Genomic organization of nucleolin gene in carp fish: Evidence for several genes

CLAUDIA QUEZADA*, CRISTINA NAVARRO, RODY SAN MARTÍN*, MARCO ÁLVAREZ, ALFREDO MOLINA and M. INÉS VERA

Departamento de Ciencias Biológicas, Universidad Andrés Bello, Millennium Institute for Fundamental and Applied Biology, Santiago, Chile.

ABSTRACT

The protein nucleolin, functionally involved in the main steps of ribosome biogenesis, is codified by a single copy gene in mammals. Here we report that at least three different genes codify for this protein in carp fish (*Cyprinus carpio*). This is the first description of the genomic organization of nucleolin in a teleost. The carp nucleolin gene includes 8.8 kb and contains 16 exons. Promoter *cis* regulatory elements are similar to constitutive genes, i.e., a putative TATA box, three G/C boxes, and three pyrimidine-rich boxes. As in other species, carp nucleolin gene introns host three snoRNA codifying sequences: U23 from the H/ACA family and two C/D box snoRNAs, U20 and U82. Both U20 and U82 span a complementary sequence with carp 18S rRNA. Additionally, we identified two cDNAs coding for nucleolin, confirming the existence of several nucleolin genes in carp. Amino acid-derived sequence from carp cDNAs differ from mammal protein because they span additional acidic domains at the amino end, whose functional significance remains unclear. We performed amino acid sequence comparison and phylogenetic analyses showing that the three isoforms of carp nucleolin, which we describe herein, cluster in two groups. cNUC1 probably diverges from cNUC2 and cNUC3 as result of ancestral fish-specific genome duplication, indeed *C. carpio* is a tetraploid fish.

Key terms: nucleolin, nucleolus, U snoRNAs, rRNA 18S, carp fish.

INTRODUCTION

The nucleolus and its most well known major function, ribosome biogenesis, constitute a highly dynamic structure-functional entity (Platani and Lamond, 2003; Raska et al., 2004). In addition to the components of the mature ribosome, the nucleolus contains many RNAs and proteins that interact transiently with non-ribosomal constituents and at various stages of the ribosomal synthesis (Fromont-Racine et al., 2003). Nucleolin, the most abundant non-ribosomal protein linked to ribosomal biogenesis, is involved in remodeling the chromatin structure, rDNA transcription, precursor rRNA processing, ribosomal assembly, and in the nuclear-to-cytoplasm transport (Erard et al., 1988; Kharrat et al., 1991; Bouvet et al., 1998; Ginisty et al.,

1998; Ginisty et al., 1999; Roger et al., 2002; Roger et al., 2003). Recently, this 100kDa phosphoprotein also has been associated with apoptosis (Otake et al., 2005; Kito et al., 2005).

The organization of the gene encoding for nucleolin has been described only in mammals: humans and rodents. There is only one copy of the nucleolin gene per human haploid genome, which is located on chromosome 2q12-qter. The 11kb gene is comprised by 14 exons with 13 introns (Srivastava et al., 1990) and codes for a 707 amino acid protein (Srivastava et al., 1989). Nucleolin exhibits a tripartite structural organization (Ginisty et al., 1999); the N-terminal domain contains highly acidic stretches interspersed with basic regions, presenting many phosphorylation sites; the central region includes four RNA binding

Corresponding author: M. Inés Vera, Departamento de Ciencias Biológicas, Universidad Andrés Bello, República 217, Piso 4, Santiago, Chile, Tel: (56-2) 661-8101, Fax: (56-2) 698-0414, E-mail: mvera@unab.cl

Received: January 17, 2006. Accepted: April 18, 2006

* Present address: Instituto de Bioquímica, Universidad Austral de Chile, Valdivia, Chile.

domains (RBD) and the C-end, rich in glycine, arginine, and phenylalanine residues, called the GAR or RGG domain. In human nucleolin, the four acidic regions of the amino-terminal domain are coded by exons 2-4, the nuclear localization signal by exon 5, and each of the four RBD is coded by two consecutive and independent exons (exons 6 to 13) (Srivastava et al., 1990). This genomic organization is highly conserved in rodents (Bourbon and Amalric, 1990). Mouse and hamster nucleolin genes span over 9kb coding for proteins comprised of 706 and 713 amino acids, respectively (Bourbon et al., 1988a; Lapeyre et al., 1985). In mammals, the nucleolin gene promoter includes a TATA-like box, two pyrimidine-rich regions, and two CCAAT-type boxes (Bourbon et al., 1988b; Srivastava et al., 1990).

