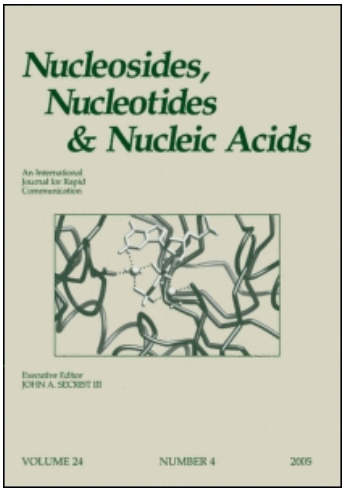


**Nucleosides, Nucleotides and Nucleic Acids**
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713597286

# cDNA, Genomic Sequence Cloning, and Overexpression of *EIF1* from the Giant Panda (*Ailuropoda Melanoleuca*) and the Black Bear (*Ursus Thibetanus Mupinensis*)

Wan-ru Hou[a]; Yun Tang[a]; Yi-ling Hou[a]; Yan Song[a]; Tian Zhang[a]; Guang-fu Wu[a]
[a] College of Life Science, China West Normal University, Nanchong, China

Online publication date: 28 June 2010

**To cite this Article** Hou, Wan-ru , Tang, Yun , Hou, Yi-ling , Song, Yan , Zhang, Tian and Wu, Guang-fu(2010) 'cDNA, Genomic Sequence Cloning, and Overexpression of *EIF1* from the Giant Panda (*Ailuropoda Melanoleuca*) and the Black Bear (*Ursus Thibetanus Mupinensis*)', Nucleosides, Nucleotides and Nucleic Acids, 29: 7, 547 — 561
**To link to this Article: DOI:** 10.1080/15257770.2010.487506
**URL:** http://dx.doi.org/10.1080/15257770.2010.487506

*Nucleosides, Nucleotides and Nucleic Acids*, 29:547–561, 2010

ISSN: 1525-7770 print / 1532-2335 online
DOI: 10.1080/15257770.2010.487506



# cDNA, GENOMIC SEQUENCE CLONING, AND OVEREXPRESSION OF *EIF1* FROM THE GIANT PANDA (*AILUROPODA MELANOLEUCA*) AND THE BLACK BEAR (*URSUS THIBETANUS MUPINENSIS*)

**Wan-ru Hou, Yun Tang, Yi-ling Hou, Yan Song, Tian Zhang, and Guang-fu Wu**
*College of Life Science, China West Normal University, Nanchong, China*

□ *Eukaryotic initiation factor (eIF) EIF1 is a universally conserved translation factor that is involved in translation initiation site selection. The cDNA and the genomic sequences of* EIF1 *were cloned successfully from the giant panda (*Ailuropoda melanoleuca*) and the black bear (*Ursus thibetanus mupinensis*) using reverse transcription polymerase chain reaction (RT-PCR) technology and touchdown-polymerase chain reaction, respectively. The cDNAs of the* EIF1 *cloned from the giant panda and the black bear are 418 bp in size, containing an open reading frame (ORF) of 342 bp encoding 113 amino acids. The length of the genomic sequence of the giant panda is 1909 bp, which contains four exons and three introns. The length of the genomic sequence of the black bear is 1897 bp, which also contains four exons and three introns. Sequence alignment indicates a high degree of homology to those of* Homo sapiens, Mus musculus, Rattus norvegicus, *and* Bos Taurus *at both amino acid and DNA levels. Topology prediction shows there are one N-glycosylation site, two Casein kinase II phosphorylation sites, and a Amidation site in the EIF1 protein of the giant panda and black bear. In addition, there is a protein kinase C phosphorylation site in EIF1 of the giant panda. The giant panda and the black bear* EIF1 *genes were overexpressed in* E. coli *BL21. The results indicated that the both EIF1 fusion proteins with the N-terminally His-tagged form gave rise to the accumulation of two expected 19 kDa polypeptide. The expression products obtained could be used to purify the proteins and study their function further.*

**Keywords** cDNA cloning; RT-PCR; *EIF1* genomic; giant panda (*Ailuropoda melanoleuca*); black bear (*Ursus thibetanus mupinensis*)

Received 21 March 2010; accepted 14 April 2010.

This work is 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), Key Discipline Construction Project in Sichuan Province (SZD0420), Key Discipline of Zoology Construction Project in Sichuan Province (404001), Application Foundation Project in Sichuan Province (2009JY0061), and Youth Fund Project of Educational Committee of Sichuan Province (09ZB088).

Address correspondence to Wan-ru Hou and Yun Tang, College of Life Science, China West Normal University, 44# Yuying Road 637002, Nanchong, China. E-mail: hwr168@yahoo.com.cn; tangyun_502@yahoo.com.cn

## INTRODUCTION

Initiation of protein synthesis in eukaryotes is a highly complex process and a key point in the regulation of gene expression.[1,2] Proper assembly of the mRNA, initiator methionyl-tRNA (Met-tRNAi), and ribosome into a functional complex capable of beginning translation at the correct codon in the mRNA requires the coordinated activities of at least 12 initiation factors (eIFs) as well as the hydrolysis of both GTP and ATP.[3] Eukaryotic translation initiation factor 1 (EIF1) was first purified as a factor stimulating binding of Met-tRNA and mRNA to the ribosome.[4,5] EIF1 is critical for proper start codon recognition. In vitro toe printing analyses have suggested that EIF1 is required for the formation of 43S·mRNA complexes capable of locating AUG codons[6] and, more specifically, that omission of EIF1 from such experiments results in assembly of 43S complexes at both cognate and near-cognate initiation codons.[7] Furthermore, biochemical analysis of the mammalian system showed that EIF1 antagonizes recognition of non-AUG codons during scanning[7] and also restrains the GAP (GTPase activating protein) function of eIF5 at non-AUG codons.[8,9] A genetic screen for suppressor mutations that restore expression of *HIS4* in the absence of a start codon identified several point mutations in EIF1 that result in the ability to initiate translation at UUG codons.[10,11] Taken together, this evidence strongly suggests that EIF1 plays a critical role in regulating the start codon recognition event. Mapping the binding site of EIF1 on the 40S subunit by hydroxyl radical probing suggested that EIF1 is incapable of directly monitoring the interaction between the mRNA and the tRNA.[12] Based on this result, it was hypothesized that EIF1 might be involved in a conformational change that is somehow important in regulating the start codon recognition event.



