Cloning of cDNA for the γ -subunit of mammalian translation initiation factor 2B, the guanine nucleotide-exchange factor for eukaryotic initiation factor 2

Nigel T. PRICE*, Scot R. KIMBALL†, Leonard S. JEFFERSON† and Christopher G. PROUD*‡

*Department of Biochemistry, University of Bristol, Bristol BS8 1TD, U.K., and †Pennsylvania State University, College of Medicine, Hershey, PA 17033, U.S.A.

Peptide sequence data were obtained from rabbit protein synthesis initiation factor subunit eIF2B γ . Searching the database of expressed sequence tags (dbEST) revealed nucleotide sequences potentially encoding human eIF2B γ that contained peptides corresponding to those from the rabbit subunit. PCR primers were derived from these sequences and used to generate a probe. This was used to screen a rat skeletal muscle cDNA library, and a clone encoding rat eIF2B γ was isolated. This cDNA gave a product in coupled transcription/translation that co-migrated with the γ -subunit of purified eIF2B under SDS/PAGE. The sequence of this rat eIF2B γ cDNA is reported. The protein sequence shows homology with that of yeast eIF2B γ (the GCD1

gene product). We have also identified an open reading frame from the *Caenorhabditis elegans* genome project that probably encodes the γ -subunit of *C. elegans* eIF2B. All these sequences show similarity to nucleotidyl- and acyltransferases, as previously reported for GCD1 [Koonin (1995) Protein Sci. 4, 1608–1617], and contain conserved motifs potentially involved in nucleotide binding. They also contain 'I-patch' motifs: isoleucine-rich hexamer repeats that have been associated with the binding of acyl groups in bacterial acyltransferases. The roles of these motifs are discussed in relation to the known properties of eIF2B.

INTRODUCTION

The protein synthesis initiation factor eIF2B performs an essential role in mammalian protein synthesis by recycling another initiation factor, eIF2, responsible for delivering the charged initiator Met-tRNA^{Met} to the ribosome as a ternary complex of eIF2, Met-tRNA^{Met} and GTP. After a successful round of protein synthesis initiation, the eIF2 is released as an inactive complex with GDP. The guanine nucleotide-exchange factor eIF2B is responsible for recycling eIF2 to its active GTP-bound state. This is a key regulatory step in eukaryotic protein synthesis (reviewed in [1]). eIF2B has five different subunits. As a heteropentameric protein, eIF2B is more complex than all other known guanine nucleotide-exchange factors. The reasons for this complexity are as yet unclear. eIF2B might have other roles in translation, besides simply recycling eIF2, or it might be involved in other cellular processes. Alternatively the number of subunits in eIF2B might reflect a variety of regulatory inputs: the protein's guanine nucleotide exchange activity is known to be subject to regulation via phosphorylation of eIF2B or of its substrate eIF2 and by a range of allosteric effectors (reviewed in [1]).

The individual roles of the five subunits of mammalian eIF2B in its known function and its regulation remain unclear [1]. The γ subunit is implicated in binding pyridine dinucleotides. Purified eIF2B has been shown to contain NADPH [2], and photoaffinity-labelling studies with 8-azido-ATP result in labelling of the γ and δ subunits of eIF2B, presumably at the NADPH binding site [3]. NAD(P) [3,4] and ATP [5] are reported to inhibit eIF2B activity. The sequence of eIF2B γ from yeast (the GCD1 gene product) is available [6]; however, it is not known whether adenine or pyridine nucleotides regulate the activity of yeast eIF2B. Although GCD1 does not contain the consensus nucleotide-binding

motif [AG]XXXXGK[ST], it does show similarity to the nucleotidyl and acyl transferase family of proteins [7]. In addition, the protein sequences of eIF2B γ and ϵ from yeast (GCD1 and GCD6) show similarity to one another [8]. The sequences of eIF2B ϵ from mammals and yeast (GCD6) also show sequence similarity to the transferases.

Although cDNA sequences are available for the other four subunits of eIF2B from mammals [9–13] and for all five from yeast (reviewed in [1,14]), this is the first report of the sequence eIF2B γ from a metazoan species. Motifs present in the protein sequence, and its similarity to eIF2B γ from yeast and to eIF2B ϵ from mammals and yeast, are also discussed. We have also identified an open reading frame (ORF) from *Caenorhabditis elegans* that probably encodes the γ subunit of *C. elegans* eIF2B. The product of this ORF had previously been assigned a different role. A sequence that probably encodes *C. elegans* eIF2B ϵ is also available [15].

EXPERIMENTAL

Materials

Materials were obtained as described previously [9,12,16], except where otherwise indicated.

Sequencing of peptides derived from eIF2B γ

Rabbit reticulocyte eIF2B was purified as described [4]. Attempted N-terminal sequencing of the γ subunit was performed as described [9]. N-terminal sequences for several CNBr-derived fragments of eIF2B γ were obtained as described previously [9,12].

Abbreviations used: eIF, eukaryotic initiation factor; EST, expressed sequence tag; dbEST, database of ESTs; ORF, open reading frame.

‡ To whom correspondence should be addressed at the Department of Biosciences, University of Kent at Canterbury, Canterbury, Kent CT2 7NJ, U.K.

The nucleotide sequence of rat eIF2B γ reported will appear in EMBL and GenBank Nucleotide Sequence Databases under the accession number U38253.