The adaptive adjustments of carp fish to the seasonal cycles of habitat conditions (temperature, photoperiod) involve a clear modulation of gene expression, among other cellular and molecular changes (Alvarez et al., 2004; San Martin et al., 2004; Pinto et al., 2005). The most dramatic phenotypical demonstration is the segregation of the nucleolar components during the winter acclimatization process, which is reversed when the fish adapts to summer conditions (Sáez et al., 1984). In the carp, the molecular mechanisms that sustain the winter nucleolar rearrangement involve reduced ribosomal biogenesis, concurrently with a decrease in the expression of other nucleolar protein factors, i.e., ribosomal protein L41 and protein kinase CK2 β (Vera et al., 2000; Alvarez et al., 2001; Molina et al., 2002). Previously, we reported that in winter-adapted carp, nucleolin protein content is up-regulated (Alvarez et al., 2003), supporting the idea that nucleolin protein plays a primary role in repressing rRNA synthesis (Roger et al., 2002, 2003).

In this context, nucleolin seems to have a particularly relevant role in the process of seasonal acclimatization in ectotherms. Thus, we deemed it important to study the carp nucleolin gene characterizing both its genomic organization and its cDNA structure. We described previously the

cloning of a full-length carp nucleolin cDNA encoding for a 693 amino acid protein (cNUC1), which depicted a higher number of acidic repeats in the N-terminal region than mammal protein (Alvarez et al., 2003). Here, we report the characterization of the complete carp nucleolin gene organization and a third cDNA sequence. The introns of this gene codify for three small nucleolar RNAs (snoRNAs) as in other species, even though carp snoRNAs are localized in different introns. Because the genomic sequence we isolated (cNUC3) encodes for a different nucleolin protein of 637 amino acids, with only one, additional, highly acidic sequence at the N-terminal region, we examined different carp nucleolin cDNAs and identified a third, distinct, nucleolin expressing for a 643 amino acid protein (cNUC2).

MATERIALS AND METHODS

Animals

Male carp (*Cyprinus carpio*) were captured and maintained under summer (20-22°C) and winter (8-10°C) temperatures (Alvarez et al., 2003). When necessary, the tissues were dissected and frozen at -80°C.

Genomic library screening

The coding sequence of carp U23 snoRNA (GenBank Accession Number AJ009731) was amplified by PCR (pU23' clone) and simultaneously labeled with [α -P³²] dCTP, using gene specific primers (U23S 5'-ttcttctcatgagctcctct-3' and U23A2 5'-tcacatcagacatgggcatg-3') (Mertz and Rashtchian, 1994). Using this probe, a λ FIX II Carp Genomic Library (Stratagene, USA) was screened, yielding one positive clone (Mertz and Rashtchian, 1994). The recombinant DNA, containing a 12.5kb insert (λ U23cc2), was analyzed by restriction mapping, and its nature was confirmed by Southern blot using the U23' probe. The resulting 4.0kb, 3.0kb, 2.0kb, 2.0kb and 1.5kb *Sac* I fragments were subcloned (pGNU4, pGNU3, pGNU2, pGNU68, pGNU1.5) in pBluescript KS+

(Stratagene, USA) and sequenced.

Genomic Southern analyses

Carp genomic DNA (30µg) was digested with *Hind* III, *Pst* I and *Sac* I restriction enzymes, fractionated in a 0.8% agarose gel, blotted onto nylon membrane Immobilon-Ny+ (Bedford, USA) and covalently cross-linked by UV irradiation. The membrane was hybridized with a ³²P-labeled probe, corresponding to the U23 snoRNA, using the manufacturer-recommended conditions. After overnight hybridization at 42°C, the membrane was washed twice with 2X SCC 0.1% SDS for 15 min at room temperature, followed by two washes, each with 1X SCC 0.1% SDS, 0.5X SCC 0.1% SDS at 42°C, and 0.1X SCC 0.1% SDS at 65°C. Membrane was briefly blot-dried and autoradiographed with intensifying screens.

