPS14*; Giant panda; Overexpression; *Ailuropoda melanoleuca*; Genomic cloning

## INTRODUCTION

Ribosomes are vital organelles, which catalyze protein synthesis in all living organisms (Ramakrishnan and Moore, 2001). Eukaryotic ribosomes consist of four RNA molecules (rRNAs) and approximately 80 different proteins and can be dissociated into a small and large subunit (Wool et al., 1995; Yoshihama et al., 2002; Hwang et al., 2004). The small subunit protein RPS14, the yeast homolog of the bacterial S11 protein, directly binds helix 23 of 18S rRNA (Fewell and Woolford Jr., 1999; Mayolo and Woolford Jr., 2003). Recently, it was reported that RPS14 is an essential protein necessary for the maturation of the 40S ribosomal subunit (Larkin et al., 1987; Moritz et al., 1990; Paulovich et al., 1993; Ford et al., 1999; Jakovljevic et al., 2004) and the processing of 18S pre-rRNA (Ferreira-Cerca et al., 2005; Ebert et al., 2008). And in the 40S subunit, RPS14 helps to form the mRNA exit channel along with its interacting protein RPS5 (Yusupov et al., 2001). Moreover, the RPS14 protein in yeast regulates its own expression, via binding to the 5' splice-site of *RPS14B* pre-mRNA, which forms a stem-loop structure with a bulged region caused by imperfect complementary base pairs (Li et al., 1995). In humans, *RPS14* maps to the commonly deleted region at 5q31-q32 (Boultonwood et al., 2002) and is identified as a primary candidate gene, and its loss may be a causative factor of 5q- syndrome (Boultonwood et al., 2007), which is a hematologic malignancy characterized by a distinct phenotype and interstitial deletion of the long arm of chromosome 5 (Boultonwood et al., 1994; Giagounidis et al., 2004). Since RPS14 plays such an important role and its primary structure and function in giant panda has not been defined, it is significant to clone and analyze the *RPS14* gene of the giant panda and its significance lies not only in the protection of the giant panda, but also in the therapy for several kinds of human hereditary diseases.

The giant panda (*Ailuropoda melanoleuca*) is a rare species currently found only in China. It is at high risk of extinction because of human population expansion and destruction of its habitat. At present, there is very little genetic information for the panda (Liao et al., 2003; Du et al., 2007; Hou et al., 2008, 2009; Zhang et al., 2009), which is an essential tool for detailed understanding of the biology of this organism. Recently, the giant panda's draft genome sequence, which has good coverage and completeness for genes with unique sequences, has been generated and assembled (Li et al., 2010), but some genes are still incomplete in this project. Using conventional methods, we obtained the ribosomal protein (r-protein) S14 cDNA clone and genomic sequence clone from the skeletal muscle of the giant panda. In

further analysis, our data are intact and helped us clearly understand the extent of giant panda *RPS14*. We then analyzed the sequence characteristics of the protein encoded by the cDNA and compared it with vertebrate reported. We also overexpressed it in *Escherichia coli* using pET28a plasmids. Consequently, our study provides scientific data for inquiring into the hereditary traits of the gene from giant panda and these data will enrich and supplement the information about *RPS14*.

## MATERIAL AND METHODS

### Material

Skeletal muscle was collected from a dead giant panda at the Wolong Conservation Center of the Giant Panda, Sichuan, China. The collected skeletal muscle was frozen in liquid nitrogen and then used for RNA and DNA isolation.

### DNA and RNA isolation

A total of 500 mg muscle tissue from giant panda was ground in liquid nitrogen to a fine powder, and the powder was suspended completely in 15 mL lysis buffer containing 10 mM Tris-HCl, pH 8.0, 100 mM EDTA and 0.5% SDS. After treatment with proteinase K (100 mg/mL, final concentration) at 55°C for 3 h, the mixture was then cooled to room temperature and mixed with an equal volume of saturated phenol (pH 8) before being centrifuged at 5000 g at 4°C for 20 min. The supernatant was pooled and then mixed with an equal volume of 1:1 (v:v) phenol-chloroform and then centrifuged as above and the supernatant collected, from which the DNA was precipitated by ethanol. The DNA obtained was then dissolved in TE buffer and kept at -20°C.

Total RNAs were isolated from about 400 mg muscle tissue using the Total Tissue/Cell RNA Extraction Kits (Waton Inc., Shanghai, China) according to manufacturer instructions. The total RNAs extracted were dissolved in DEPC (diethyl pyrocarbonate) water, and kept at -70°C.

### Primer design, RT-PCR, cloning of RT-PCR products, and sequencing

The polymerase chain reaction (PCR) primers were designed by Primer Premier 5.0, based on the mRNA sequence of *RPS14* from *Homo sapiens* (NM\_001025070), *Mus musculus* (NM\_020600), *Rattus norvegicus* (NM\_022672), and *Gallus gallus* (NM\_001030619). The specific primers of cDNA sequences were as follows: *RPS14* forward: 5'-GGAGTCTGGAGAAGACGCTCA-3'; *RPS14* reverse: 5'-GAGAAGTCTTGTTACAGACG-3'.