PCR amplification of human eIF2B γ

DNA sequences of candidate clones encoding human eIF2B γ were obtained by searching the database of expressed sequence tags (dbEST) [17]. This database was searched with the sequences derived from rabbit eIF2B γ by using the tblastn program [18]. Overlapping sequences were assembled and suitable regions were used to derive PCR primers. By using PCR, a 227 bp fragment of human eIF2B γ was amplified from skeletal muscle cDNA.

cDNA library screening

The 227 bp fragment of human eIF2B γ was purified from an agarose gel and labelled with [α - 32 P]dCTP by random priming (Amersham Multiprime Kit). This probe was purified by gel filtration (Pharmacia Nick Column) and used to screen a rat skeletal muscle cDNA library in Uni-ZAP XR (Stratagene no. 937510). Putative positive clones were isolated by screening approx. 10^6 recombinant phages, and excised as described previously [16].

DNA sequence determination

DNA was sequenced on both strands by using a custom primer walking strategy with Applied Biosystems 373A sequencing gels.

Transcription/translation *in vitro*

Rat eIF2B γ cDNA in pBluescript was expressed by coupled transcription/translation in rabbit reticulocyte lysate (Promega TNT system). Transcription was driven by the T3 promoter and translation products were labelled in the presence of [35 S]methionine. Samples were analysed by SDS/PAGE alongside purified eIF2B, and the resulting gel was subjected to autoradiography.

Isolation of RNA and Northern blotting

Total RNA was prepared from various rat tissues by using TRIzol reagent (GibcoBRL). Poly(A)⁺ RNA was isolated from this total RNA with the polyAtract mRNA isolation system (Promega). The denatured RNA was subjected to electrophoresis on a formaldehyde gel and transferred to Hybond-N membrane (Amersham). The membrane was probed with a 952 bp PCR product (corresponding to nt 318–1269) amplified from the rat eIF2B γ clone. The probe was 32 P-labelled by random priming as described above. After autoradiography the blot was stripped and probed again with a 32 P-labelled probe for β -actin.

DNA sequence analysis

Sequence assembly and alignment was performed with the Seqman and Megalign packages (DNASTar Lasergene Software). Database searches were performed with the BLAST family of programs [18] and BLITZ [19].

RESULTS

Sequencing of eIF2B γ peptides and cloning strategies

The N-terminus of rabbit eIF2B γ was resistant to the Edman degradation. Internal sequence data were therefore obtained by sequencing peptides generated by CNBr fragmentation as described previously [20]. A large number of CNBr peptides were observed on Tricine/SDS gels, some of which were isolated for sequencing by reverse-phase HPLC (Aquapore RP-300 column, Applied Biosystems) [20]. The peptide sequences thus obtained were used to design redundant oligonucleotide primers for use in PCR from rabbit cDNA. This approach has been used suc-

Table 1 EST clones for human eIF2B γ

Sequences for potential human eIF2B γ clones were obtained by searching dbEST with peptide sequence data from rabbit eIF2B γ by using the tblastn program [18]. The clones are listed in descending order of length. The source column indicates the organization producing the sequence data. Abbreviation: HCD, human cDNA database.

Clone identification	Source	Accession no. (5' end)	Accession no. (3' end)	Tissue
71g09	Genethon	Z24924	F00408	Skeletal muscle
c-zqf04	Genethon	Z45650	Z41323	Brain
yc97g06	WashU/Merck	T78051	—	Brain
c-08a05	Genethon	F05355	F01614	Brain
yc89h04	WashU/Merck	T75394	R38705	Brain
yd42b04	WashU/Merck	T82858	T90276	Liver
—	HCD	THC80184	THC109661	—

cessfully before for the β and δ subunits of eIF2B [9,12] and was aided by comparing the sequences of the mammalian peptides with those of the corresponding subunits of the yeast factor, which indicated the correct orientation of the primers and the approximate size of the expected product. However, in the present case this approach was unsuccessful and no correct PCR products were obtained. It is likely that this was in large part due to the low similarity between GCD1 and mammalian eIF2B γ (see below) and also to the fact that the peptide sequence data obtained allowed the design of only highly redundant PCR primers.

Use of human expressed sequence tag (EST) data to derive an eIF2B γ clone

To try to extend the peptide sequences further, the rabbit eIF2B γ peptide sequence data were used to search dbEST [17] with the tblastn algorithm [18]. A potential sequence for the 5' end of a clone for human eIF2B γ was identified. Subsequent searching of databases with this nucleotide sequence revealed the presence of six overlapping sequences derived from six original EST clones from brain, liver and skeletal muscle cDNA libraries (see Table 1). For five of these clones, corresponding sequences for their 3' ends were also retrieved from dbEST. These sequences are listed in Table 1. All of these sequences were assembled, and edited where necessary. Translation of the two resultant contiguous sequences revealed a protein sequence that contained regions corresponding very closely to six of the rabbit CNBr-derived peptides (see Figure 1). Thus it seemed highly probable that these six human clones represented truncated cDNAs encoding human eIF2B γ .

The assembled human ESTs (with the exception of Genethon clone c-08a05) were later found in the human cDNA database (HCD) [21] (level 1 access). These are assigned HCD numbers THC80184 (5' end) and THC109661 (3' end) (where THC stands for tentative human consensus). These clones were tentatively identified as encoding an isologue of a putative pyrophosphorylase from *C. elegans* (see the Discussion section).