Rapid amplification of cDNA ends (RACE)

A partial nucleolin cDNA clone of carp (pFNUC) was obtained from the screening of a carp liver cDNA library (Álvarez et al., 2003). The full-length 5'-region, including the transcription start site, was obtained using the Firstchoice RLM-RACE Kit (Ambion, USA). The gene-specific primers utilized for nested PCR reactions were NUCext (5'-cctcgtcttcttcagattcc-3') and NUCint (5'-cttcgcgttcaccattcctg-3'). The 5'RACE RT-PCR products (783bp and 616bp) were cloned into the pGEM-T-Easy vector (Promega, USA), and both clones were fully sequenced (pcNUC1 and pcNUC2, respectively). cNUC1 completes the 5'-region of pFNUC, and the cNUC2 sequence represents a different nucleolin cDNA. The full-length 3'-region of cNUC2 was obtained by 3' RACE using the gene-specific primers J1ext (5'-aggaggacgaggaagatgac-3') and J1int (5'-gatgatgatggaagaggagat-3'), cloned into pGEM-T-Easy vector (Promega, USA) and sequenced.

Analyses of carp nucleolin sequences

The nucleotide sequence homology was searched in the GenBank database by using BLAST. Program ClustalX (Thompson et al., 1997) allowed the comparison of nucleotide

and deduced amino acid sequences through the use of the multisequence alignment. The evolution distances were used to construct a phylogenetic tree by the neighbor-joining method (NJ) provided by the ClustalX program according to Saitou and Nei (1987). The sequences considered for these analyses were the following:

Organism	GenBank access number
<i>Cyprinus carpio cNUC1</i> (carp)	AAO22235
<i>Cyprinus carpio cNUC2</i> (carp)	AAQ17064
<i>Cyprinus carpio cNUC3</i> (carp)	AAQ17065
<i>Danio rerio</i> (zebrafish)	NP001003839
<i>Xenopus laevis-a</i> (African clawed frog)	CAA51460
<i>Xenopus laevis-b</i> (African clawed frog)	CAA44805
<i>Mus musculus</i> (mouse)	NP035010
<i>Rattus norvegicus</i> (rat)	NP036881
<i>Cricetulus griseus</i> (hamster)	A27441
<i>Homo sapiens</i> (human)	NP005372
<i>Gallus gallus</i> (chicken)	NP990596

RESULTS

The sequence analyses of the 12.5kb insert of genomic clone λU23cc2 showed that it contains the full length of a carp nucleolin gene, which spans approximately 8.8kb (Fig 1). The sequence was deposited in GenBank, Accession N° AY330169. In addition, we cloned two PCR products of 783bp and 616bp, respectively, containing the 5'-end of two different nucleolin cDNAs from carp liver RNA. The longer amplicon (783bp) fulfils the sequence of a partial cDNA clone, named cNUC1, which codes for a 693 aa protein (GenBank Accession N° AY166587), isolated from a carp liver cDNA library, which we described previously (Álvarez et al, 2003). The 616bp PCR product was completed by 3' RACE experiments, and we confirmed that it corresponds to a second carp nucleolin cDNA (cNUC2), which spans for 2,619bp and codes for a derived protein of 643 residues (GenBank Accession N° AY330167). Comparison analysis demonstrated 86% of homology between cNUC1 and cNUC2 cDNAs, thus, we used both sequences to derive exon/intron organization at the carp nucleolin gene. From the genomic clone, we derived a third

nucleolin cDNA sequence, denominated cNUC3, with a 88% and 97% identity with cNUC1 and cNUC2, respectively, that codifies for a protein of 637 amino acids (GenBank Accession N° AY330168).

