Molecular cloning and characterization of a hypoxia-responsive CITED3 cDNA from grass carp

P.K.S. Ng, R.S.S. Wu, Z.P. Zhang, H.O.L. Mok, D.J. Randall, R.Y.C. Kong*

Department of Biology and Chemistry and Centre for Coastal Pollution and Conservation, City University of Hong Kong, 83 Tat Chee Avenue, Kowloon Tong, Hong Kong, PR China

Received 25 February 2003; received in revised form 4 June 2003; accepted 4 June 2003

Abstract

We have isolated a 1586-bp full-length CITED3 cDNA from grass carp which specifies for a cAMP-responsive element-binding protein/p300-interacting transactivator with glutamic acid (E)/aspartic acid (D)-rich C-terminal domain protein. The cDNA, designated as gcCITED3, has an open reading frame of 762 bp and encodes a protein of 253 amino acids with a predicted molecular mass of 28.3 kDa and pI of 6.4. Pairwise comparison showed that gcCITED3 shares high sequence identity with the CITED3 of zebrafish (94%), chicken (72%) and Xenopus (59%). Northern blot analysis indicated that gcCITED3 is most highly expressed and responsive to hypoxia in the carp kidney. Hypoxic induction was also observed in heart, albeit at a lower level. This is the first report on the isolation of a hypoxia-responsive CITED3 gene from fish.

Keywords: Ctenopharyngodon idellus; CITED; Grass carp; Hypoxia; Gene regulation; Co-transactivator; RACE PCR

1. Introduction

The CITED (cAMP-responsive element-binding protein (CBP)/p300-interacting transactivators with glutamic acid (E) and aspartic acid (D)-rich C-terminal domain) proteins are members of a new family of transcriptional co-activators that are involved not only in the modulation of a variety of cellular and developmental processes (Dunwoodie et al., 1998; Yahata et al., 2001) but also in responses to various biological (Sun et al., 1998) and environmental (Bhattacharya et al., 1999) stimuli. CITEDs do not contain DNA-binding motifs nor possess any significant DNA-binding activity but interact with other DNA-binding regulatory proteins and function as transcriptional co-activators (Yahata et al., 2002).

So far, four different CITEDs have been cloned and characterized from various vertebrates and they share a highly conserved 32-amino acid sequence motif (designated as CR2 domain) at the C-terminal transactivating domain that binds to the CBP/p300 transcriptional activators (Bhattacharya et al., 1999). During mouse embryogenesis, it has been demonstrated that CITED1 (formerly MSG1) is predominantly expressed in the heart tube, limb bud and sclerotome; CITED2 (formerly MRG1)
is expressed largely in the precardiac and anterior visceral mesoderm (Dunwoodie et al., 1998) and neuroectoderm (Schlange et al., 2000); and CITED-ED4 (formerly MRG2) is strongly and selectively expressed in hematopoietic tissues and endothelial cells (Yahata et al., 2002). In contrast, CITED3 is highly expressed during early stages of embryonic development in the mesonephric tubules and eye of chick (Andrews et al., 2000) and the pronephros and eye of frog (Gawantka et al., 1998).

CBP/p300 are ubiquitously expressed homologous nuclear proteins which function as transcriptional co-activators linking stimulus-responsive DNA-binding proteins to the basal transcriptional machinery (for a review, see Goodman and Smolik, 2000). Direct binding of the CH1 region of CBP/p300 to the hypoxia-inducible factor, HIF-1, was found to be crucial for cellular response to hypoxia (Arany et al., 1996). Recently, it has been demonstrated that CITED2 is induced by hypoxia and acts as a negative regulator of HIF-1α functions by competitive binding to the CH1 domain of CBP/p300 (Bhattacharya et al., 1999).

While much is known about the distribution and regulation of the CITED genes in birds and mammals, not much is known about fish. As one of the most important and largest group of vertebrates, fish often have to contend with hypoxic stress in order to survive in environments with low and/or variable oxygen levels. Whereas much is known about the physiological and biochemical responses of fish to hypoxia, there is little information at the molecular level. cDNA microarrays have been used to study hypoxia-induced gene responses in zebrafish (Ton et al., 2002) and the euryoxic fish, *Gillichthys mirabilis* (Gracey et al., 2002). Our group is interested in understanding the molecular basis of hypoxia tolerance in the grass carp and using subtractive libraries and cDNA microarrays, several novel hypoxia-inducible genes have been identified including a CITED-like partial cDNA (Zhang, 2002). Here, we describe for the first time the cloning of a hypoxia-inducible CITED3 cDNA from the grass carp and its expression and response pattern to short- and long-term hypoxia.