The giant panda (*A. melanoleuca*) known as a "living fossil" is a rare species currently found only in China. The black bear (*Ursus thibetanus mupinensis*) is listed in the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES), the appendix I species; the National 2 levels of key protections wild animals; and China Red Data Book of Endangered Animals, V species.[13] Studies on the giant panda and the black bear have been concentrated on fields from breeding and propagation, ecology, et al.[14–19] to molecular level.[20–33] EIF1 is therefore an essential and universally conserved translation factor with established functions and the study on the EIF1 already quite rich, but the mechanism of its action is still controversial, which has drawn our interest in the study, especially on the unreported EIF1of giant panda and black bear. We address this question by describing the structure of the protein and providing new data for genetic diversity, parentage, phylogenesis of these two species.

This study was conducted using reverse transcription polymerase chain reaction (RT-PCR) to amplify the cDNA of EIF1 gene from the total RNAs,

which were extracted from the skeleton muscle of the giant panda and the black bear. The genomic sequence of *EIF1* was cloned successfully from the giant panda and the black bear using touchdown-polymerase chain reaction (PCR), respectively. We also analyzed the sequence characteristics of the protein encoded by the cDNA and compared it with those of human and other mammalian species reported. The study is of significance to provide sequence data for the Giant panda and the black bear.

## MATERIALS AND METHODS

### Materials

Skeletal muscle was collected from a dead giant panda at the Wolong Conservation Center of the Giant Panda, Sichuan, China. Skeletal muscle tissue samples of the Asian black bear Sichuan subspecies (*Ursus thibetanus mupinensis*) were obtained from black bears collected from Sichuan Province Traditional Chinese Medicinal Materials Company Dujiangyan Raising Deer Field. The collected skeletal muscle was frozen in liquid nitrogen and then used for DNA and RNA isolation.



### DNA and RNA Isolation

The genomic DNA were isolated from the giant panda and the black bear muscle tissue according to the literature.[28] The DNA obtained was dissolved in TE buffer 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 and then dissolved in RNase-free dd$H_2O$, and kept at −70°C.

DNA and RNA sample quality was checked using Experion (Bio-Rad, China) and quantification was performed spectrophotometrically.

### Primers Design, RT-PCR, Cloning of cDNA Sequence, and Sequencing

The PCR primers were designed by Primer Premier 5.0, based on the mRNA sequence of *EIF1* from *Homo sapiens* (NM_005801.3), *Mus musculus* (NM_011508.3), *Bos Taurus* (NM_001014884.1), and *Rattus norvegicus* (NM_001105837.1). The primer sequences are as follows:

*EIF1*-F: 5′-CGCAGGCCGT TTCCACCGAG-3′
*EIF1*-R: 5′-AGGAAATCCT CACTTAAGCT-3′

First-stranded cDNAs were synthesized using a reverse transcription kit with Oligo dT as the primers followed by PCR amplification according to

the manufacturer's instructions (Promega, Shanghai, China). Reverse transcription reactions were performed in duplicate. Lack of genomic DNA contamination was confirmed by PCR amplification of RNA samples in the absence of cDNA synthesis. After amplification, PCR products were separated by electrophoresis in a 1.5% agarose gel with 1 × TAE buffer, stained with ethidium bromide and visualized under ultraviolet (UV) light. The expected fragments of PCR products were harvested and purified from gel using a DNA harvesting kit (Omega Bio-Tek, USA), and then ligated into pMD19-T vector (TaKaRa Dalian, China) at 16°C for 2 hours. The recombinant molecules were transformed into *E. coli* competent cells (DH5$\alpha$) and then spread on the LB-plate containing 50 $\mu$g/mL ampicillin, 200 mg/mL IPTG, and 20 mg/mL X-gal. 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 *EIF1*

The genomic DNA of the *EIF1* gene were amplified using primers *EIF1*-F and *EIF1*-R by touchdown-PCR with the following conditions: 94°C for 30 seconds, 56°C for 45 seconds, and 72°C for 2 minutes in the first cycle and the anneal temperature deceased 0.2°C per cycle; after 20 cycles conditions changed to 94°C for 30 seconds, 52°C for 45 seconds, and 72°C for 2 minutes for another 20 cycles. The fragment amplified was also purified, ligated into the clone vector and tansformed into the *E. coli* competent cells. Finally, the recombinant fragment was sequenced by Huada Zhongsheng Scientific Corporation.



### Construction of the Expression Vector and Overexpression of Recombinant EIF1

PCR fragments of both the giant panda and the black bear corresponding to the EIF1 polypeptide were amplified from the EIF1 cDNA clone with the same forward primer, 5′- AAGGTCGAC GAATCGTATCGT ATG -3′ (Sac I) and the same reverse primer, 5′- CAC CTCGAG TTA AAA CCC ATG AAC -3′ (XhO I), respectively. Both the PCR were performed at 94°C for 2 minutes; 35 cycles of 30 seconds at 94°C, 45 seconds at 55°C, and 1 minute at 72°C; 7 minutes at 72°C. The amplified PCR product were cut and ligated into corresponding site of pET28a vector (Stratagen). The resulting construct were transformed into *E.coli* BL21(DE3) strain (Novagen) and used for the induction by adding IPTG (isopropyl-b-D-thiogalactopyranoside) at an OD600 of 0.6 and culturing further for 5 hours at 37°C, using the empty vector transformed BL21(DE3) as a control. The recombinant protein samples were induced after 1 and 2 hours, and then separated by SDS-PAGE (27) and stained with Coomassie Brilliant Blue dye.