This carp nucleolin genomic sequence is organized in 16 exons codifying for an mRNA that spans for 2,595nt (Fig 1). The exon/intron organization boundaries follow the GT-AG rule, starting with GT at the

donor site and ending with AG at the acceptor site (Table I), preceded by a polypirimidine tract (Mount, 1982). All introns contain a potential acceptor site for the intermediate lariat formation, located upstream from the 3' splice site (Ruskin and Green, 1985). Introns 1 and 2 correspond to phase 0; introns 3-7, 9, 11, 13, and 15 to phase I, and introns 8, 10, 12, and 14 are phase II splice type (Rogers, 1985).

Figure 1: Organization of carp and human nucleolin gene and cDNA. a) and c) represent the genomic organization of carp and human, respectively. Gray boxes correspond to exons and white boxes represent 5' and 3' noncoding sequences, respectively. Introns and flanking regions are depicted with thin lines. b) and d) represent the cDNA and the deduced protein organization of both species. The domains codified within the exons are denoted by squared boxes: acidic domains, black boxes: RBDs, dot boxes: nuclear localization sequence.

TABLE I

Exon/intron distribution of carp nucleolin gene

Exon/intron Number	Exon size (bp)	Splicing donor sequence	Intron	Splicing acceptor sequence	Intron size (bp)
1	105	GCCAAG/gtaag.....	ataaaacaacccctggcttttctcgacgtttgcag/GCAGCT		1434
2	93	GAAGAG/gtgag.....	ttcaatctcgattcggcag/GCGCCA		89
3	106	AATCTG/gtgag.....	tcttaattcctctcaatag/AAGAGG		90
4	108	ATGAAG/gtgtg.....	ctaacaactcggtttctctgtgtgtgtgtgtag/ATGACG		80
5	126	ATGATG/gtgtg.....	tctgttaactttttaaagtgtgtcctgattcag/AAGAGG		111
6	114	AAGAAG/gtgag.....	ttattgactttttctgtgttctctctcag/AAACCC		636
7	90	GTGAAG/gtgag.....	tatgccgcactaaagtgtgttgaactttcag/GTTTCA		171
8	124	GACCAA/gtaag.....	tcgtaactttgttttgaag/GAAATT		94
9	137	AAAAAG/gtaca.....	tttaattcttctctcctcag/AGAGAG		183
10	127	CAGAGG/gtaag.....	cttgacactcactcgtcctcctctgtgggtacag/CATCGC		77
11	131	GCAGAG/gtgag.....	gtgttaacacgcctctgtttgttctccttttag/CTACTG		82
12	130	GAAGGG/gtaca.....	tttggctcaactctgtgtgtgtttttttgttag/ATTTCG		1593
13	149	GCTCTG/gttag.....	tgtttgtcaactgttctcctctctgtttgttag/GACCTA		99
14	127	TAAAGG/gtaca.....	tctgaacgtttgtgtgatactgttctcctcctcag/GTTTGG		228
15	275	GAGGCG/gtggg.....	tttaaccggttctctctgtctctctctctcag/GTGGAC		501
16	56				
Consensus sequence		AG/gtaag.....yray(y) _n r(y) _n ag/GT			

The organization of the 5'-untranslated region and the position of the first translating exon were determined by comparison of genomic clone λ U23cc2 nucleotide sequence together with cDNAs sequences of cNUC1 and cNUC2. This genomic clone contains the full length of the codifying region of nucleolin and 319bp from the promoter region, where the putative start site was predicted (Fig 2). In this region, an atypical TATA box (TAAAA) is located at -29 from the +1 site; three G/C boxes at -52, -65, -235; and three pyrimidine-rich nucleotide stretches at the positions -229 to -211, -185 to -163, and -11 to +3. CAAT consensus sequences are not evident in the analyzed region. The GC boxes at -52 and -65 are potential binding sites for the transcription factor Sp1. The pyrimidine-rich stretch located at -11 to +3 includes the putative +1 site, similar to the mouse nucleolin gene organization (Bourbon et al., 1988b). In

addition, we identified numerous potential transcription factors binding sites at the promoter region, like GATA, Myc, Mef-2, LSF, and ERE (Fig 2). The first exon contains 87bp from the non-translated 5'-end and the coding region for the first 6 amino acids, starting at the ATG codon. Exons 2 to 6 codify for 5 acidic domains, exon 7 contains a bipartite nuclear localization signal, and four RNA binding domains (RBD) are codified for exons 8 to 15. The 3'-end of exon 15 and the 5'-end of exon 16 codify for the acidic glycine/arginine-rich (GAR) domain (Fig 1).