The assembled human sequences were used to design exact-match PCR primers, which were used in PCR with human cDNA generated from skeletal muscle total RNA. The product obtained was gel-purified and used to screen a cDNA library.

cDNA library screening

Because only in rat are cDNAs available for all of the other four subunits of mammalian eIF2B, it was decided to screen a rat

```

1                               80
MEFOAVVYVAGGSGRMDLTSSIPKPLLPVGNKPLIWPYLNLLERVGFEEIVVTTQDVQKALCAEFKMKLKPDI VCI PD
                               160
DADMTADSLXYIYPKLTDLVLSDLITDVALHEVVDLFRAYDASLMLMRKGQDSIEPVPGQKGGKKKAVEGRDFIGV
EADMTADSLRHYIYPKLTDLVLGCDLITDVALHEVVDLFRAYDASLMLMRKGQDSETPVPGQKGGKKKTVEGRDFIGV
Rabbit peptide: (MRKGQDGLPVPQK)

161                               240
DSTGKRLLFMANEADLDEELVIGSILQKHPRIRFHTGLVDHLYCKKYIVDFLMENGSIITSIRSELIPYLVRKOFSSA
DSTGKRLLFMANEADLDEELVIGSILQKHPRIRFHTGLVDHLYCKKYIVDFLMENKSLITSIRSELIPYLVRKOFSSA

241                               320
SSOQCTRKEKEGGSXKKXKLSLDIYSFY.....
SSOQ...ROEDKEEDLKKKEPKSLDIYSFIKKONTLTLPYDACWNAFRROKWDLSRQVRCYVHIMKEGLCSRSTLGL

321                               400
.....LIGPETQIGEKSSI KRSVIGSSCLIKDRVTITNCLLMNSVTY
YMEANRQVPKLLSVLCPEESMHPSAQIANKHLIGADSLIGSDTVQGEKSSIKRSVIGSSCVRDRVTITNCLLMNSVTY

401
EEGNSIQGSVINCNAVIEKGADIKDCLIGSGORIEAKAKRYNEVINYNDOLMEI •
EEGSSIHGSVINCNAVIEAGAEIRDCLIGSGORIEAKAKRMNEYINYNDOLMEI •

```

Figure 1 Partial cDNA-derived protein sequence of human eIF2B γ

The human EST-derived DNA sequence (upper sequence; see also Table 1) was assembled and edited (results not shown). The translated sequence is shown, aligned with that of the rat subunit (lower sequence). It should be noted that the two regions of sequence are derived in each case from only one strand of the DNA; however, this sequence is assembled from up to six overlapping ESTs in some regions. Underlining in the rat protein sequence indicates regions corresponding to peptides derived from the rabbit subunit by CNBr cleavage. Gaps in the underlining indicate differences between the rabbit and rat sequence data or uncertainties in the original protein sequence data. The rabbit peptide (shown below the sequence) corresponding to that with double underlining was used to generate a polyclonal antibody (see text).

cDNA library for eIF2B γ . As a result of screening approx. 10⁶ recombinant phages, 18 clones were isolated and excised. On the basis of restriction analysis, these clones were assigned to two groups. Sequence analysis revealed that the first group consisted of sibling clones encoding eIF2B γ . Clones from the second group contained part of the sequence of eIF2B γ in reverse orientation. However, this sequence was truncated at its 3' end; the clone also contained approx. 180 bp that were not derived from eIF2B γ . A clone from the first group was sequenced fully on both DNA strands. The complete sequence of rat eIF2B γ is shown in Figure 2.

Characteristics of the rat eIF2B γ sequence

The rat eIF2B γ sequence is 1553 bp in length. The first ORF (nt 89–1447) contains all the peptide sequences corresponding to those obtained from rabbit eIF2B γ . There is an in-frame termination codon (nt 70–73; UAA) upstream of the proposed initiator AUG. It is also noteworthy that the sequence (M)EFQAVV was not among those derived from CNBr-fragmented eIF2B γ , in agreement with this sequence corresponding to the blocked N-terminus of the subunit. This AUG has a G at the +4 position and a purine (A) at position –3, and is thus in a favourable context for initiation. The predicted ORF encodes a protein of 452 amino acids with a molecular mass of 50440. This is smaller than the observed size of 55700 for the subunit of the purified rat factor [22]. The β , δ and ϵ subunits of eIF2B also migrate more slowly than predicted from their cDNA sequences [8,9,12,22,23]. The predicted and observed pI values for eIF2B γ (6.8 and 6.9, respectively [22]) are in close agreement.