Total RNAs were synthesized into the first-stranded cDNAs using a reverse transcription kit with Oligo dT as the primers according to manufacturer instructions (Promega, USA). Twenty microliters of the first-strand cDNA synthesis reaction system was included in 1 µg total RNAs, 5 mM MgCl<sub>2</sub>, 1 mM dNTPs, 0.5 µg Oligo dT<sub>15</sub>, 10 U/µL RNase inhibitor, and 15 U AMV reverse transcriptase, and 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 MgCl<sub>2</sub>, 200 µM of

each of dATP, dGTP, dCTP and dTTP (Omega, China), 0.3  $\mu$ M of each primer, 5.0 units Taq plus DNA polymerase (Sangon Co., Shanghai, China). DNA amplification was performed using an MJ Research thermocycler, Model PTC-200 (Watertown, MA, USA) with a program of 4 min at 94°C, followed by 30 cycles of 1 min at 94°C, 0.5 min at 48°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 on 1.5% agarose gel with 1X TAE (Tris-acetate-EDTA) buffer, stained with ethidium bromide and visualized under UV light. The expected fragments of PCR products were harvested and purified from gel using a DNA harvesting kit (Omega), and then ligated into a pMD19-T vector at 4°C for 8 h. The recombinant molecules were transformed into *E. coli* competent cells (DH5 $\alpha$ ), and then spread on an LB-plate containing 50  $\mu$ g/mL ampicillin, 200 mg/mL IPTG (isopropyl-beta-D-thiogalactopyranoside), and 20 mg/mL X-gal. Plasmid DNA was isolated and digested by *Pst*I and *Sca*II to verify the insert size. Plasmid DNA was sequenced by Huada Zhongsheng Scientific Corporation (Beijing, China).

### Cloning the genomic sequence of *RPS14*

The PCR primers were the same as the *RPS14* forward and *RPS14* reverse presented above. The genomic sequence of the *RPS14* gene was amplified using Touchdown-PCR with the following conditions: 94°C for 30 s, 55°C for 45 s, 72°C for 4 min in the first cycle and the annealing temperature decreased 0.5°C per cycle; after 20 cycles conditions changed to 94°C for 30 s, 45°C for 45 s, 72°C for 4 min for another 15 cycles, and then ended with the final extension for 10 min at 72°C. The fragment amplified was also purified, ligated into the clone vector and transformed into *E. coli* competent cells. Finally, the recombinant fragment was sequenced by Sangon.

### Construction of the expression vector and overexpression of recombinant *RPS14*

The PCR fragment corresponding to the RPS14 polypeptide was amplified from the *RPS14* cDNA clone with the forward primer, 5'-ACTGGATCCATGGCGCCTCGAAAG-3' (*Bam*HI) and reverse primer, 5'-GACGCGGCCGCTCACAGACGGCGACCA-3' (*Not*I), respectively. PCR was performed at 94°C for 3 min; 30 cycles of 30 s at 94°C, 45 s at 52°C, 1 min at 72°C, and 10 min at 72°C. The amplified PCR product was cut and ligated into corresponding site of the pET28a vector (Stratagene, USA). The resulting construct was transformed into the *E. coli* BL21 (DE3) strain (Novagen, USA) and used for induction by adding IPTG at an OD 600 of 0.6 and culturing 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, 1, 1.5, 2, 2.5, and 3 h and then separated by SDS-PAGE and stained with Coomassie blue R 250.

### Data analysis

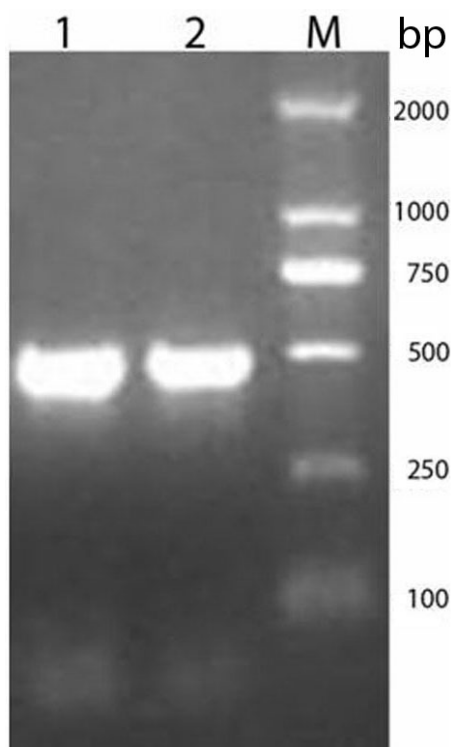
The sequence data were analyzed by the GenScan software (<http://genes.mit.edu/GENSCAN.html>). Homology research of the giant panda *RPS14* compared with the gene sequences of other species was performed using Blast 2.1 (<http://www.ncbi.nlm.nih.gov/blast/>). Open-reading frame (ORF) of the DNA sequence was searched using the ORF finder software (<http://www.ncbi.nlm.nih.gov/gorf/gorf.html>). Protein structure of the *RPS14* sequence

cloned was deduced using the PredictProtein software (<http://cubic.bioc.columbia.edu/predictprotein/>). Prediction of the RPS14 sequence was simulated by the SWISS-MODEL software (<http://swiss-model.expasy.org/>).