One distinctive attribute of nucleolin genes is that some of the introns encode for different snoRNAs. In mammals, intron 5 hosts the sequence of U82, intron 11 for U20, and intron 12 for U23 snoRNA, respectively (Rebane and Metspalu, 1999; Nicoloso et al., 1994; Ginisty et al., 1999). Because carp nucleolin gene contains 16 exons compared to 14 in mammals, carp

Figure 2: Regulatory region of carp nucleolin gene. Recognized *cis* elements for binding of transcription factors are underlined. The putative TATA box, the presumed transcription initiation +1 site, the initial ATG codon, and the first five coded amino acid of this ORF are indicated.

snoRNAs sequences are codified in different localization, U82 at intron 7, U20 at intron 9, and U23 at intron 14, respectively (Fig 3). The derived primary structure and length of carp U82 snoRNA (GenBank Accession N° DQ133600) is highly conserved when compared with other species, we identified the consensus elements C, D and D' boxes that perform the 2'-O-methylation of precursor rRNA (Rebane and Metspalu,

1999; Bachellerie and Cavaille, 1997; Kiss, 2001). Derived carp U20 snoRNA (GenBank Accession N° DQ133601) is 10 nucleotides shorter (71nt) than in mammals (81nt), and putative C and D boxes are distinguishable (Fig 3). Sequence comparison analysis of carp and mammals U20 snoRNA coding regions shows that the carp-derived sequence is conserved at the 5'-and 3'-ends and in the central area (Nicoloso et al., 1994).

Figure 3: Three snoRNAs are codified in nucleolin gene introns. A) Carp nucleolin gene structure showing the localization of coding sequences for U82, U20 and U23 snoRNAs (black boxes) in introns 7, 9 and 14, respectively. Gray boxes correspond to exons. B) snoRNAs sequence comparison among different species. Gray boxes represent the consensus C, D and C' boxes of U82 and U20 snoRNAs and H/ACA boxes of U23 snoRNA.

At carp rDNA cistron sequence (GenBank Accession N° AF133089; Vera et al., 2003), we identified complementary regions between U82 and U20 snoRNAs with 18S rRNA, which could potentially be 2'-O-methylated (Bachellerie et al., 1995). Figure 4 shows two complementary regions, span 12 and 11nt, between carp 18S rDNA and both snoRNAs with adenine in position 1673 and uracil in position 1798 being potentially 2'-O-methylated in carp 18S rRNA, by U82 and U20 snoRNA, respectively.

When we performed PCR using carp genomic DNA and primers derived from carp U23 snoRNA (GenBank Accession N°

AJ009731), we isolated an amplicon which differs in two nucleotides with the sequence previously reported, deposited as carp U23 snoRNA (GenBank, Accession N° DQ133602). Both carp U23 snoRNA codifying sequences are well conserved regarding other species, and boxes H and ACA are easily distinguished, consistent with pre-rRNA pseudouridilation function (Ginisty et al., 1999; Kiss, 2001).

Genomic analyses by Southern blot, using the promoter region of the genomic clone as a probe, identified four or more hybridization bands (Fig 5). These results strongly suggest that there are at least four genes coding for carp nucleolin in carp genome.

Figure 4: Complementary sequences between snoRNAs and 18S rRNA in carp fish. Complementary nucleotides: A) between U82 snoRNA and 18S rRNA and B) between U20 snoRNA and 18S rRNA, respectively. Arrows indicate adenine (A) in position 1673 and uracil (U) in position 1798 of 18S rRNA, which could be methylated.