The presence of sequences corresponding to all of the peptide sequence data obtained from eIF2B γ confirms the identity of the clone. A synthetic peptide derived from one of the rabbit sequences was used to generate a polyclonal antibody to eIF2B γ (see Figure 1). This antibody recognizes both rat and rabbit proteins (results not shown), despite differences in the corresponding regions in the sequences of the rat and rabbit polypeptides (see Figure 1).

```

1 GCGGCTACGGCGT 13
CGGGGAGCTGAGGTAAAGCCGGGCTAGTACGTTACCAAGCGGTGCACAGCTGGTAAAGCCAGATTGTGTAAGA 88
ATGGAGTTTCAAGCCGCTAGTTATGGCCGCTGGGCGGATCTCGGATGACAGACTTGCATCTCCAGCATTCGCAAG 163
1 H E F I Q A V V H A V G G S R M T D L T S S I P K
26 CTTCTGCTCCAGTGTGGAACAACCTTTAATTGGTATCCGCTGAACCTGCTGAGCGAGTTGGATTGGAAGAA 238
P L L P V G N K P L I W Y P L N L L E R V G F E E
51 GTCATCGTGGTTACAACCAAGATGTCCAAAGGCTCTGTGTGCAGAGTCAAGATGAATTAAGCCAGATATT 313
V I V V T T K D V O K A L C A E F K H K L K P D I
76 GTGTGCTTCCTGATGAGGACAGACATGGGACTGACAGATTCTACGCCCATATATCCAAACCTTAAGACTGAT 388
V C I P D E A D M G T A D S L R H I Y P K L K T D
101 GTGCTGGCTGGGCTGCGATCTGATACAGATGTTGCTTACAGAGGTTGTGGACCTGTTCAAGCTTACGAT 463
V L V L G C D L I T D V A L H E V I V D L F R A Y D
126 GCATCACTTGCCATGTTAATGAGGAAGGACAGAGACAGACAGACAGTTCAGGTCAGAGGGAAGGAAAAA 538
A S L A M L M R K G O E S T E P V P G Q K G K K K
151 ACAGTGGAGCAGCGTGAATTCATTGGAGTGGACAGCAGCAGGAAAAAGGCTGCTTTATGGCTAATGAAGCAG 613
T V E Q R D F I G V D S T G K L L F H A N E A D
176 TTGGATGAAGAGCTGGTCTTAAGGGATCCATCTGCAGACAGCCTAGGATACATTCAGACAGGCTTGTG 688
L D E E L V I K G S I L O K H P R I H F O T G L V
201 GATGCTCATCTCTGCTGCTGAAGAAATATGTTGGATTCTTATGAAGAAATGAAGTCAATCTTCTATCCGA 763
D A H L Y C L K K Y V V D F L M E N K S I T S I R
226 AGCGAATGATTCGCTAGTAAAGAAAGCTTTCTCAGCTCCTCAGCAGAAAGGACAGGAGACAAAGAA 838
S E L I P Y L V R K O F S S A S S O Q R Q E D K E
251 GAGGATCAAAAGAAAGGAGCCCAAGTCTTTAGATATTTACAGTTTATAAAAAAGACAACTCTGACTTTG 913
E D L K K K E P K S L D I Y S F I K K D N T L T L
276 GCTCCATGATGCTGCTGGAATGCTTTTGAAGAGACAAAGTGGGAAGATTTGTCTAGATCAGAGTTGCTGC 988
A P Y D A C W N A F R R D K W E D L S R S Q V R C
301 TATGCTCCATCATGAGGAAGGACTTGTCTCGAGTAAAGCAGCTGGGACTTACATGGAAGCAACAGACAG 1063
Y V H I M K E G L C S R V S T L G L Y H E A N R Q
326 GTACCAATTTGCTGCTGTTTGTCTGAGGAATCAATGATCCATCAGCCAGATTCGCAACAAACAC 1138
V P K L L S V L C P E E S M I H C P S A Q I A N K H
351 CTGATTGAGCCGACGCTTATCGGCTCAGATACAAAGTTGGAGAGAAGTCACTATTAAACGATCTGCTATT 1213
L I G A D S L I G S D T O V G E K S S I K R S V I
376 GGTCTATCTGCTGATCAGAGACAGAGTAACTGTTACCACTGCTCTCATGACTGTCACGCTGGAAGAA 1288
G S S C V I R D R V I V T N C L L M N S V T V E E
401 GGGAGCAGCATCCAGGAGCGTCTATCTCAACACBCCTGGTGGAGCGGGCGGAGATCAGGAGCTGTTA 1363
G S S I H G S V I C N N A V V E A G A E I R D C L
426 ATTGGAAGTGGTCAAGGATTGAAGCCAAAGCTAAACGATGAACAGGTTGATCTGSSAAATGACAGCTCATG 1438
I G S G O R I E A K A K R M N E V I V G N D Q L M
451 GAGATCTGAATCCTGAACAGTCTGACCCCTCCCTTTTGGCCCAAACTACAGATGTTGGCTGCCACCTGTT 1513
E I • 452
TGACTCTGATTTATTTTCAATAAAGAAGATTTCAGG 1553

```

Figure 2 Nucleotide and deduced amino acid sequence of rat eIF2B γ

The Figure illustrates a number of features of the sequence of rat eIF2B γ . The potential P-loop and magnesium-binding regions of the sequence are boxed (residues 4–30 and 100–117). These boxes correspond to the regions of sequence similarity between GCD1, GCD6, eIF2B ϵ and various transferases in the alignments shown by Koonin [7]. The aspartate residue 107, which is potentially involved in the chelation of Mg²⁺, is underlined [7]. This residue is conserved in the human, yeast and (putative) *C. elegans* eIF2B γ sequences; see Figure 5. Another potential nucleotide-binding domain conforming to the consensus [AG]XXXXGK is present. This sequence (residues 159–165) is also underlined. The groups of boxed residues towards the C-terminus indicate the so-called I-site hexad repeats (see text). It should be noted that the N-terminus of the mature protein is blocked and it is therefore unknown whether it actually starts with the methionine indicated here.

Figure 3 Expression of rat eIF2B γ cDNA in rabbit reticulocyte lysate

Rat eIF2B γ cDNA in pBluescript was expressed by coupled transcription/translation in rabbit reticulocyte lysate (Promega TNT system) in the presence of [³⁵S]methionine. Transcription was driven by the T3 promoter. Samples were removed at various time points as indicated and analysed by SDS/PAGE and autoradiography. The position of migration of the γ subunit of purified eIF2B is shown.

Figure 4 Probing for eIF2B γ on Northern blots of poly(A)⁺ RNA from different rat tissues

Total RNA was prepared from various rat tissues. Poly(A)⁺ RNA was isolated from approx. 250 μ g of each total RNA sample and subjected to electrophoresis in a formaldehyde gel. The RNA was transferred to Hybond N membrane and probed with ³²P-labelled rat eIF2B γ cDNA (A). The membrane was stripped and probed again with β -actin cDNA (B). RNA was isolated from brain, heart, liver, skeletal muscle and testes. The Figure shows the resulting autoradiographs.

The rat eIF2B γ has 88 bp of 5' non-coding sequence and 106 bp of 3' non-coding sequence. The nucleotide sequence of human eIF2B γ derived from EST data has a similar amount of 3' untranslated sequence (110 bp; results not shown).

Expression of rat eIF2B γ *in vitro*

To verify that the clone contained all the coding region for eIF2B γ , coupled transcription/translation *in vitro* was performed. The resulting major [³⁵S]methionine-labelled protein co-migrated with the γ subunit of purified eIF2B (Figure 3).