2. Materials and methods

2.1. Fish

Grass carp, *Ctenopharyngodon idellus*, were obtained from a commercial hatchery and acclimated in 300-l fiberglass tanks with circulating, filtered and well-aerated tap water at 20 °C for 1 week prior to experimentation. Fish were reared under normoxia (7.0 ± 0.2 mg O₂/l) or hypoxia (0.5 ± 0.3 mg O₂/l) in a continuous flow system described by Zhou et al. (2001). Dissolved oxygen were monitored continuously using a YSI Model 580 dissolved oxygen meter.

2.2. RNA isolation and cloning of full-length cDNAs

Total RNA was isolated from grass carp tissues using the Trizol reagent (Invitrogen) and further purified using the SV total RNA isolation kit (Promega) according to the manufacturer’s instructions. Poly (A)⁺ RNA was purified from total RNA using the PolyATtract System kit (Promega). A 1.4-kb cDNA fragment of grass carp CITED3 was isolated from a grass carp liver cDNA library prepared in λTriplEx2 in our laboratory using the Smart cDNA library construction kit (Clontech). 5’-RACE was performed using the Marathon cDNA amplification kit (Clontech). The adaptor primers AP1 and AP2 were purchased from Clontech. Gene-specific nested primers for RACE were: primer L7B6-R2, 5’-CTGATGCTGCTGAACCGT-GATGTTGC-3’ and primer L7B6-R1, 5’-GGTGACCCCATCTGCTCCATCTAGGG-3’ and were designed based on the CITED-like partial cDNA. First-strand cDNAs were synthesized with Superscript II RNase H⁻ reverse transcriptase (Invitrogen) and 35 cycles of PCR were performed using Advantage2 Taq polymerase (Clontech) according to the manufacturer’s recommendations. RACE products were cloned into a pGEM-T vector (Promega) for DNA sequencing. Full-length cDNAs were obtained by reverse-transcription PCR using gene-specific primers: L7B6-5’, 5’-GTGTCGGTCCGCTGGAGTTAT-3’ and L7B6-3’, 5’-CAAGTGCACATAGGTGTTTG-3’.

2.3. Northern blot analysis

Total RNA (20 μg) from different tissues was electrophoresed on 1% agarose/formaldehyde gels and blotted onto nylon membrane (Hybond-XL, Amersham Biosciences). DNA probes were radiolabeled by the random priming method and 2.0×10⁶ cpm/ml were used in northern hybridizations and were carried out at 65 °C for 2 h in ExpressHyb solution (Clontech). Blots were washed thrice in 2× standard sodium citrate.
Fig. 1. Nucleotide and deduced amino acid sequences of grass carp CITED3. Nucleotide positions are on the left and amino acid positions are indicated on the right. The conserved CR1, CR2 and CR3 domains are boxed. The potential phosphorylation sites at serine-12 and serine-180 are indicated by arrows. The putative CBP/p300-binding site within the CR2 domain is highlighted in gray. A consensus polyadenylation signal (AATAAA) is in bold type and underlined.

2.4. DNA sequencing and sequence analysis

Plasmid DNA was purified using the Miniprep Concert kit (Invitrogen) and DNA sequencing reactions were performed using the ABI PRISM Dye Terminator Kit (Applied Biosystems). DNA sequences were analysed on ABI PRISM 377 automated sequencer (Applied Biosystems).

2.5. Phylogenetic analysis

Phylogenetic analysis was performed by maximum parsimony using the PROTPARS program of the PHYLIP package version 3.57c (Felsenstein, 1995). Phylogenetic tree was displayed using Treeview (Page, 1996). Sequence analyses and
Fig. 2. Northern blot analysis of gcCITED3. Total RNA (20 μg) from different tissues of grass carp subjected to normoxia (N) and hypoxia (H) for 4 and 96 h were analysed by Northern hybridization. A representative Northern blot derived from the tissues of one normoxic and one hypoxic fish from a total of three in each group is shown. Hybridization signals were normalized against the 28S rRNA stained with ethidium bromide.