Figure 5: Carp genomic Southern blot analyses. 30µg of carp genomic DNA were digested with the restriction enzymes *Pst* I, *Sac* I and *Hind* III and hybridized with a probe that contains the 5'-region of cNUC2 clone. Asterisks show the positive hybridization signals, numbers indicate the size of DNA fragments (kb).

The nucleotide sequence comparison analysis of the three carp nucleolin cDNA, using ClustalX program (Thompson et al., 1997) showed 97% homology between cNUC2 and cNUC3, 88% between cNUC1 and cNUC3, and 86% homology between cNUC1 with cNUC2, respectively. Amino acid-derived sequence comparison of the three carp nucleolin cDNAs shows that cNUC2 and cNUC3 share 84% and 85% identity with cNUC1 respectively (Table II). All three carp protein sequences display the classical tripartite domain distribution of nucleolin, involving several acidic regions at the amino end, four central RNA-binding domains, and the GAR domain at the C-terminus. Nevertheless, cNUC2 and cNUC3 differ from cNUC1 because both contain five acidic regions at the N-terminal domain

instead of six regions contained by cNUC1 (Fig 6A). Figure 6B shows the sequence alignment of carp nucleolin amino acid sequences with those of other vertebrate species, using ClustalX (Thompson et al., 1997). The three carp nucleolin protein sequences exhibit approximately 50% of identity with mammals (human, rat, mouse and hamster) and birds (chicken), 55-58% with amphibians (*Xenopus laevis*), reaching up to 78-81% with zebra fish (Table II). The higher degree of conservation entails the GAR domain (98 to 70%) and ribosomal binding domains, RBDs (94 to 60%). On the basis of the evolutionary distances, a phylogenetic tree was constructed by the NJ method (neighbor-joining) provided by the ClustalX software according to Saitou and Nei, (1987) (Fig 6C).

TABLE II

Nucleolin amino acid sequence comparison among vertebrates, showing the percentage of identity of the total protein sequence and among the three protein domains

	cNUC1	cNUC2	cNUC3	Zebrafish	Xenopus A	Xenopus B	Human	Rat	Hamster	Mouse	Chicken
A. Complete sequence											
cNUC1 (carp)	-	84	85	78	56	58	51	50	50	51	51
cNUC2 (carp)	-	-	96	80	55	56	51	51	50	52	52
cNUC3 (carp)	-	-	-	81	57	57	53	52	51	52	53
Zebrafish	-	-	-	-	58	54	50	52	49	50	50
<i>Xenopus A</i>	-	-	-	-	-	84	51	49	51	49	55
<i>Xenopus B</i>	-	-	-	-	-	-	54	52	54	50	56
Human	-	-	-	-	-	-	-	83	84	81	61
Rat	-	-	-	-	-	-	-	-	91	93	59
Hamster	-	-	-	-	-	-	-	-	-	89	59
Mouse	-	-	-	-	-	-	-	-	-	-	57
B. N-terminal domain											
cNUC1 (carp)	-	78	77	74	44	43	36	35	32	37	31
cNUC2 (carp)	-	-	97	76	36	37	39	40	38	43	37
cNUC3 (carp)	-	-	-	75	36	36	38	39	36	42	36
Zebrafish	-	-	-	-	40	35	35	40	37	37	33
<i>Xenopus A</i>	-	-	-	-	-	75	38	35	38	35	42
<i>Xenopus B</i>	-	-	-	-	-	-	44	40	38	33	44
Human	-	-	-	-	-	-	-	73	76	71	51
Rat	-	-	-	-	-	-	-	-	84	87	50
Hamster	-	-	-	-	-	-	-	-	-	81	51
Mouse	-	-	-	-	-	-	-	-	-	-	45
C. Central domain											
cNUC1 (carp)	-	87	88	82	66	66	61	60	60	61	62
cNUC2 (carp)	-	-	94	81	64	63	57	55	55	55	58
cNUC3 (carp)	-	-	-	82	66	66	59	57	57	57	60
Zebrafish	-	-	-	-	66	65	58	57	57	57	60
<i>Xenopus A</i>	-	-	-	-	-	92	62	60	61	60	64
<i>Xenopus B</i>	-	-	-	-	-	-	63	61	62	61	64
Human	-	-	-	-	-	-	-	89	90	87	67
Rat	-	-	-	-	-	-	-	-	98	96	65
Hamster	-	-	-	-	-	-	-	-	-	96	65
Mouse	-	-	-	-	-	-	-	-	-	-	65
D. C-terminal domain											
cNUC1 (carp)	-	96	97	92	70	70	76	77	77	77	75
cNUC2 (carp)	-	-	98	88	68	71	77	79	78	79	75
cNUC3 (carp)	-	-	-	91	74	72	77	77	77	77	75
Zebrafish	-	-	-	-	77	71	80	80	78	80	77
<i>Xenopus A</i>	-	-	-	-	-	75	70	70	77	70	72
<i>Xenopus B</i>	-	-	-	-	-	-	61	56	63	56	74
Human	-	-	-	-	-	-	-	98	98	98	82
Rat	-	-	-	-	-	-	-	-	100	100	80
Hamster	-	-	-	-	-	-	-	-	-	100	74
Mouse	-	-	-	-	-	-	-	-	-	-	80