Quantification of mRNA for eIF2B γ in different tissues

Northern blot analysis of various rat tissues was performed (Figure 4). A single major species was detected for all tissues examined. Its position of migration, intermediate between those of the two forms of actin found with a β -actin probe in skeletal muscle, suggests a size of 1750–1800 nt, indicating that the 1553 bp eIF2B γ clone is near full length.

In all of these tissues, expression of eIF2B γ message was approximately the same, relative to that of the β -actin message. Large amounts of mRNA for eIF2B α , β and δ , and to a smaller extent eIF2B ϵ , have been observed in rat testis relative to other rat tissues [23]. This was not observed for eIF2B γ .

Human eIF2B γ

The partial cDNA-derived protein sequence of human eIF2B γ predicted from assembled and edited EST data is shown in Figure 1 along with the aligned rat sequence. The regions corresponding to CNBr peptides from rabbit eIF2B γ are indicated. It should be noted that although the DNA sequence was derived from several overlapping sequencing runs, these were all performed on only one strand of the DNA: caution should therefore be exercised in using these data for the human sequence. It can, however, be seen that rat and human eIF2B γ show a high degree of amino acid sequence identity.

DISCUSSION

Relationship of the rat eIF2B γ protein sequence to those of other proteins

Searching protein sequence databases with the rat eIF2B γ sequence revealed the expected homology with eIF2B γ (GCD1) from *Saccharomyces cerevisiae*, and with eIF2B ϵ from mammals. [GCD1 and GCD6 (eIF2B γ and eIF2B ϵ from yeast) have previously been reported to show limited sequence identity [8] (see below).] The aligned rat and yeast eIF2B γ protein sequences show 19% similarity (see Figure 5). Rat eIF2B γ was also found to have similarity to nucleotidyl transferases such as NDP-hexose pyrophosphorylases and dTDP-glucose synthase, and this is discussed further below.

To gain an insight into the possible functions of the γ subunit of eIF2B, the protein sequence of rat eIF2B γ was used to search the Blocks database of protein families [24]. Sequence similarity with another nucleotidyl transferase, ADP-glucose pyrophosphorylase, was revealed, both in the potential nucleotide-binding region and at several positions in the C-terminus of rat eIF2B γ . The same C-terminal regions were also found to possess sequence similarity to isoleucine-rich hexapeptide repeats found in bacterial transferases. Exactly the same was found when searching the Blocks database with the sequence of rat eIF2B ϵ . These features have recently been shown to be present in the sequences of yeast eIF2B γ and of eIF2B ϵ from yeast and mammals [7], as discussed further below.

Identification of a putative sequence for *C. elegans* eIF2B γ

Rat eIF2B γ was also found to have strong similarity to the product of a predicted ORF from *C. elegans*. This ORF was predicted to encode a pyrophosphorylase designated PPP-1 and has similarity to glucose-1-phosphate thymidyltransferases [25]. Recently PPP-1 has been included in a family of nucleotidyl-transferases [7]; however, we propose that PPP-1 in fact encodes *C. elegans* eIF2B γ .

The ORF in the sequence of PPP-1 is considerably shorter than that of rat eIF2B γ (lengths 360 and 452 amino acid residues respectively). However, editing of the PPP-1 sequence by removal of a single A in the sequence GAAAT (nt 1074–1078) gives a theoretical protein with an extended C-terminus (new length 405 residues). This altered C-terminal sequence, shown in Figure 5, shows strong similarity to both the yeast and rat eIF2B γ sequences. It seems likely, in view of the high degree of similarity of the 'new' C-terminus to other eIF2B γ sequences, that the extra A, which results in the introduction of a premature termination codon, represents a sequencing artifact. Hereafter we have therefore used the edited PPP-1 sequence.

The DNA-derived protein sequences of eIF2B γ from rat and yeast (GCD1) are shown aligned with PPP-1 in Figure 5. The rat eIF2B γ sequence shows a greater degree of similarity to PPP-1 than it does to that of yeast eIF2B γ (24% and 19% similarity respectively). The yeast eIF2B γ protein sequence is longer than those of *C. elegans* or rat [lengths 405 (edited sequence, see above), 579 and 452 residues respectively] due to insertions in the N-terminus and a much extended C-terminus (see alignment in Figure 5). In particular, the highly acidic C-terminal region of GCD1 is not present in the rat or *C. elegans* sequences. PPP-1 also has a deletion of 50–60 residues (after residue 244) in its sequence, relative to both the rat and yeast sequences.

Sequence similarity between eIF2B γ and eIF2B ϵ

It has previously been reported that the protein sequences of eIF2B γ (GCD1) and eIF2B ϵ (GCD6) from yeast show regions of