## RESULTS

### Analysis of the cDNA of *RPS14* from the giant panda

About 500 bp of the cDNA fragment was amplified from the giant panda with the primers *RPS14* forward and *RPS14* reverse (Figure 1). The length of the cDNA cloned was 492 bp.



**Figure 1.** Reverse transcription polymerase chain reaction products of giant panda *RPS14*. Lanes 1 and 2 = Amplified *RPS14*; M = molecular ladder DL2000.

Blast research showed that the cDNA sequence cloned is highly homologous with the *RPS14* sequences from human and other mammalian species. On the basis of the high identity, we concluded that the cDNA we isolated is the cDNA encoding the giant panda RPS14 protein. The *RPS14* cDNA sequence was submitted to GeneBank (accession No. HM153519). The giant panda *RPS14* sequence we cloned contains 24-bp 5'-untranslated sequence and 12-bp 3'-untranslated region. An ORF of 456 bp encoding 151 amino acids was found in the cDNA (Figure 2).



```

1   GGAGTCTGGAGAAGACGCTCAGAA ATG GCG CCT CGA AAG GGG AAG GAA AAG AAG GAA GAA CAG GTC
1   M   A   P   R   K   G   K   E   K   K   E   E   Q   V
68  ATC AGC CTT GGA CCT CAG GTA GCT GAA GGA GAA AAT GTG TTT GGT GTC TGC CAC ATC TTT GCA
15  I   S   L   G   P   Q   V   A   E   G   E   N   V   F   G   V   C   H   I   F   A
130 TCC TTC AAT GAC ACT TTT GTG CAT GTC ACC GAT CTT TCT GGC AAG GAA ACC ATC TGC CGC GTG
36  S   F   N   D   T   F   V   H   V   T   D   L   S   G   K   E   T   I   C   R   V
193 ACT GGT GGA ATG AAG GTG AAG GCT GAC CGA GAT GAG TCC TCT CCC TAC GCT GCC ATG TTG GCT
57  T   G   G   M   K   V   K   A   D   R   D   E   S   S   P   Y   A   A   M   L   A
256 GCC CAG GAT GTA GCC CAG AGG TGC AAG GAG CTG GGC ATC ACT GCC CTC CAC ATC AAA CTA CGG
78  A   Q   D   V   A   Q   R   C   K   E   L   G   I   T   A   L   H   I   K   L   R
319 GCC ACA GGA GGA AAT AGA ACC AAG ACC CCT GGA CCT GGG GCC CAG TCA GCC CTC AGA GCC CTT
99  A   T   G   G   N   R   T   K   T   P   G   P   G   A   Q   S   A   L   R   A   L
382 GCC CGC TCA GGA ATG AAG ATC GGG CGG ATT GAG GAT GTC ACC CCC ATC CCC TCC GAT AGC ACC
120 A   R   S   G   M   K   I   G   R   I   E   D   V   T   P   I   P   S   D   S   T
445 CGC AGG AAG GGG GGT CGC CGT GGT CGC CGT CTG TGA ACAAGACTTCTC
141 R   R   K   G   G   R   R   G   R   R   L   *

```

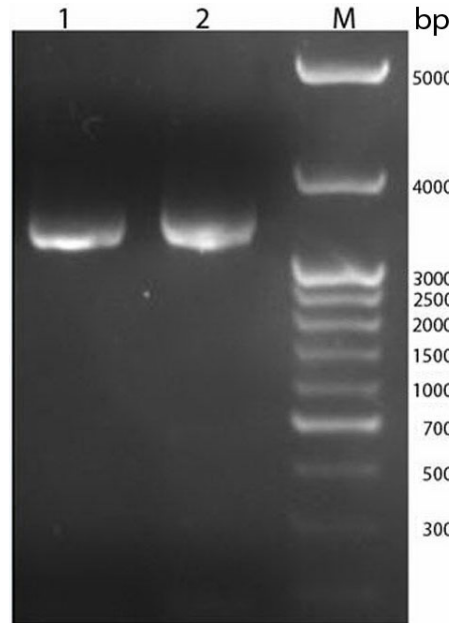
**Figure 2.** Nucleotide sequence of cDNA encoding the giant panda (*Ailuropoda melanoleuca*) *RPS14* and the amino acid sequence deduced from its ORF. The nucleotide and amino acid sequences are numbered on the left. The translation is given under the center of each codon (asterisk is the termination codon).