Homology searches were performed using the online BLAST suite of programs (NCBI, USA).

2.6. Statistical analysis

A non-parametric χ² was used to test the null hypothesis that the ratio of hypoxic:normoxic expression level was not significantly different from 1 (Siegel, 1956).

3. Results and discussion

3.1. Cloning of a full-length CITED3 cDNA and sequence analysis

A grass carp cDNA microarray that was fabricated from a liver cDNA library (constructed in the λTriplEx2 vector) was used to screen for genes that are responsive to hypoxia in the grass carp (Zhang, 2002). A 1.4-kb cDNA (PA-cDNA1) which showed significant hypoxia upregulation and high amino acid sequence identity (>90%) to zebrafish CITED3 (accession number AAK43715) was one of several novel cDNA genes identified in the screen. DNA sequence analysis showed that PA-cDNA1 contained a poly (A) stretch at the 3’ end but lacked the start codon. Using 5’-RACE, a 0.4-kb cDNA was obtained and DNA sequencing showed that it contained sequences at its 3’ end that overlapped with the 5’ end of PA-cDNA1, and this information was used to assemble the full-length cDNA sequence. In order to verify the authenticity of the cDNA sequence, a pair of primers that corresponded to the 5’- and 3’-ends of the putative full-length cDNA were designed and tested by RT-PCR using as template grass carp kidney poly (A+) RNA that was purified using the PolyATtract mRNA isolation kit (Promega). A single RT-PCR product of approximately 1.6 kb was obtained (data not shown) and cloned into a plasmid vector. The cDNA was completely sequenced on both strands, and pair-wise comparison using BLAST 2 (NCBI) showed that it is 100% identical to the putative full-length PA-cDNA1 sequence. The full-length cDNA is 1586 bp in length (including the polyA tail) and contains 5'- and 3'-untranslated (UTR) regions of 133 and 690 bp, respectively (Fig. 1) and is in agreement with the detection of a single mRNA transcript of approximately 1.6 kb by Northern blot analysis (data not shown).

The open reading frame (ORF) of PA-cDNA1 is 762 bp long and encodes a protein of 253 amino acids, with a predicted molecular mass of 28.3 kDa and pI of 6.4. A BLASTX homology search showed that the ORF shares high sequence identity with the CITED3 of zebrafish (94%), chicken (72%) and Xenopus (59%) and moderate (CITED2; 45–53%) to low (CITED1 and CITED4; < 40%) sequence identity with other CITED proteins from various animal species. The analysis strongly indicated that PA-cDNA1 encodes for the grass carp CITED3 protein and is hereupon designated as gcCITED3.

3.2. In vivo expression and response pattern of gcCITED3 to short and long-term hypoxia

To study the in vivo expression and response pattern of gcCITED3 to hypoxia, grass carps were exposed to normoxic (7 mg O₂/l) and hypoxic (0.5 mg O₂/l) conditions and fish were sampled from each treatment group and control after 4 and 96 h. Total RNA was isolated from six different tissues (brain, gill, heart, kidney, liver and muscle) of each of three fish from the normoxic and hypoxic groups at each time point for Northern blot analysis. By and large, the expression and hypoxia induction patterns of the gcCITED3 gene in different grass carp tissues were consistent amongst the replicate blots and a representative autoradiogram is shown in Fig. 2. Under normoxic
conditions, the highest expression of gcCITED3 mRNA was detected in kidney; and low expression levels were detected in brain, gill, heart and liver. This result is quite similar to earlier observations by Gawantka et al. (1998) and Andrews et al. (2000) in frog and chicken, respectively, where CITED3 expression was detected in various embryonic cell types associated with kidney develop-
opment. No expression was detectable in muscle. Exposure to hypoxia for 4 and 96 h resulted in a marked and persistent increase in gcCITED3 expression in kidney. The fish kidney is not only involved in ionic and osmotic regulation, but is also the site of erythropoiesis and the production of erythropoietin (Wickramasinghe, 1993), a hormone regulated by HIF-1 in mammals and probably also in fish. It is possible that gcCITED3 may play a role in modulating the actions of HIF-1 in

| gcCITED3 | 166 | HMKCNSHCNGAUKM3AQILSNM |
| gcCITED3 | 167 | PMKAEKNTKQPMNRHCM |
| gcCITED3 | 167 | PMKAEKNTKQPMNRHCM |
| gcCITED3 | 168 | HMKCNSHCNGAUKM3AQILSNM |
| gcCITED3 | 169 | HMKCNSHCNGAUKM3AQILSNM |
| gcCITED3 | 170 | HMKCNSHCNGAUKM3AQILSNM |
| gcCITED3 | 171 | HMKCNSHCNGAUKM3AQILSNM |