### Data Analysis

The sequence data were analyzed by GenScan software (http://genes.mit.edu/GENSCAN.html). Homology research of the giant panda *EIF1* compared with the gene sequences of other species ware performed using Blast 2.1 (http://www.ncbi. nlm.nih.gov/blast/). ORF of the DNA sequence was searched using ORF finder software (http://www.ncbi.nlm.nih.gov/gorf/gorf.html). Protein structure of the *EIF1* sequence cloned was deduced using PredictProtein software (http://cubic.bioc. columbia.edu/predictprotein/) and (http://swissmodel.expasy.org/). Multiple sequence alignment was performed by software DNAstar Lasergene and DNAMAN 6.0.

## RESULT

### Analysis of the cDNA of *EIF1* from the Giant Panda and the Black Bear



The cDNA fragments of about 0.4 kp were amplified from the giant panda and the black bear with primers *EIF1*-F and *EIF1*-R. The cDNA sequence is 418 bp long. Blast research showed that the cDNA sequence cloned shares a high degree of homology with the *EIF1* from other mammals, including *Homo sapiens, Mus musculus Rattus norvegicus,* and *Bos taurus*. On the basis of the high level of sequence identity, we concluded that we had cloned the cDNA encoding the giant panda and the black bear EIF1 protein, respectively. The *EIF1* cDNA sequences of the giant panda and black bear were submitted to Genbank (accession numbers: GU295661 and GU295662).

The 418 bp of the giant panda and the black bear *EIF1* sequence contains 38 bp untranslated sequence at the 5′ and 3′ of the coding region. An open reading frame (ORF) of 342 bp encoding 113 amino acids was found in this cDNA (Figures 1 and 2).

### Analysis of the Genomic Sequences of *EIF1* from the Giant Panda and the Black Bear

The DNA fragments of about 2000 bp were amplified with primers *EIF1*-F and *EIF1*-R (Figure 3). The length of the DNA fragment cloned from the giant pandais 1909 bp. The length of the DNA fragment cloned from the black bear is1897 bp. They were found to possess four exons and three introns. Comparison between their cDNA sequence and their DNA fragment sequence of the *EIF1* amplified from the giant panda and the black bear were performed by DNAMAN version 6.0. The result indicated their cDNA sequences are in full accord with four fragments in their DNA fragment, suggesting that these DNA fragments amplified are the genomic sequence of the *EIF1* from the giant panda and the black bear**,** respectively. The genomic

```
1    CGCAGGCCGTTTCCACCGAGGAAAAGGAATCGTATCGT ATG TCC GCT ATC CAG AAC
                                             M  S  A  I  Q  N
63   CTC CAC TCT TTC GAC CCC TTT GCT GAT GCA AGT AAG GGT GAT GAT CTG CTT CCT
      L  H  S  F  D  P  F  A  D  A  S  K  G  D  D  L  L  P
111  GCT GGC ACT GAG GAT TAT ATC CAT ATA AGA ATT CAA CAG AGA AAC GGC AGG
      A  G  T  E  D  Y  I  H  I  R  I  Q  Q  R  N  G  R
162  AAG ACC CTT ACT ACT GTC CAA GGG ATC GCT GAT GAT TAC GAT AAA AAG AAA
      K  T  L  T  T  V  Q  G  I  A  D  D  Y  D  K  K  K
213  CTT GTG AAG GCG TCT AAG AAG AAA TTT GCC TGC AAT GGT ACT GTA  ATT GAG
      L  V  K  A  S  K  K  K  F  A  C  N  G  T  V  I  E
264  CAC CCA GAA TAT GGA GAA GTA ATT CAG CTA CAG GGT GAC CAG CGC AAG AAC
      H  P  E  Y  G  E  V  I  Q  L  Q  G  D  Q  R  K  N
315  ATA TGC CAG TTC CTG GTA GAG ATT GGA CTG GCT AAG GAC GAC CAG CTG AAG
      I  C  Q  F  L  V  E  I  G  L  A  K  D  D  Q  L  K
366  GTT CAT GGG TTT TAA  GTGCTTTTGGCTCACTGAAGCTTAAGTGAGGATTTCCT
      V  H  G  F  *
```

**FIGURE 1** Nucleotide sequence of cDNA encoding the giant panda *EIF1* and the amino acid sequence deduced from its ORF (*the stop codon).



```
1    CGCAGGCCGTTTCCACCGAGGAAAAGGAATCGTATCGT ATG TCC GCT ATC CAG AAC
                                             M  S  A  I  Q  N
63   CTC CAC TCT TTC GAC CCC TTT GCT GAT GCA AGT AAG GGT GAT GAT CTG CTT CCT
      L  H  S  F  D  P  F  A  D  A  S  K  G  D  D  L  L  P
111  GCT GGC ACT GAG GAT TAT ATC CAT ATA AGA ATT CAA CAG AGA AAC GGC AGG
      A  G  T  E  D  Y  I  H  I  R  I  Q  Q  R  N  G  R
162  AAG ACC CTT ACC ACT GTC CAA GGG ATC GCT GAT GAT TAC GAT AAA AAG AAA
      K  T  L  T  T  V  Q  G  I  A  D  D  Y  D  K  K  K
213  CTA GTG AAG GCG TTT AAGACG ACA TTT GCC TGC AAT GGT ACT GTA ATT GAG
      L  V  K  A  F  K  T  T  F  A  C  N  G  T  V  I  E
264  CAT CCA GAA TAT GGA GAA GTG ATT CAG CTA CAG GGT GAC CAG CGC AAG AAC
      H  P  E  Y  G  E  V  I  Q  L  Q  G  D  Q  R  K  N
315  ATA TGC CAG TTCCTG ATA GAG ATT GGA CTG GCT AAG GAC GAC CAG CTG AAG
      I  C  Q  F  L  I  E  I  G  L  A  K  D  D  Q  L  K
366  GTT CAT GGG TTT TAA  GTGCTTTTGGCTCACTGAAGCTTAAGTGAGGATTTCCT
      V  H  G  F  *
```

**FIGURE 2** Nucleotide sequence of cDNA encoding the black bear *EIF1* and the amino acid sequence deduced from its ORF (*the stop codon).