### Analysis of the genomic sequence of *RPS14* from the giant panda

A DNA fragment of about 3500 bp was amplified with primers *RPS14* forward and *RPS14* reverse (Figure 3). The length of the DNA fragment cloned is 3421 bp. Comparison between the cDNA sequence and the DNA fragment sequence of the *RPS14* amplified from giant panda was performed by the Lasergene software. These results indicated that the cDNA sequence is in total accordance with four short fragments in the DNA fragment, which shows that the DNA fragment amplified is the genomic sequence of the *RPS14* from giant panda. The genomic sequence of the *RPS14* reported in this paper has been submitted to GenBank and assigned the accession No. HM153520.

### Prediction and analysis of protein functional sites in *RPS14* protein of the giant panda

Primary structure analysis revealed that the molecular weight of the putative *RPS14* protein of the giant panda is 16.27 kDa with a theoretical pI = 10.07. Topology prediction shows that there are two N-glycosylation sites, three protein kinase C phosphorylation sites, two casein kinase II phosphorylation sites, four N-myristoylation sites, two amidation sites, and one ribosomal protein S11 signature in the *RPS14* protein of the giant panda (*A. melanoleuca*) (Figure 4).



**Figure 3.** PCR products of complete genomic sequence of RPS14 from giant panda. Lanes 1 and 2 = Complete sequences of amplified RPS14; M = molecular ladder BM5000.

```

PD MAP_RKGGKKEEQVISLGPQVABGENVPGVCHIFASFNRIIEVHVYTDLGGSTGTTICRVTGGMHVKAADRESSPYAAMLAA 79
HM MAP_RKGGKKEEQVISLGPQVABGENVPGVCHIFASFNRIIEVHVYTDLGGSTGTTICRVTGGMHVKAADRESSPYAAMLAA 79
BT MAP_RKGGKKEEQVISLGPQVABGENVPGVCHIFASFNRIIEVHVYTDLGGSTGTTICRVTGGMHVKAADRESSPYAAMLAA 79
MS MAP_RKGGKKEEQVISLGPQVABGENVPGVCHIFASFNRIIEVHVYTDLGGSTGTTICRVTGGMHVKAADRESSPYAAMLAA 79
RT MAP_RKGGKKEEQVISLGPQVABGENVPGVCHIFASFNRIIEVHVYTDLGGSTGTTICRVTGGMHVKAADRESSPYAAMLAA 79
GG MAP_RKGGKKEEQVISLGPQVABGENVPGVCHIFASFNRIIEVHVYTDLGGSTGTTICRVTGGMHVKAADRESSPYAAMLAA 79
XL MAP_RKGGKKEEQVISLGPQVABGENVPGVCHIFASFNRIIEVHVYTDLGGSTGTTICRVTGGMHVKAADRESSPYAAMLAA 79
DR MAP_RKGGKKEEQVISLGPQVABGENVPGVCHIFASFNRIIEVHVYTDLGGSTGTTICRVTGGMHVKAADRESSPYAAMLAA 79

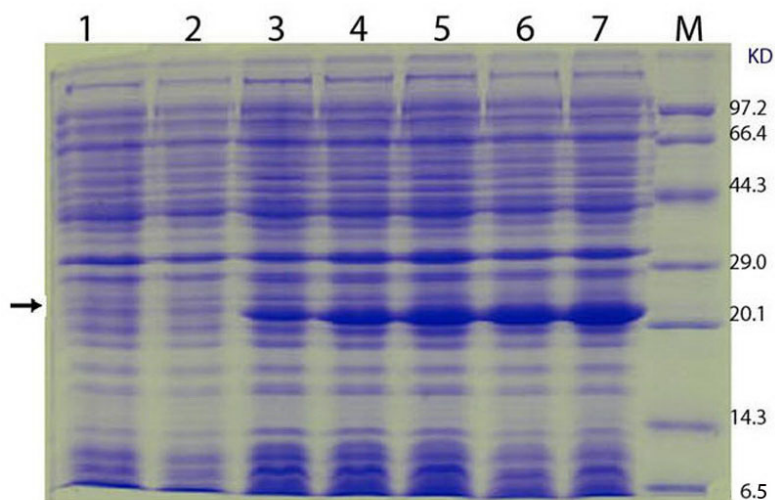
PD QDVAQRCKELGITALMHLKLRATGGRRTITPGFGAASATRALARSEMKIGRIEDVTFIPSD...STTRQGG_RKGGRL 151
HM QDVAQRCKELGITALMHLKLRATGGRRTITPGFGAASATRALARSEMKIGRIEDVTFIPSD...STTRQGG_RKGGRL 151
BT QDVAQRCKELGITALMHLKLRATGGRRTITPGFGAASATRALARSEMKIGRIEDVTFIPSD...STTRQGG_RKGGRL 151
MS QDVAQRCKELGITALMHLKLRATGGRRTITPGFGAASATRALARSEMKIGRIEDVTFIPSD...STTRQGG_RKGGRL 151
RT QDVAQRCKELGITALMHLKLRATGGRRTITPGFGAASATRALARSEMKIGRIEDVTFIPSD...STTRQGG_RKGGRL 151
GG QDVAQRCKELGITALMHLKLRATGGRRTITPGFGAASATRALARSEMKIGRIEDVTFIPSD...STTRQGG_RKGGRL 151
XL QDVAQRCKELGITALMHLKLRATGGRRTITPGFGAASATRALARSEMKIGRIEDVTFIPSD...STTRQGG_RKGGRL 151
DR QDVAQRCKELGITALMHLKLRATGGRRTITPGFGAASATRALARSEMKIGRIEDVTFIPSD...STTRQGG_RKGGRL 151