Percentage of identity were obtained through the comparison of primary structure using ClustalX program. Sequence and numbers are described in Material and Methods.

DISCUSSION

Since it was identified by Orrick (1973), nucleolin has been associated with several steps in ribosomal biogenesis process. These include chromatin structure remodeling, regulation of rDNA transcription, rRNA maturation, ribosomal assembly, and nucleus-cytoplasm transport (Erard et al., 1988; Kharrat et al., 1991; Ginisty et al., 1998, 1999; Bouvet et al., 1998; Roger et al., 2002; 2003).

The isolated genomic clone λ U23cc2 host, the carp nucleolin gene that contains 16 exons and 15 introns, contrasts with mammalian nucleolin genes, which span only 14 exons (Bourbon et al., 1988a; Srivastava et al., 1990). In mammals, a unique gene copy codifying for nucleolin has been described (Bourbon et al., 1988a; Srivastava et al., 1990). In contrast, carp nucleolin could be expressed from several genes, four at the very least (Figure 5). The 5'-regulatory region of the carp nucleolin gene, which contains the basal promoter, revealed that it encloses several features of constitutive genes, such as Sp1 binding sites and pyrimidine-rich regions (Figure 2). Although consensus TATA is not observed, a TATA-like box is found at position -29. Similarly, the human and mouse nucleolin promoter does not present a TATA and CAAT consensus boxes. A GATTACTG sequence found at position -23 and -16 in mouse genes and a GTTACTG sequence at position -49 in human gene, seems to act as a TATA-like box (Bourbon et al., 1988b; Srivastava et al., 1990).

As in other species, U82, U20 and U23 snoRNAs are codified in the introns of the carp nucleolin gene (Rebane and Metspalu, 1999; Ginisty et al., 1999). Nevertheless, as a consequence of the additional exons, U82 and U23 are displaced in two introns when compared with mammalian genes. Thus, U82 is located in the intron 7 between exons codifying for the nuclear localization signal (NLS) and the first RBD. U23 snoRNA is located in the intron 14 between exons codifying for the fourth RBD. Unexpectedly, U20 snoRNA was found within intron 9 instead of intron 13 as in

mammalian nucleolin genes. Both C/D box snoRNA (U20 and U82) are complementary to 18S rRNA of carp. The meaning of these changes in intron location has not been described until now.

With regard to U20 snoRNA, a sequence of 12nt is complementary to a region located in the position 1795-1806 of the 18S rRNA. Although in *S. cerevisiae* and human genes this complementary region involves 21nt (Bachelierie et al., 1995; Bachelierie and Cavaille, 1997), it has been described that, in general, C/D box snoRNAs are able to form a duplex with rRNA sequences that comprise between 10 to 21nt (Lafontaine and Tollervey, 1998). A specific function associated with U20 snoRNA consists in methylation of the 18S rRNA (Bachelierie and Cavaille, 1997). In carp, we propose that this modification involves uracil at position 1798, differing from other species where the modification is in position 1804 (Bachelierie and Cavaille, 1997).