Rat	M EFOAVVMAVGG-----GSRMTDLTSS-----IPKPLLPVGNKPLIWIYPLNLL	43
<i>C. elegans</i>	MHEMOGIILLSSG-----GTRMPVLTIRH-----VOKCLLPVGVNPNFLYPLSSL	44
<i>S. cerevisiae</i>	M-SIOAFVFGGKSNLAPFTQDPFPQTQNKDSTAATSGLKLNELVNSALDSTVINEFMQHSTRLPKALLPIGNRPMIEVLDWC	84
Rat	ERVSGFEVIVTT---KDVCKALCA--EFKM-KL-----KPDIVCIPDEADMGTADSRHIIYPKIKTIV	101
<i>C. elegans</i>	LRTGITDKIFVR---EVLQLTNEK--EYKKSLLLEKY-----FAHIEYICVNOE--DFGTADLLKNHHSKITQDA	108
<i>S. cerevisiae</i>	DQADFKELSVYAPVDETELIESGLTSFLSLRQGFELIYKALSNSNHSHLQDKKINF--ISKANSTGESLQKELLPRINGDF	167
Rat	LIVGCDLITVALHEVVDLFRAYASLAMMRKGQESTEVPGQKGGKKTVEQRDFIGVDSTGRRL-----LMANEADLDE	178
<i>C. elegans</i>	LIVSCDFISDASLIPLVDFFRATNSTLVALI-ADTCYNAPAGSKSKKPKAT--DVMAIVESTGQL-----AFLCGDDFDA	182
<i>S. cerevisiae</i>	VILPCDFVITIPPQVLVDQFRNRDNNLAMTIYYKNLDSSTGKQKQAKNNSNLSLFIQKTRTQRGSQYFVNVYSQRDVTKIK	252
Rat	ELVYKGSILQKHPRTHFOTGLVDAHLV---CLKKYVVDLFMENK--SITSIRSELIPYLVRKOFSSASQQRQEDKEELKKKEPK	259
<i>C. elegans</i>	PLVMEKS-LKIFPSIKLTSKYNDCHVY---AIRHKVLLNLSKSK--HSSFKADVPILLIDKQEPDS-----	245
<i>S. cerevisiae</i>	YLQIRSHLLWNYPNLTIVSTKLLNSFIFCFSELCLLKLGPQMSRQASFQDPITGNQQQNPPPTTQDDEDRNHDDDDYKPSAT	337
Rat	SLDIYSFIKKDNTLTARFYDAC---WNAFR---RDKMEDLSRSQVRQYVHIMKEG---LCSRVSTLGLYMEANRVPKLLSVLC	334
<i>C. elegans</i>	SIQPTYFKKKNDLILDFIN-CKSLSKVFRDLSSRSQVQH-SKPREPIGIFILPNETLFI--RANNNAYMDANRFVLKI----	285
<i>S. cerevisiae</i>	SIQPTYFKKKNDLILDFIN-CKSLSKVFRDLSSRSQVQH-SKPREPIGIFILPNETLFI--RANNNAYMDANRFVLKI----	413
Rat	PEESMIHPSADTANKHIGADSLTIGSDTVGEGKSSIKRSVIGSSCVTRDVTVTNCLLMNSVTVEEGSSIHGSVICNNAVVEAGA	419
<i>C. elegans</i>	MEYRGNGKNFNKYTKIAIHESRIEESAETDKDSVIKRSFHSNCRIGEKTKLKESI IAKGVVIGMBASISNSTICDGVETIGENA	369
<i>S. cerevisiae</i>	KSQTMFTKNIDTOSAAI-GADAIIDPKCOISAHSNVMSVLTQANIGSCRVAGSLLPFGVHLGDEVILENCIIGPMAKIGSKC	496
Rat	EIRDCLIGSGORTAKAKRMNEVI-----VANDQLM---EI	452
<i>C. elegans</i>	DVTNCIYAKDQKVPKAKGVONEV-----EDGEDEEWTD	404
<i>S. cerevisiae</i>	KLTCGYEGHYVVEPKNNFKGETILANVHLEDEDEELIYDDSVIAGESEIAEETDSDDRSDESDSSEYTDIEYEDDGLFER	579

Figure 5 Alignment of the protein sequences of eIF2B γ from mammals (rat), yeast and *C. elegans*

The sequences used are those of rat eIF2B γ (present study), *S. cerevisiae* eIF2B γ (the *GCD1* gene product, accession number P09032), and *C. elegans* PPP-1 (accession number P80361). The theoretical ORF PPP-1 has previously been putatively assigned the role of pyrophosphorylase. The likelihood that PPP-1 encodes eIF2B γ is discussed in the text. It should be noted that the PPP-1 sequence has been edited as described in the text, and thus the C-terminus of the protein sequence differs from that currently in the database. Sequences were aligned by using the Clustal algorithm with the Megalign package in Lasergene software from DNASTar. Identical residues are shown in white on black and conserved residues are in boxes. Several regions of the alignment are ambiguous, particularly those containing the repeated I-patch domains (see text).

identity [8], although overall the sequences show only 11% similarity. The mammalian counterparts are more closely related, showing 17% similarity to one another. The aligned rat eIF2B γ and eIF2B ϵ sequences are shown in Figure 6. An ORF that probably encodes *C. elegans* eIF2B ϵ has also been identified recently [17]. This shows only 14% identity to PPP-1.

Sequence similarity between eIF2B subunits and nucleotidyl transferases

Whether PPP-1 really is eIF2B γ from *C. elegans* or is a pyrophosphorylase, it is clear that sequence similarity exists between the eIF2B γ/ϵ sequences and the nucleotidyl transferase family [7]. This sequence similarity extends across three conserved domains: (1) the P-loop (putative phosphate binding site), (2) putative Mg²⁺ binding sites, which are both thought to be involved in nucleotide binding in the nucleotidyl transferases, and (3) the so-called I-patches (see below). It should be noted that some of the PPP-1 I-patch motifs are in regions altered in the edited sequence (see above), and this editing reveals additional hexads not present in the alignments presented in [7], in addition to slightly altering one of the motifs present in that reference.

There is now evidence that eIF2B can be directly activated by sugar phosphates such as fructose 1,6-bisphosphate [26]. The sequence elements conserved between eIF2B and nucleosidyl-transferases might be involved in binding sugar phosphates, because the latter group of enzymes also bind these molecules.