```

**Figure 4.** Comparison of the amino acid sequence of RPS14 between the giant panda and seven other vertebrate species. PD = the giant panda; HM = *Homo sapiens*; BT = *Bos taurus*; MS = *Mus musculus*; RT = *Rattus norvegicus*; GG = *Gallus gallus*; XL = *Xenopus laevis*; DR = *Danio rerio*. Shaded letters = N-glycosylation site; double-line squares = protein kinase C phosphorylation site; underlined dots = casein kinase II phosphorylation site; single-line squares = N-myristoylation site; solid lines = amidation site; double solid lines = ribosomal protein S11 signature; thin dotted lines = no amino acids.

### Overexpression of the *RPS14* gene in *E. coli*

The *RPS14* gene was overexpressed in *E. coli* using pET28a plasmids carrying strong promoter and terminator sequences derived from phage T7. For this purpose, the *RPS14* gene was amplified individually by PCR and cloned in a pET28a plasmid, resulting in a gene fusion coding for a protein bearing a His-tag extension at the N-terminus. Expression was tested by SDS-PAGE analysis of protein extracts from recombinant in *E. coli* BL21 strains (Figure 5). The results indicated that the protein RPS14 fusion with the N-terminally His-tagged form gave rise to the accumulation of an expected 22-kDa polypeptide that formed inclusion bodies. Apparently, the recombinant protein was expressed after half an hour of induction and after 2 h reached the highest level. These results suggested that the protein is active and just the protein encoded by the *RPS14* from the giant panda. The expression product obtained could be used to purify the protein and further study its function.



**Figure 5.** Proteins extracted from recombinant *Escherichia coli* strains were analyzed by SDS-PAGE gel stained with Coomassie blue R 250. Numbers on the right show the molecular weight, and the arrow indicates the recombinant protein bands induced by IPTG for 0, 1, 1.5, 2, 2.5, and 3 h (lanes 2-7), respectively. Lane 1 = products of the *E. coli* strains with the empty vectors. M = molecular marker.

### DISCUSSION

As determined by BLAST analysis, the nucleotide sequence *RPS14* cloned from the giant panda shares a high homology with those of *H. sapiens*, *Bos taurus*, *M. musculus*, *R. norvegicus*, *G. gallus*, *Xenopus laevis*, and *Danio rerio* (93.64, 83.37, 92.54, 91.89, 87.28, 84.21, and 84.87%, respectively). Comparison of the deduced amino acid sequences of giant panda with those of these species showed that the RPS14 shares a high homology with that of *B. taurus*, *R. norvegicus*, and *D. rerio* (85.99, 99.34, and 99.34%, respectively), and is entirely identical with the rest. This striking pattern of evolutionary conservation is reasonable, as ribosomal protein genes are a group of highly conserved housekeeping genes (Wool,



1979). Furthermore, we observed that the stop codon TGA is used in giant panda, *H. sapiens*, *M. musculus* and *R. norvegicus*, while *B. taurus* and *X. laevis* is terminated with TAA and two other species terminated with TAG. Physical and chemical analysis showed that the molecular weight of the putative protein among the five mammals is very close and that of the theoretical pI is exactly identical (Table 1).

**Table 1.** Molecular weight and pI of RPS14 of the giant panda and seven other vertebrate species.

	<i>A. melanoleuca</i>	<i>H. sapiens</i>	<i>B. taurus</i>	<i>M. musculus</i>	<i>R. norvegicus</i>	<i>G. gallus</i>	<i>X. laevis</i>	<i>D. rerio</i>
Molecular weight (kDa)	16.2727	16.2727	16.7764	16.2727	16.2586	16.2727	16.2727	16.2446
Theoretical (pI)	10.07	10.07	10.07	10.07	10.07	10.07	10.07	10.02

*A. melanoleuca* = *Ailuropoda melanoleuca*; *H. sapiens* = *Homo sapiens*; *B. taurus* = *Bos taurus*; *M. musculus* = *Mus musculus*; *R. norvegicus* = *Rattus norvegicus*; *G. gallus* = *Gallus gallus*; *X. laevis* = *Xenopus laevis*; *D. rerio* = *Danio rerio*.