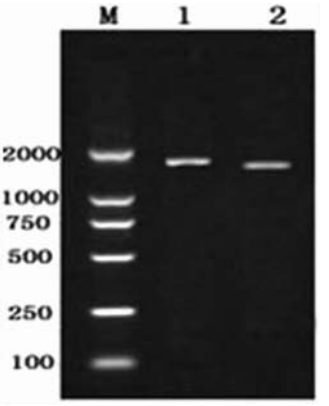


**FIGURE 3** PCR products of complete genomic sequence of *EIF1* from giant panda and the black bear (M: molecular marker DL2000; 1: amplified *EIF1* genomic of the giant pand; 2: amplified *EIF1* genomic of the black bear).



sequence of the *EIF1* of the giant panda has been submitted to Genbank (accession number: GU295663). The genomic sequence of the *EIF1* of the black bear has been submitted to Genbank (accession number: GU295664).

## Comparison of Nucleotide Sequences, Molecular Weight and pI of *EIF1* among six Mammal Species

Alignment analysis of the cDNA sequences of *EIF1* and the deduced amino acid sequence among the giant panda, the black bear and other

```
Pa  MSAIQNLHSFDPFADASKGDDLLPAGTEDYIHIRIQQRNGRKTLTTVQGIADDYDKKKLV 60
Hx  MSAIQNLHSFDPFADASKGDDLLPAGTEDYIHIRIQQRNGRKTLTTVQGIADDYDKKKLV 60
Ho  MSAIQNLHSFDPFADASKGDDLLPAGTEDYIHIRIQQRNGRKTLTTVQGIADDYDKKKLV 60
Mu  MSAIQNLHSFDPFADASKGDDLLPAGTEDYIHIRIQQRNGRKTLTTVQGIADDYDKKKLV 60
Ra  MSAIQNLHSFDPFADASKGDDLLPAGTEDYIHIRIQQRNGRKTLTTVQGIADDYDKKKLV 60
Bo  MSAIQNLHSFDPFADASKGDDLLPAGTEDYIHIRIQQRNGRKTLTTVQGIADDYDKKKLV 60

Pa  KASKKKFACNGTVIEHPEYGEVIQLQGDQRKNICQFLVEIGLAKDDQLKVHGF 113
Hx  KAFKTTFACNGTVIEHPEYGEVIQLQGDQRKNICQFLIEIGLAKDDQLKVHGF 113
Ho  KAFKKKFACNGTVIEHPEYGEVIQLQGDQRKNICQFLVEIGLAKDDQLKVHGF 113
Mu  KAFKKKFACNGTVIEHPEYGEVIQLQGDQRKNICQFLIEIGLAKDDQLKVHGF 113
Ra  KAFKKKFACNGTVIEHPEYGEVIQLQGDQRKNICQFLIEIGLAKDDQLKVHGF 113
Bo  KAFKKKFACNGTVIEHPEYGEVIQLQGDQRKNICQFLVEIGLAKDDQLKVHGF 113
```

**FIGURE 4** Analyzed the functional sites of the amino acid sequence encoded by *EIF1* gene of six mammal species (*Pa: A. melanoleuca; Hx: Ursus thibetanus mupinensis; Ho: H. Sapiens; Mu: M. Musculus; Ra: R. Norvegicus; Bo: B.Taurus).* ___: N-glycosylation site; □: Casein kinase II phosphorylation site; ═══: Amidation site; ------: Protein kinase C phosphorylation site; ■: polymorphic sites.

**TABLE 1** Comparison of nucleotide and amino acid sequences among 6 mammal species

| | *A. melanoleuca* | *Ursus thibetanus mupinensis* | *H.sapiens* | *M. musculus* | *R. norvegicus* | *B. Taurus* |
|---|---|---|---|---|---|---|
| *A. melanoleuca* | | 97.7% | 97.1% | 97.7% | 98.0% | 99.1% |
| *Ursus thibetanus mupinensis* | 95.6% | | 96.5% | 98.2% | 98.2% | 98.5% |
| *H. sapiens* | 99.1% | 96.5% | | 97.7% | 98.0% | 98.0% |
| *M. musculus* | 98.2% | 97.3% | 99.1% | | 99.4% | 98.5% |
| *R. norvegicus* | 98.2% | 97.3% | 99.1% | 100% | | 98.8% |
| *B. taurus* | 99.1% | 96.5% | 100% | 99.1% | 99.1% | |

*Note*: The homology matrix of EIF1 encoding sequence is above the diagonal, the homology matrix of protein sequence is below the diagonal.

mammals reported including *H. sapiens, M.musculus, R. norvegicus,* and *B. Taurus,* was performed by software DNAstar Lasergene. This analysis indicated that both the nucleotide sequence and the deduced amino acid sequence are highly conserved. The giant panda shares the highest homology for nucleotide sequence with *Bos taurus* (99.1%); and the same highest homology for amino acid sequence from *H. sapiens* and *B. Taurus* (Table 1). The black bear shares the highest homology for nucleotide sequence with *Bos taurus* (98.5%); and the highest homology for amino acid sequence from *M. musculus* and *R. norvegicus* (97.3%) (Table 1).