Nucleotide-binding motifs

The newly cloned sequence of mammalian eIF2B γ also shows similarity to the nucleotide/nucleotide-sugar binding regions of transferase enzymes. This includes the P-loop region, thought to

bind phosphate, and the potential Mg²⁺ binding region, including the invariant aspartate residue (residue 107 in rat eIF2B γ ; see Figures 1, 2, 3 and 6). Figure 6 shows that these regions of rat eIF2B γ and ϵ are well conserved. In the P-loop region, the optimal alignments for the N-terminal sequences of eIF2B γ (Figure 5) are ambiguous, and those shown differ slightly from those of Koonin [7]. However, all these sequences contain features that suggest that the γ and ϵ subunits of the guanine nucleotide-exchange factor eIF2B might be involved in nucleotide binding.

Mammalian eIF2B is known to interact with NAD(P)⁺ and NAD(P)H and with ATP, all of which are allosteric modifiers of the factor's GTP/GDP exchange activity. The ability of ATP analogues to label the γ and δ subunits of eIF2B might reflect roles for these subunits in binding these nucleotides. Because the δ subunit is devoid of nucleotide-binding motifs, it might be that the γ subunit is the one to which these ligands bind. Although the region with sequence similarity to P-loop binding domains is also found in the eIF2B ϵ [7], this subunit was not labelled by ATP or GTP analogues [3]. eIF2B can also be photoaffinity-labelled by a GTP analogue [3] but the subunit labelled in these experiments was the β , and not the γ or ϵ .

Examination of the rat eIF2B γ sequence in regions other than those with similarity to the P-loop region of nucleotidyl transferases reveals the presence of the sequence GVDSTGK (residues 159–165). This motif is also conserved in the human sequence (Figure 1), but is not apparently conserved in eIF2B γ from *C. elegans* or yeast (Figure 5). It is possible that this motif in the rat sequence is involved in the binding of NADPH by mammalian eIF2B, a property that has yet to be demonstrated for eIF2B from yeast or *C. elegans*.

The I-patch domain

The I-patch, or isoleucine hexamer repeat domain, is a feature of

eIF2B gamma	ME-----	FOAVVMVVGSSRM TDLTSSITKPLLPVGNKPLTWYPLNLERVGFEEVIVV-	55
eIF2B epsilon	MAATAAVPSAVGGRANKRGGSGGGTQGAEEPPPP	LQAVLVADSGFDRRFISKDQRLVLLPLANVALIDYTFETATVQETEFVC	90
eIF2B gamma	--TTKDVQKAL-----GAEFKMKLKPDIVCIPDEADMTADSLRHITPL--	LKTQVLVLGCDLITDVAQHEVVD--	132
eIF2B epsilon	CWKAQQLKEHLQSKWCHPTSLNV---MRITTSQLYRSLGQVLRQVDAKALVRSDFLL	LYGVVSNINISKALEEHRRLRKLEKNVSVMT	177
eIF2B gamma	R--KQDESTIEVPQKQK-----KKTVEORDI-----GVDSIGKPL-----	LFMAN-----EADL	176
eIF2B epsilon	MVFKESSPSHPTCHEDNVVLAVDSTTNRILHFQKTQGLRHISFPLGLFGQSLDQVEI	RYDLLDCHISICSPQVAQLFTDNFDYQTRDIF	267
eIF2B gamma	DEELVITKGSLLQKHPRTH-----FDTGLVDALY-----LKKYV-----	VD-----	216
eIF2B epsilon	VRGLLVNNEELGNQIHLVTSREYGSRVNLLMYSAVCTDVIRRYMYPLTPEVNT	DSSTQSYTHSRHNIYRGPEVSLGHGSVLEENVLL	357
eIF2B gamma	ENKSTIT---SIRSELI-----PYL---VRKQFSSASSO-----QROEDKEDKKKE-----	PKSDDTYSFTIKKDNLTIA	276
eIF2B epsilon	GAGTVYVGSNCSITNSVIGPNCHIGDNVLDQAYLWQGVRAAGAQIHISLLCDRAEVKERVIL	KPHCVLTQSVVGGFDITLPEGSVLSH	447
eIF2B gamma	PYDACWN---AFRRKWEQLSRQVR-----GYV-----HIMKEGLCSRVTGLYMEANRQVKKLLSVLCP	ESMIH	341
eIF2B epsilon	PPDAEEDDDGQSDSGADQEKVKLKGYNPAEVGPEGQGYLWKAEDVDEKEDERQSLWGLMINNEE	SESESESVDPEDLSRA	537
eIF2B gamma	PSAQTANKHLIGADSLIGSDTQVGGKSKIKRSVIGSSCV-----	TRDRVTITNCLLMSVITVEEGSSITHGVCN-----	412
eIF2B epsilon	GSFQDDIRVFQNEVGLTQRGREENISCDNLVLEINSLKYAYNISLKEVMOVL	SHVVLEFRLQQVDGVLDPNRYC	627
eIF2B gamma	-----NAVVEAGAEIRDCLTIGSGQRTEAKAKRM-----NEVIVG-----	NDOLM-----EII	452
eIF2B epsilon	RNYIKRAADHLEALAAIEDFFLEHETLMPSLAKVLMFAFYQLEILAEETLS	WFSQSDITDKGQQLRKNOOLGRFIQWLREAEESDD	716

Figure 6 Alignment of the protein sequences of mammalian eIF2B γ and eIF2B ϵ .

The sequences used are those of rat eIF2B γ (present study) and rat eIF2B ϵ (accession number U19511). Sequences were aligned as in Figure 3 and show 17% overall similarity.

bacterial acyltransferases that has been proposed to bind acyl groups, as discussed by Koonin [7]. Several of these enzymes are multimers [7] and these motifs might also be involved in subunit/subunit interactions. The sequence of rat eIF2B γ also contains these motifs (see Figure 2) although their function is not clear. Apart from recycling eIF2, eIF2B has no known cellular function. It is possible that these I-patches are involved in protein/protein interactions involving the five subunits of eIF2B, and/or other translational components. The sequences of eIF5 and eIF2B ϵ exhibit similarity, particularly in their C-termini [7]. eIF5 and eIF2B each form stable complexes with eIF2, and it is tempting to speculate that conserved regions are involved in interaction with eIF2. These conserved regions are distinct from the hexapeptide-repeat-rich regions of eIF2B γ/ϵ .