Fig. 3 (Continued).
the kidney. Whilst hypoxic induction of gcCITED3 in liver was only marginal at 4 h, a dramatic induction in expression was observed at 96 h. Hypoxic induction of gcCITED3 was also observed in gill at 4 h, and in heart at 4 and 96 h, although at a lower level. In vivo expression of gcCITED3 was seemingly unaffected by both short and long-term hypoxia in brain and muscle of grass carp.

The expression level of gcCITED3 in all replicates of each tissue under hypoxic conditions were normalized against 28S rRNA and compared with the normoxic counterpart. If expression levels were the same in both the hypoxia treatment and normoxia control groups, the hypoxia:normoxia expression ratio should theoretically be equal to 1; whereas the ratio should be significantly greater than 1 if hypoxia induction was observed. The 4-h and 96-h datasets of each tissue were lumped whereas the ratio should be significantly greater than 1 if hypoxia induction was observed. The 4-h and 96-h datasets of each tissue were lumped and a non-parametric $\chi^2$ test was performed to test the null hypothesis that the hypoxia:normoxia expression ratio was not significantly different from 1 (Siegel, 1956). The analysis showed that gcCITED3 expression levels in kidney and liver were significantly higher than the normoxia controls ($P<0.05$).

3.3. Multiple sequence alignment and comparison of gcCITED3 with other CITEDs

In order to obtain more CITED sequences for comparative analysis, the Fugu rubripes (Takifugu rubripes) genome database (version 3 at http://genome.jgi-psf.org/fugu6.home.html) was searched for Cited genes not previously identified in this species. A total of three (two of which were previously reported as FCITED1 and FCITED3 by Braganca et al., 2002) were identified and assembled using the BLAST suite of programs. Interestingly, two Fugu clones showed high sequence identity (59–74%) with the CITED3 proteins from different sources including the Xenopus CITED3 (AI031460), and are designated here as fCITED3a and fCITED3b (FCITED3 equivalent). fCITED3a is a new isoform that was identified in our bioinformatic search and shares 71 and 69% identity with gcCITED3 and fCITED3b, respectively. gcCITED3 and fCITEDb shares 73.6% sequence identity. Homology searches of the Drosophila, Caenorhabditis elegans and Saccharomyces cerevisiae databases with various CITED-EDs did not produce any sequence match of appreciable significance.

The deduced amino acid sequence of gcCITED3 was multiply aligned with known sequences of all four classes of CITED proteins available to date from various nucleotide/protein databases and the result is shown in Fig. 3. The alignment indicated that the signature motifs represented by the CR1, CR2 and CR3 domains (which are characteristic features of the CITED family of proteins) are all conserved in gcCITED3. Incidentally, the CR1, CR2 and CR3 domains are also conserved in the CITED2 proteins. In contrast, CITED4 (Yahata et al., 2002) and CITED1 (Dunwoodie et al., 1998) do not contain the CR1 and CR3 domains, respectively.