The genomic sequence of *RPS14* is 3421 bp in size. A comparison of the nucleotide sequences of the genomic and cDNA sequences indicated that the genomic sequence of *RPS14* possesses four exons and three introns, which is also supported by restriction mapping of the genomic and cDNA sequences. Compared with some mammals, including *H. sapiens* (NC\_000005), *M. musculus* (NC\_000084), *R. norvegicus* (NC\_005117), and *Canis familiaris* (NC\_006586), the length of the genomic sequence, the first, second and third introns, and the 5'- and 3'-untranslated sequences are different (see Table 2). From the table we can see that *D. rerio*'s introns and *B. taurus*'s third intron differ dramatically with respect to their lengths. These diverse intron patterns in the table facilitate the evaluation of characteristics in *RPS14* evolution, although the origin and significance of various intron models remain a mystery (Rogers, 1990; Roy et al., 2003). Furthermore, we compared our data with the giant panda's draft genome sequence counterpart (accession No. GL192580), which contains only exon 1 to exon 3 and even its exon 3 is truncated, showing that our data are intact and can supplement information on *RPS14*.

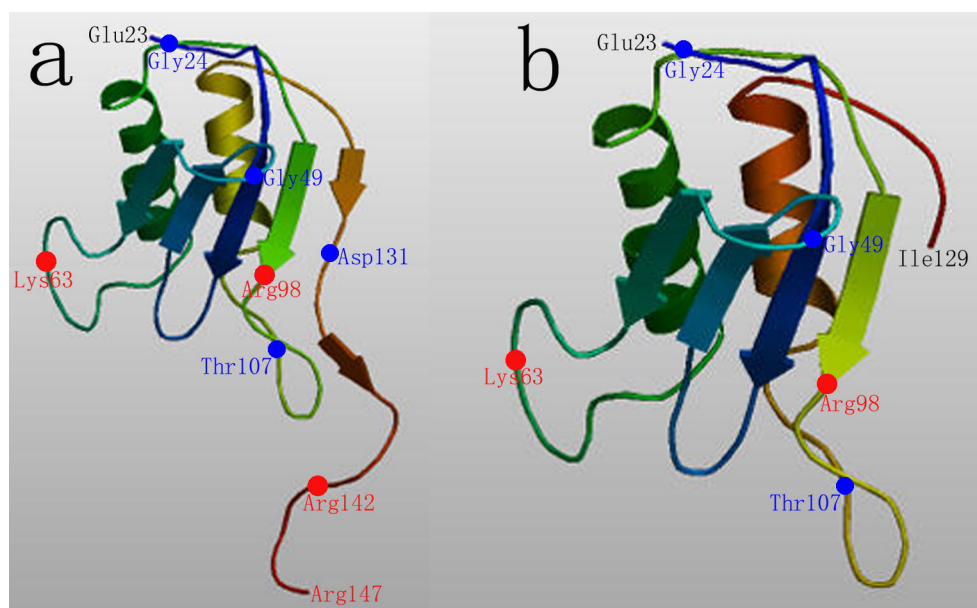
**Table 2.** Comparison of *RPS14* genomic sequence among eight vertebrate species.

Species	Genomic length	Number of exons	Number of introns	5' UTR (bp)	3' UTR (bp)	Intron 1	Intron 2	Intron 3	GenBank accession Nos.
<i>A. melanoleuca</i> ★	3421	4	3	24	12	617	1347	965	HM153520
<i>A. melanoleuca</i> ◆	2375	3	2	0	0	617	1347	0	GL192580
<i>H. sapiens</i>	5528	4	3	2023	57	621	1116	1255	NC_000005
<i>B. taurus</i>	20465	4	3	848	3039	493	1076	14535	NC_007305
<i>M. musculus</i>	3896	4	3	1754	46	378	721	541	NC_000084
<i>R. norvegicus</i>	4785	4	3	1837	0	711	1250	531	NC_005117
<i>C. familiaris</i>	6845	4	3	2592	56	1186	1561	994	NC_006586
<i>G. gallus</i>	4506	4	3	1062	1528	203	484	773	NC_006100
<i>D. rerio</i>	4292	4	3	1114	38	2493	112	79	NC_007132

The genomic sequence of giant panda cloned is marked with a star, while the lozenge represents the counterpart published in the draft genome sequence. *C. familiaris* = *Canis familiaris*. For the complete name of the other species, see legend to Table 1.