Physical and chemical analysis revealed that the molecular weight of the putative EIF1 protein for the giant panda is 12.67239 kDa and its theoretical isoelectric point (pI) is 6.90, the molecular weight of the putative EIF1 protein for the black bear is 12.65836 kDa and its theoretical isoelectric point (pI) is 6.04. Physical and chemical analysis shows that the molecular weight and that the theoretical pI of the putative EIF1 protein among the six mammalians is very close (Table 2).



## Prediction and Analysis of Protein Functional Sites in *EIF1* Protein

EIF1 is necessary for scanning and is involved in initiation site selection. It promotes the assembly of 48S ribosomal complexes at the authentic initiation codon of a conventional capped mRNA.Topology prediction shows there are one N-glycosylation site site, two Casein kinase II phosphorylation

**TABLE 2** Molecular weight and pI of EIF1 of six mammals species

| | *A. melanoleuca* | *Ursus thibetanus mupinensis* | *H. sapiens* | *M. musculus* | *R. norvegicus* | *B. taurus* |
|---|---|---|---|---|---|---|
| Molecular weight(kDa) | 12672.39 | 12658.36 | 12732.49 | 12746.52 | 12746.52 | 12732.49 |
| pI | 6.90 | 6.04 | 6.90 | 6.90 | 6.90 | 6.90 |

sites, and an amidation site in the *EIF1* protein of the giant panda and the black bear. But there is a protein kinase C phosphorylation site in panda's *EIF1* protein (Figure 4). Alignment analysis of *EIF1* among those protein revealed that the functional sites are entirely identical in *EIF1* proteins of these mammalians. The structure and function of the *EIF1* protein of the giant panda and the black bear are highly conserved.

## Comparison of *EIF1* Genomic among 6 Mammal Species

The genomic sequence of *EIF1*of the giant panda is1909 bp in size. The genomic sequence of *EIF1*of the black bear is1897 bp in size. A comparison of the nucleotide sequences of the genomic and cDNA sequences indicated that their genomic suquence of *EIF1* possesses four exons and three introns, which is also supported by restriction mapping of the genomic and cDNA sequences. Compared with some mammals incluing *Homo sapiens* (NC_000017.10), *Mus musculus* (NC_000077.5), *Rattus norvegicus* (NC_005195.2), and *Bos Taurus* (NC_007317.3), the four exons, which comprise the cDNA sequence of *EIF1* gene after RNA splicing, is highly conserved and remain essentially the same. The restriction sites in the exons are the same in both the cDNA and the gennomic sequences. On the contrary, the genomic, the introns, the 5′-untranslated sequence and the 3′-untranslated sequence are different in length (see Table 3). The variations in lengths of the introns determine the lengths of the *EIF1* genes.



## Overexpression of the *EIF1* Gene in *E. coli*

Next, we intended to overexpress the giant panda and the black bear *EIF1* genes in *E. coli* using pET28a plasmids carrying strong promoter and terminator sequences derived from phage T7. For this purpose, both the *EIF1* genes were amplified individually by PCR and cloned in a pET28a plasmid, resulting in two genes fusion coding for two proteins bearing a His-tag extension at the N terminus. Expressions were tested by SDS-PAGE analysis of protein extracts from recombinant *E. coli* strains BL21 (Figure 5). Data showed that the both EIF1 fusion proteins with the N-terminally His-tagged form gave rise to the accumulation of two expected 19 kDa polypeptide that formed inclusion bodies. Apparently, the recombinant proteins were expressed after half an hour of induction and then after 2 hours reached the highest level. These results suggested that the proteins are active and they are just the proteins encoded by the *EIF1* from the giant panda and the black bear, respectively. The expression products obtained could be used to purify the proteins and study their function further.

**TABLE 3** Comparison of *EIF1* genomic among six mammal species

| | Length of genomic (bp) | exons exons | introns introns | Length of 5′-untranslated sequence (bp) | Length of 3′-untranslated sequence (bp) | Join sites in the CDS | GenBank accession Numbers |
|---|---|---|---|---|---|---|---|
| *Ailuropoda melanoleuca* | 1909 | 4 | 3 | 38 | 38 | 39–68, 772–936, 1084–1185, 1827–1871 | GU295663 |
| *Ursus thibetanus mupinensis* | 1897 | 4 | 3 | 38 | 38 | 39–68, 789–955, 1102–1201, 1815–1859 | GU295664 |
| *Homo sapiens* | 2772 | 4 | 3 | 164 | 820 | 165..195, 904..1067, 1214..1315, 1908..1952 | NC_000017 |
| *Mus musculus* | 2103 | 4 | 3 | 136 | 735 | 137..167, 673..836, 955..1056, 1324..1368 | NC_000077 |
| *Rattus norvegicus* | 2169 | 4 | 3 | 245 | 564 | 246..276, 873..1036, 1151..1252, 1561..1605 | NC_005109 |
| *Bos taurus* | 2558 | 4 | 3 | 133 | 804 | 134..164, 869..1032, 1177..1278, 1710..1754 | NC_007317 |