An unusual feature of gcCITED3 and a few of the other CITED3 proteins (except chicken and Xenopus), is the presence of a glutamine-rich domain extending from residue Q-78 to Q-135 (Fig. 3). Interestingly, glutamine-rich domains have also been detected in CBP/p300 (where direct interaction with SRC-1 has been demonstrated; Yao et al., 1996), and in a number of newly-described transcriptional regulators such as TIG-1 (Abraham and Solomon, 2000) and PCQAP (Berti et al., 2001), which exhibit co-activator function in RNA polymerase II transcription. Apart from the Q-rich domain, major sequence variabilities between the four classes of CITED proteins can be detected in two regions, the N-terminal region...
between CR3 and CR1, and the short region between CR1 and CR2 (Fig. 3). Notwithstanding this, members of each CITED class show ‘intra-class’ conserved sequence motifs in these two regions, and this raises the possibility that these short motifs of conserved amino acids may dictate the interaction/regulatory specificity and in turn the functional roles of the different classes of CITED proteins. Analysis of gcCITED3 with the NetPhos program (http://www.cbs.dtu.dk/services/NetPhos) predicted two potential phosphorylation sites, one at serine-12 and the other at serine-180 (Fig. 1). A similar analysis performed on other CITED3 members predicted that c-CITED3 and ICITED3a, respectively, contain three (serine-10, -18 and -145) and one (threonine-63) potential phosphorylation sites (Fig. 3), and none in ICITED3b and zCITED3. In contrast, a more variable number of potential phosphorylation sites (between 3 and 12) were predicted in both human and mouse CITED1, CITED2 and CITED4, and are consistent with the diverse biological roles ascribed to these proteins, which range from cellular responses to cytokines (Sun et al., 1998) and hypoxia (Bhattacharya et al., 1999), to signal transduction (Shioda et al., 1998) and vertebrate development (Dunwoodie et al., 1998).

Hydropathy analysis showed that the hydrophobicity profile of gcCITED3 and its pI (≈6.4) are very similar to those of other CITED3 proteins (data not shown). Although members of the different CITEDs are very hydrophilic in character, the hydrophobicity profile of each CITED class is distinct (as represented by the grass carp and human proteins in Fig. 4). Moreover, the pIs of CITED1 (4.53) and CITED4 (5.38) are comparatively lower than those of CITED2 and CITED3 which are >6.3. The overall sequence identity between members of CITED1–CITED4 is not high which range between 31 and 43%.

 Parsimony analysis of 19 CITED proteins from different animal species showed that they were grouped as expected, into four respective clades with high bootstrap support (Fig. 5). The tree shows that gcCITED3 is most closely related to the zebrafish CITED3 and is supported by a bootstrap value of 100%. The genomic sequence data sets for human and mouse are complete, and homology searches using BlastX with various CITED3 proteins failed to turn up any CITED3-like sequences from their genome databases. This together with the fact that no mammalian CITED3 has been reported to date, suggest that mammals lack the gene for CITED3. Conversely, whilst CITED3 has so far been detected only in *Xenopus*, chicken, puffer fish (torafugu), zebrafish and grass carp, the gene for CITED4 has not been reported in these animals. Bioinformatic searches of the *Takifugu rubripes* genome with various mammalian CITED4 also failed to turn up any CITED4-like sequence, which suggest that CITED4 genes may not be present in the genomes of non-mammalian vertebrates. It is, therefore, tempting to speculate that CITED4 and CITED3 are orthologs that may have arisen subsequent to lineage divergence between mammalian and non-mammalian vertebrates.

To date, CITED2 has been the most extensively studied and is known to possess many functions. It is a hypoxia-inducible gene that negatively regulates HIF-1 activity by competing for the CH1 region of CBP–p300 (Bhattacharya et al., 1999). Moreover, gene knockout studies have shown that CITED2 is required for normal embryonic development (Bamforth et al., 2001), and in vitro
binding assays have indicated that CITED2 enhances the transcription of the LH/FSH α-glycoprotein subunit (Glen and Maurer, 1999) and co-activates different isoforms of the transcription factor AP2 (Bamforth et al., 2001). The physiological functions and molecular interactions of CITED3 have not as yet been described and are still largely unknown. It is likely that novel co-activator molecules will be discovered that would interact physiologically with these molecules to modulate the expression of a number of hypoxia-responsive genes. Therefore, much work has yet to be done to determine its specific interaction, in particular with HIF-1, and its physiological roles in relation to hypoxia tolerance in fish.

Acknowledgments

This work was funded by a Central Earmarked Research Grant (Project No. CityU1057/99M)
from the Research Grants Council of Hong Kong Special Administrative Region, People’s Republic of China.

References


Glen, D.J., Maurer, R.A., 1999. MRG1 binds to the LIM domain of Lhx2 and may function as a coactivator to stimulate glycoprotein hormone α-subunit gene expression. J. Biol. Chem. 274, 36159–36167.