RPS14 protein from the giant panda is extremely basic (pI = 10.07), and its Arg and Lys residues are concentrated in both termini of the protein, in which the 10 amino-terminal residues contain four lysines and one arginine, while the 11 carboxyl-terminal amino acids include six arginine residues and one lysine residue. Previous experimental studies have dem-

onstrated that S14 protein's amino terminal (amino acid residues 1-20) was unnecessary for biological activity, and the amino acids 24-49 and 107-131 were crucial for the ribosomal protein's biological activity, as those two segments are required for the nucleolar functions that mediate S14 incorporation into pre-ribosomal particles and for the normal function of mature cytoplasmic ribosomes (Diaz and Roufa, 1992). In addition, Cys-31, His-43, Thr-45, Gly-49, Gly-111, Ser-114, Gly-123, Met-124, Ile-126, Gly-127, Asp-131, and Thr-133 in these two domains are the structural determinants that mediate S14 incorporation into pre-ribosomal particles (Martin and Roufa, 1997). Recently, it was found that the Asn-38 is also important for ribosome assembly, while Lys-63, Arg-98 and the 10 carboxy-terminal amino acid residues, RKGGRGRRL, are responsible for RPS14 binding to helix 23 of 18S rRNA (Mayolo and Woolford Jr., 2003). Subsequently, we enquired further into the functional sites of the amino acid sequences encoded by *RPS14* genes and found that there are two N-glycosylation sites, three protein kinase C phosphorylation sites, two casein kinase II phosphorylation sites, four N-myristoylation sites, two amidation sites, and one ribosomal protein S11 signature in the RPS14 protein of the giant panda, *H. sapiens*, *M. musculus*, *R. norvegicus*, *G. gallus*, *X. laevis*, and *D. rerio*. As RPS14's C-terminus in cattle varies widely, its functional sites are quite distinct from those species, and the discrepancy is that there is only one protein kinase C phosphorylation site, five N-myristoylation sites, and neither amidation site nor ribosomal protein S11 signature exist (Figure 4). Meanwhile, the variation near the carboxyl terminus also led to the difference of tertiary structure among all the compared species. In this paper, we described only the tertiary structure of giant panda and cattle (Figure 6) for the structures that are exactly the same among those species, with cattle excepted.



**Figure 6.** Stereo ribbon representation of panda (a) and cattle (b) were formed by partial sequences of RPS14. Compared with its homology model, the conserved sequences of giant panda are amino acid residues from 23 to 147 (E value:  $5.77e-62$ ), while the residues of cattle are confined to the area from 23 to 129 (E value:  $1.21e-52$ ).

In addition, the *RPS14* cDNA was expressed efficiently in a prokaryotic organism using pET28a plasmids, and the resulting fusion protein was in accordance with the expected 22-kDa polypeptide. These results suggest that the protein is active and is just the protein encoded by the *RPS14* from the giant panda (Figure 5). The expression product obtained could be used for purification and study of its function.

During the course of the research described above, the complete primary sequence for giant panda *RPS14* has been determined and published. The primary purpose for the accumulation of these data is to use them to arrive at a solution for the structure of this ribosomal subunit. However, the information may also help in understanding the evolution of ribosomes, in unraveling the function of proteins, in defining the rules that govern the interaction of proteins and rRNAs. And this gene structure also provides important first steps toward understanding and carrying out further research on the genetic basis of panda's biology, and contributes to disease control and conservation efforts for this endangered species.

## ACKNOWLEDGMENTS

Research supported by the Key Chinese National Natural Science Foundation (#30470261), Application Technology Project in Sichuan Province (#2006J13-057), Key Scientific Research Foundation of Educational Committee of Sichuan Province (#07ZA120), and Key Discipline Construction Project in Sichuan Province (#SZD0420). Sichuan key discipline zoology construction funds subsidization project (#404001). Application infrastructure projects in Sichuan Province (#2009JY0061), Education Department Youth Fund Project of Sichuan Province (#09ZB088).

## REFERENCES

- Boultonwood J, Lewis S and Wainscoat JS (1994). The 5q-syndrome. *Blood* 84: 3253-3260.
- Boultonwood J, Fidler C, Strickson AJ, Watkins F, et al. (2002). Narrowing and genomic annotation of the commonly deleted region of the 5q- syndrome. *Blood* 99: 4638-4641.
- Boultonwood J, Pellagatti A, Cattani H, Lawrie CH, et al. (2007). Gene expression profiling of CD34+ cells in patients with the 5q- syndrome. *Br. J. Haematol.* 139: 578-589.
- Diaz JJ and Roufa DJ (1992). Fine-structure map of the human ribosomal protein gene *RPS14*. *Mol. Cell Biol.* 12: 1680-1686.
- Du YJ, Luo XY, Hao YZ, Zhang T, et al. (2007). cDNA cloning and overexpression of acidic ribosomal phosphoprotein P1 gene (RPLP1) from the giant panda. *Int. J. Biol. Sci.* 3: 428-433.
- Ebert BL, Pretz J, Bosco J, Chang CY, et al. (2008). Identification of *RPS14* as a 5q- syndrome gene by RNA interference screen. *Nature* 451: 335-339.
- Ferreira-Cerca S, Poll G, Gleizes PE, Tschochner H, et al. (2005). Roles of eukaryotic ribosomal proteins in maturation and transport of pre-18S rRNA and ribosome function. *Mol. Cell* 20: 263-275.
- Fewell SW and Woolford JL Jr (1999). Ribosomal protein S14 of *Saccharomyces cerevisiae* regulates its expression by binding to *RPS14B* pre-mRNA and to 18S rRNA. *Mol. Cell Biol.* 19: 826-834.
- Ford CL, Randal-Whitis L and Ellis SR (1999). Yeast proteins related to the p40/laminin receptor precursor are required for 20S ribosomal RNA processing and the maturation of 40S ribosomal subunits. *Cancer Res.* 59: 704-710.
- Giagounidis AA, Germing U, Haase S, Hildebrandt B, et al. (2004). Clinical, morphological, cytogenetic, and prognostic features of patients with myelodysplastic syndromes and del(5q) including band q31. *Leukemia* 18: 113-119.
- Hou WR, Sun GL, Chen Y, Wu X, et al. (2008). Molecular cloning of ribosomal protein L26 (RPL26) cDNA from *Ailuropoda melanoleuca* and its potential value in phylogenetic study. *Biochem. Syst. Ecol.* 36: 194-200.
- Hou YL, Hou WR, Ren ZL, Hao YZ, et al. (2009). cDNA cloning and overexpression of ribosomal protein S19 gene (RPS19) from the giant panda. *DNA Cell Biol.* 28: 41-47.