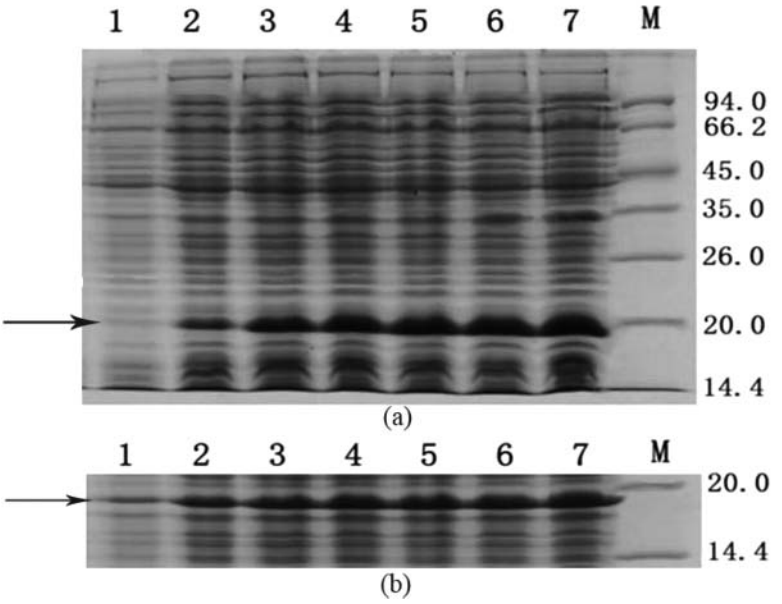


**FIGURE 5** Protein extracted from recombinant *E. coli* strains were analyzed by SDS-PAGE gel stained with Commassie blue R 250. Numbers on right shows the molecular weight and the arrow indicates the recombinant protein bands induced by IPTG with 0, 0.5, 1, 1.5, 2, 3, and 4 hours (lanes 1–7), respectively (M: molecular marker a:giant panda; b: black bear).



## DISCUSSION

Alignment analysis of *EIF1* among the giant panda and the black bear and those of the 6 mammal species, indicated that both the nucleotide sequence and the deduced amino acid sequence are highly conserved. There is not any deletion and insertion of nucleotide and amino acid residue. There are four polymorphic sites in these deduced amino acid sequences of EIF1 proteins (63, 65, 66, and 98; Figure 4). These polymorphic sites are located irregularly in the amino acid sequences all of which result from the transversion or transition of the corresponding codons. It is noteworthy that the Ser-63, Lys-65 of giant panda EIF1 has led to the emergence of the protein kinase C phosphorylation site, while Thr-65 of black bear EIF1 did not lead to the emergence and deletion of any site.

Although the functional sites are not conserved, secondary and tertiary structure analysis shows that residues 29–113 of these sequences form the same structure: a tightly folded domain with two $\alpha$-helices on one side of a five- stranded parallel and antiparallel $\beta$-sheet (Figure 6). No other protein domain is known to have an identical fold. However, structures with a general similarity to EIF1, namely a $\beta$-sheet with $\alpha$-helices on one side, are found in three classes of small protein domains: ribosomal protein S6

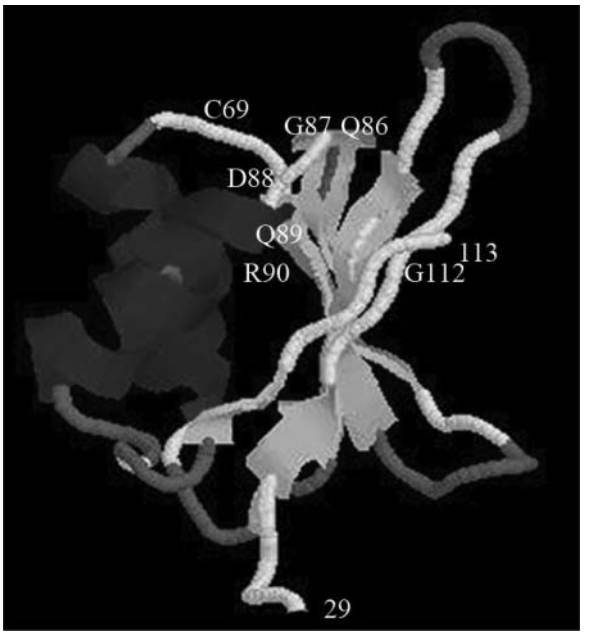


**FIGURE 6** Structure of the folded region of EIF1.



and RNP RNA-binding domains;[34,35] ribosomal protein S3 and KH RN-Abinding domains,[36,37] and double-stranded RNA-binding domains, which have a similar fold to ribosomal protein S5.[38–40] EIF1 differs from all of these structures in the number and order of secondary structure elements in the sequence and the arrangement of the strands in the $\beta$-sheet. Only one feature of the topology is common: the $\beta\alpha\beta\beta$ segment with $\beta$-strand connections[41] formed by the second, third and fourth $\beta$-strands and the first $\alpha$-helix of EIF1. The same topology is found in the first three $\beta$-strands and first $\alpha$-helix of the RNP and S6 proteins and in the three $\beta$-strands and first $\alpha$-helix of the KH and S3 proteins. This may reflect an ancient evolutionary relationship between these proteins. Although this part of the RNP domains contains much of the RNA-binding site, which is located on the surface of the $\beta$-sheet, there is no evidence for a binding site in the equivalent region of EIF1 or the KH and S3 proteins. The N-terminal first 28 residues of the protein have no folded structure. The lack of structure seems to be reflected in the evolution of the sequence: alignment of EIF1 homologs from eukaryotes, archaea and bacteria reveals much lower sequence conservation at the N-terminus than in the folded region.[42]

In addition, it was reported that mutations in yeast EIF1 implicate this protein in maintaining the accuracy of initiation site selection. The *sui1* mutations D88Y, D88G and Q89P allow initiation at non-AUG codons,[31] while the *mof2* mutation G112R causes an increase in programmed ribosomal

frameshifting.[32] All three mutated residues are conserved among the 6 Mammal Species EIF1 (Figure 6). In the structure of EIF1, D88, Q89, and G112 are found close together on the surface of the protein. Furthermore, the same region includes the side chains of C69, Q86, G87, and R90, residues that are almost perfectly conserved among EIF1 homologs. These data suggest that this area of the surface is directly involved in the initiation site selection function of EIF1, most likely as a binding site for another molecule (Figure 6).