We have also observed a high degree of sequence similarity between the α , β and δ subunits of eIF2B from yeast and mammals, especially in their C-terminal sections [23]. In particular there are regions with regular repeats of isoleucine or other hydrophobic residues, with a periodicity of six or seven residues [23]. It is possible that these regions interact with the I-patches on eIF2B γ/ϵ . The motifs in eIF2B α , β and δ are not detected as being related to hexapeptide repeats by searching the Blocks database.

Concluding remarks

The availability of cDNAs for all five mammalian eIF2B subunits from a single species should allow further molecular studies on the structure, function and regulation of mammalian eIF2B. In particular this should facilitate the investigation of the roles of the potential nucleotide-binding regions of eIF2B γ and ϵ (the regions of sequence similarity to nucleotidyl transferases) and the additional motif in rat eIF2B γ . Experiments are also required to elucidate the role of the I-patch motifs in eIF2B γ and ϵ , and their possible interactions with regions of other subunits of eIF2B, its substrate, eIF2, or other components of the translational machinery.

We thank Peter Martin for performing automated DNA sequence analysis. This work was supported by grants from the Wellcome Trust (to C.G.P.) and the National Institutes of Health (DK 15656 and DK13459 to L.S.J.).

REFERENCES

- Price, N. T. and Proud, C. G. (1994) *Biochimie* **76**, 748–760.
- Dholakia, J. N., Mueser, T. C., Woodley, C. L., Parkhurst, L. J. and Wahba, A. J. (1986) *Proc. Natl. Acad. Sci. U.S.A.* **83**, 6746–6750.
- Dholakia, J. N., Francis, B. R., Haley, B. E. and Wahba, A. J. (1989) *J. Biol. Chem.* **264**, 20638–20642.
- Oldfield, S. and Proud, C. G. (1992) *Eur. J. Biochem.* **208**, 73–81.
- Kimball, S. R. and Jefferson, L. S. (1995) *Biochem. Biophys. Res. Commun.* **212**, 1074–1081.
- Hill, D. E. and Struhl, K. (1988) *Nucleic Acids Res.* **16**, 9253–9265.
- Koonin, E. V. (1995) *Protein Sci.* **4**, 1608–1617.
- Bushman, J. L., Asuru, A. I., Matts, R. L. and Hinnebusch, A. G. (1993) *Mol. Cell. Biol.* **13**, 1920–1932.
- Craddock, B. L., Price, N. T. and Proud, C. G. (1995) *Biochem. J.* **309**, 1009–1014.
- Flowers, K. M., Mellor, H., Kimball, S. R. and Jefferson, L. S. (1995) *Biochim. Biophys. Acta* **1264**, 163–167.
- Flowers, K. M., Kimball, S. R., Feldhoff, R. C., Hinnebusch, A. G. and Jefferson, L. S. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 4274–4278.
- Price, N. T., Francia, G., Hall, L. and Proud, C. G. (1994) *Biochim. Biophys. Acta* **1217**, 202–210.
- Henderson, R. A., Krissansen, G. W., Yong, R. Y. Y., Leung, E., Watson, J. D. and Dholakia, J. N. (1994) *J. Biol. Chem.* **269**, 30517–30523.
- Hinnebusch, A. G. (1994) *Trends Biochem. Sci.* **19**, 409–414.
- Wilson, R., Ainscough, R., Anderson, K., Baynes, C., Berks, M., Burton, J., Connell, M., Bonfield, J., Copsey, T., Cooper, J. et al. (1994) *Nature (London)* **368**, 32–38.
- Jackson, V. N., Price, N. T. and Halestrap, A. P. (1995) *Biochim. Biophys. Acta* **1238**, 193–196.
- Boguski, M. S., Tolstoshev, C. M. and Bassett, Jr., D. E. (1994) *Science* **265**, 1993–1994.
- Altschul, S. F., Boguski, M. S., Gish, W. and Wootton, J. C. (1994) *Nature Genet.* **6**, 119–129.
- Collins, J. F. and Coulson, A. F. W. (1990) *Methods Enzymol.* **183**, 474–486.
- Price, N. T., Redpath, N. T., Severinov, K. V., Campbell, D. G., Russell, J. M. and Proud, C. G. (1991) *FEBS Lett.* **282**, 253–258.
- Adams, M. D., Kerlavage, A. R., Fleischmann, R. D., Fuldner, R. A., Bult, C. J., Lee, N. H., Kirkness, E. F., Weinstock, K. G., Gocayne, J. D., White, O. et al. (1995) *Nature (London)* **377** (suppl.) 3–174.
- Kimball, S. R., Karinch, A. M., Feldhoff, R. C., Mellor, H. and Jefferson, L. S. (1994) *Biochim. Biophys. Acta* **1201**, 473–481.
- Price, N. T., Mellor, H., Craddock, B. L., Flowers, K. M., Wilmer, T., Kimball, S. R., Jefferson, L. S. and Proud, C. G. (1996) *Biochem. J.* **318**, 637–643.
- Henikoff, S. and Henikoff, J. G. (1994) *Genomics* **19**, 97–107.
- Kuwabara, P. E. and Shah, S. (1994) *Nucleic Acids Res.* **22**, 4414–4418.
- Singh, L. P. and Wahba, A. J. (1996) *Biochem. Biophys. Res. Commun.* **217**, 616–623.