- Hwang KC, Cui XS, Park SP, Shin MR, et al. (2004). Identification of differentially regulated genes in bovine blastocysts using an annealing control primer system. *Mol. Reprod. Dev.* 69: 43-51.
- Jakovljevic J, de Mayolo PA, Miles TD, Nguyen TM, et al. (2004). The carboxy-terminal extension of yeast ribosomal protein S14 is necessary for maturation of 43S preribosomes. *Mol. Cell Biol.* 14: 331-342.
- Larkin JC, Thompson JR and Woolford JL Jr (1987). Structure and expression of the *Saccharomyces cerevisiae* CRY1 gene: a highly conserved ribosomal protein gene. *Mol. Cell Biol.* 7: 1764-1775.
- Li R, Fan W, Tian G, Zhu H, et al. (2010). The sequence and *de novo* assembly of the giant panda genome. *Nature* 463: 311-317.
- Li Z, Paulovich AG and Woolford JL Jr (1995). Feedback inhibition of the yeast ribosomal protein gene CRY2 is mediated by the nucleotide sequence and secondary structure of CRY2 pre-mRNA. *Mol. Cell Biol.* 15: 6454-6464.
- Liao MJ, Zhu MY, Zheng X, Zhang ZH, et al. (2003). cDNA cloning of growth hormone from giant panda (*Ailuropoda melanoleuca*) and its expression in *Escherichia coli*. *Comp. Biochem. Physiol. B. Biochem. Mol. Biol.* 135: 109-116.
- Martin NJ and Roufa DJ (1997). Functional analysis of human *RPS14* null alleles. *J. Cell Sci.* 110: 955-963.
- Mayolo PA and Woolford JL Jr (2003). Interactions of yeast ribosomal protein rpS14 with RNA. *J. Mol. Biol.* 333: 697-709.
- Moritz M, Paulovich AG, Tsay YF and Woolford JL Jr (1990). Depletion of yeast ribosomal proteins L16 or rp59 disrupts ribosome assembly. *J. Cell Biol.* 111: 2261-2274.
- Paulovich AG, Thompson JR, Larkin JC, Li Z, et al. (1993). Molecular genetics of cryptopleurine resistance in *Saccharomyces cerevisiae*: expression of a ribosomal protein gene family. *Genetics* 135: 719-730.
- Ramakrishnan V and Moore PB (2001). Atomic structures at last: the ribosome in 2000. *Curr. Opin. Struct. Biol.* 11: 144-154.
- Rogers JH (1990). The role of introns in evolution. *FEBS Lett.* 268: 339-343.
- Roy SW, Fedorov A and Gilbert W (2003). Large-scale comparison of intron positions in mammalian genes shows intron loss but no gain. *Proc. Natl. Acad. Sci. U. S. A.* 100: 7158-7162.
- Wool IG (1979). The structure and function of eukaryotic ribosomes. *Annu. Rev. Biochem.* 48: 719-754.
- Wool IG, Chan YL and Gluck A (1995). Structure and evolution of mammalian ribosomal proteins. *Biochem. Cell Biol.* 73: 933-947.
- Yoshihama M, Uechi T, Asakawa S, Kawasaki K, et al. (2002). The human ribosomal protein genes: sequencing and comparative analysis of 73 genes. *Genome Res.* 12: 379-390.
- Yusupov MM, Yusupova GZ, Baucom A, Lieberman K, et al. (2001). Crystal structure of the ribosome at 5.5 Å resolution. *Science* 292: 883-896.
- Zhang T, Hou WR, Hou YL, Hao YZ, et al. (2009). cDNA, genomic sequence cloning and overexpression of ribosomal protein s20 gene (RPS20) from the giant panda (*Ailuropoda melanoleuca*). *African J. Biotech.* 8: 5627-5632