Of the 26 fully charged residues on the surface of the folded domain, 23 are grouped into clusters of residues with the same charge: three clusters of positively charged residues and two of negatively charged residues. One of these clusters is particularly striking, comprising five lysine residues on the surface of the first $\alpha$-helix (residues 56–58, 61, and 64). These residues are well conserved among EIF1 homologs. This region is located some distance away from the mutation sites and is, therefore, a possible second binding site, with the positive charges making it suitable for interacting with the phosphate backbone of an RNA molecule.

In summary, the cDNA and the complete coding sequence of *EIF1* gene has been cloned and the *EIF1* cDNA is expressed efficiently in prokaryotic organism using pET28a plasmids. The gained fusion protein is in accordance with the expected 19 kDa polypeptide (Figure 5). These results suggest that the protein is active and it is just the protein encoded by the *EIF1* from the giant panda and the black bear. The question of more complex molecular mechanism and genetic polymorphism will be addressed in future work



## REFERENCES


1. Kapp, L.D.; Lorsch, J.R. The molecular mechanics of eukaryotic translation. *Annu. Rev. Biochem.* **2004**, 73, 657–704.
2. Merrick, W.C.; Hershey, J.W.B. The pathway and mechanism of eukaryotic protein synthesis. In *Translational control* (ed. J.W.B. Hershey, M.B. Matthews, and N. Sonenberg), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1996, pp. 31–69.
3. Pestova, T.V.; Hellen, C.U.T.; Shatsky, I.N. Canonical eukaryotic initiation factors determine initiation of translation by internal ribosomal entry. *Mol. Cell. Biol.* **1996**, 16, 6859–6869.
4. Schreier, M.H.; Erni, B.; Staehelin, T. Initiation of mammalian protein synthesis. I. Purification and characterization of seven initiation factors. *J. Mol. Biol.;* **1997**, 116, 727–753.
5. Trachsel, H.; Erni, B.; Schreier, M.H.; Staehelin, T. Initiation of mammalian protein synthesis. II. The assembly of the initiation complex with purified initiation factors. *J. Mol. Biol.* **1977**, 116, 755–768.
6. Pestova, T.V.; Borukhov, S.I.; Hellen, C.U.T. Eukaryotic ribosomes require initiation factors 1 and 1A to locate initiation codons. *Nature* **1998**, 394, 854–859.
7. Pestova, T.V.; Kolupaeva, V.G. The roles of individual eukaryotic translation initiation factors in ribosomal scanning and initiation codon selection. *Genes Dev.* **2002**, 16, 2906–2922.
8. Algire, M.A.; Maag, D.; Lorsch, J.R. Release from eIF2, not GTP hydrolysis, is the step controlled by start-site selection during eukaryotic translation initiation. *Mol. Cell.* **2005**, 20, 251–262.
9. Unbehaun, A.; Borukhov, S.I.; Hellen, C.U.; Pestova, T.V. Release of initiation factors from 48S complexes during ribosomal subunit joining and the link between establishment of codon–anticodon base-pairing and hydrolysis of eIF2-bound GTP. *Genes Genes* **2004**, 18, 3078–3093.

10. Cui, Y.; Dinman, J.D.; Kinzy, T.G.; Peltz, S.W. The Mof2/Sui1 protein is a generalmonitor of translational accuracy. *Mol. Cell. Biol.* **1998**, 18, 1506–1516.
11. Yoon, H.J.; Donahue, T.F. The suil suppressor locus in Saccharomyces cerevisiae encodes a translation factor that functions during tRNA(iMet) recognition of the start codon. *Mol. Cell. Biol.* **1992**, 12, 248–260.
12. Lomakin, I.B.; Kolupaeva, V.G.; Marintchev, A.; Wagner, G.; Pestova, T.V. Position of eukaryotic initiation factor EIF1 on the 40S ribosomal subunit determined by directed hydroxyl radical probing. *Genes Dev.* **2003**, 17, 2786–2797.
13. Wang S. *China Red Data Book of Endangered Animals: Mammalia.* Beijing, Science Press, 1998
14. Montali R.J.; Causes of neonatal mortality in giant panda. *Tokyo Zoological Park Society* **1990**, 83–94.
15. Lu, Z.; Johnson, W.E.; Menotti-Raymond, M.; Yuhki, N.; Martenson, J.S.; Mainka, S.; Huang, S.Q.; Zheng, Z.; Li, G.; Pan, W.; Mao, X.; O'Brien, S.J.; Patterns of genetic diversity in remaining giant panda populations. *Conserv. Biol.* **2001**, 15, 1596–1607.
16. Xie, Z.; Gipps, J. *The 2003 international studbook for giant panda (*Ailuropoda *melanoleuca).* Beijing, China Association of Zoological Gardens, 2003.
17. Swaisgood, R.R.; Lindburg, D.G.; Zhou, X. Giant pandas discriminate individual differences in conspecific scent. *Animal Behavior*, **1999**, 57, 1045–1053.
18. Swaisgood, R.R.; Lindburg, D.G.; Zhang, H. Discrimination of oestrous status in giant pandas (*Ailuropeda melanoleuca*) via chemical cues in urine. *Journal of Zoology*, **2002**, 257, 381–386.
19. Swaisgood, R.R. Chemical communication in giant pandas. In Giant pandas: biology and conservation (eds D. G. Lindburg & K. Baragona), 2004, pp. 106–120.
20. Liao, M.J.; Zhu, M.Y.; Zhang, Z.H.; Zhang, A.J. cDNA cloning of growth hormone from giant panda (Ailuropoda melanoleuca) and its expression in Escherichia coli. *Comp Biochem Phys B.* **2003**, 135, 109–116.
21. Hou, W.R.; Luo, X.Y.; Du, Y.J.; Chen, Y.; Wu, X.; Peng, Z.S.; Yang, J.; Zhou, C.Q. cDNA Cloning and Sequences analysis of RPS15 from the giant panda. *Recent Patent on DNA Sequence*, **2007**, 2(2), 16–19.
22. Du, Y.J.; Hou, W.R.; Peng, Z.S.; Zhou, C.Q. cDNA Cloning and Sequences Analysis of Acidic Ribosomal Phosphoprotein P1 (RPLP1) from giant panda. *Acta Theriologica Sinica.* **2008**, 28(1), 75–80.
23. Hou, W.R.; Sun, G.L.; Chen, Y.; Wu, X.; Peng, Z.S.; Zhou, C.Q. Molecular cloning of ribosomal protein L26 (RPL26) cDNA from Ailuropoda melanoleuca and its potential value in phylogenetic study. *Biochem Syst Ecol*, **2008**, 36, 194–200.
24. Peng, R.; Zeng, B.; Meng, X.; Yue, B.; Zhang, Z.; Zou, F. The complete mitochondrial genome and phylogenetic analysis of the giant panda (Ailuropoda melanoleuca). *Gene* **2007**, 397, 76–83.
25. Du, Y.J.; Luo, X.Y.; Hao, Y.Z.; Zhang, T.; Hou, W.R. Cloning and Overexpression of Acidic Ribosomal Phosphoprotein P1 Gene (RPLP1) from the giant panda. *Inter. J. Bio. Sci.* **2007**, 3(7), 428–433.
26. Hou, W.R.; Chen, Y.; Peng, Z.S.; Wu, X.; Tang, Z.X. cDNA cloning and sequences analysis of ubiquinol-cytochrome c reductase complex ubiquinone-binding protein (QP-C) from giant panda. *Acta. Theriologica. Sinica.* **2007**, 27(2), 190–194.
27. Hou, W.R.; Du, Y.J.; Chen, Y.; Wu, X.; Peng, Z.S.; Yang, J.; Zhou, C.Q. Nucleotide Sequence of cDNA Encoding the Mitochondrial Precursor Protein of the ATPase Inhibitor from the giant panda (Ailuropoda melanoleuca). *DNA. Cell. Bio.* **2007**, 26(11), 799–802.
28. Hou, Y.L.; Hou, W.R.; Ren, Z.L.; Hao, Y.Z.; Zhang, T. cDNA, genomic sequence and overexpression of crystallin alpha-B Gene (CRYAB) of the giant panda. *Inter. J. Bio. Sci.* **2008**, 4, 415–421.
29. Hou, Y.L.; Du, Y.J.; Hou, W.R.; Zhou, C.Q.; Hao, Y.Z.; Zhang, T. Cloning and sequence analysis of translocase of inner mitochondrial membrane 10 homolog (yeast) gene (*TIMM10*) from the giant panda. *J. Cell. Anima l. Bio.* **2009**, 3(1), 9–14.
30. Hou, Y.L.; Hou, W.R.; Ren, Z.L.; Hao, Y.Z.; Zhang, T. cDNA Cloning and Overexpression of Ribosomal Protein S19 Gene (RPS19) from the giant panda. *DNA. Cell. Bio.* **2009**, 28(1), 41–47.
31. Jennie, P.M.; Alison, M.; Rong, H.L. Activins, inhibins, and follistatins: further thoughts on a growing family of regulator. *Biol. Med.* **1992**, 201, 1–15.
32. Liao, M.J.; Zhu, M.Y.; Zhang, Z.H.; Zhang, A.J. Cloning and sequence analysis of FSH and LH in the giant panda (Ailuropoda melanoleuca). *Anim. Reprod. Sci.* **2003**, 77, 107–116.
33. Wu, Z.A.; Liu, W.X.; Murphy, C.; Gall, J. Satellite DNA sequence from genomic DNA of the giant panda. *Nucleic. Acids. Res.* **1990**, 18(4), 1054.
34. Lindahl, M. Crystal structure of the ribosomal protein S6 from *Thermus thermophilus*. *EMBO J.* **1994**, 13, 1249–1254.

35. Allain, F.H.; Howe, P.W.; Neuhaus, D.; Varani, G. Structural basis of the RNA binding specificity of human U1A protein. *EMBO J.* **1997**, 16, 5764–5772.
36. Castiglione Morelli, M.A.; Stier, G.; Gibson, T.J.; Joseph, C.; Musco, G.; Pastore, A.; Trave′,G. The KH module has an ab fold. *FEBS Lett.;* **1995**, 358, 193–198.
37. Musco, G.; Stier, G.; Joseph, C.; Castiglione Morelli, M.A.; Nilges, M.; Gibson, T.J.; Pastore, A. Three dimensional structure and stability of the KH domain: molecular insights into the fragile X syndrome. *Cell*, **1996**, 85, 237–245.
38. Ramakrishnan, V.; White, S.W. The structure of ribosomal protein S5 reveals sites of interaction with 16S rRNA. *Nature*, **1992**, 358, 768–771.
39. Ryter, J.M.; Schultz, S.C. Molecular basis of double stranded RNA–protein interactions: structure of a dsRNA binding domain complexed with dsRNA. *EMBO J.* **1998**, 17, 7505–7513.
40. Nanduri, S.; Carpick, B.W.; Yang, Y.; Williams, B.R.; Qin, J. Structure of the double stranded RNA binding domain of the protein kinase PKR reveals the molecular basis of its dsRNA mediated activation. *EMBO J.;* **1998**, 17, 5458–5465.
41. Richardson, J.S. The anatomy and taxonomy of protein structure. *Adv. Protein Chem.* **1981**, 34, 167–339.
42. Kyrpides, N.C.; Woese, C.R. Universally conserved translation initiation factors. *Proc. Natl Acad. Sci. USA*, **1998**, 95, 224–228.