The size and primary structure of the carp U82 snoRNA coding sequence is highly conserved, containing the consensus C, D and D' boxes, responsible for 2'-O-methylation of precursor rRNA. The complementary sequence between U82 snoRNA and carp 18S rRNA corresponds to 11nt, suggesting that methylation occurs at adenine 1673, similar to human genes and mouse genes, where adenine at position 1678 is methylated (Rebane and Metspalu, 1999). Two U82 snoRNA have been detected in mouse genes: a 70nt variant codified in the fifth intron of mouse nucleolin gene and another 67nt that could be transcribed by another, as yet unidentified, gene (Rebane and Metspalu, 1999).

Here, we report the cloning of three cDNA sequences codifying for carp nucleolin, cNUC1, cNUC2 and cNUC3, consistent with Southern hybridization results that showed more than three genes codifying for nucleolin in carp genome. Amino acid-derived sequences from cNUC2 and cNUC3 code for proteins containing five acidic domains, instead of six as in cNUC1, described previously (Álvarez et al., 2003). Comparison among

three carp nucleolin cDNAs confirmed that they are codified by independent genes, which code for five or six acidic domain-containing proteins. The larger number of acidic domains in carp nucleolin cDNA is consistent with the additional number of gene exons. It has been described that nucleolin presents variable number of acidic domains depending on the species (Ginisty et al., 1999). In *X. laevis*, two different cDNAs for nucleolin have been described as yielding proteins of 90 and 95kDa, respectively (Messmer and Dreyer, 1993; Ginisty et al., 1999), both isoforms differ in the content of acidic domains (4 and 5 respectively) (Rankin et al., 1993; Ginisty et al., 1999). These acidic regions are involved in the binding of nucleolin with the nontranscribed spacer regions of rDNA gene and also histone H1. This interaction results in a chromatin structure remodeling of this region that leads to the regulation of the transcription process by RNA pol I (Srivastava and Pollard, 1999). The functional significance of the additional acidic domains in carp nucleolin remains unclear. Previously, we reported that nucleolin expression is higher in liver cells of winter-adapted carp compared to summer-acclimatized fish, concomitant with repression of rRNA transcription and processing during winter (Álvarez et al., 2003). In this context, the extra-acidic domains of carp nucleolin could be involved in the regulation of the rRNA transcription during fish seasonal adaptation process (data not shown, manuscript in preparation).

Amino acid sequence comparison of cNUCs with those of other vertebrates showed that zebrafish and carp share the higher identity, with RBD and GAR being the most conserved domains (with 82-81% and 92-88% identity, respectively). With *Xenopus*, chicken and mammals, the amino acid sequence identity reaches 66-55% in RBD domains and 79-70% in the GAR region. The evolutionary distances depicted by the phylogenetic tree indicate that carp holds three different nucleolin isoforms that group in two clusters: cNUC1 is separated from cNUC2 and cNUC3. As *C. carpio* is a tetraploid fish, cNUC1 probably diverges

from cNUC2 and cNUC3 as a result of the ancestral fish-specific genome duplication (Taylor et al., 2001), which contributes to increase the number of genes for this and other proteins (Larhammar and Risinger, 1994).

ACKNOWLEDGEMENTS

This work was supported by grant 1040197 from FONDECYT, DI 45-04 and DI 25-03 from the Research Fund of UNAB, ECOS-CONICYT C02B01 and Iniciativa Científica Milenio ICM P04-071-F.

REFERENCES

- ÁLVAREZ M, KAUSEL G, FIGUEROA J, VERA MI (2001) Environmental reprogramming of the expression of protein kinase CK2 beta subunit in fish. *Mol Cell Biochem* 227: 107-112
- ÁLVAREZ M, QUEZADA C, NAVARRO C, MOLINA A, BOUVET P, KRAUSKOPF M, VERA MI (2003) An increased expression of nucleolin is associated with a physiological nucleolar segregation. *Biochem Biophys Res Commun* 301: 152-158
- ÁLVAREZ M, MOLINA A, QUEZADA C, PINTO R, KRAUSKOPF M, VERA MI (2004) Eurythermal fish acclimatization and nucleolar function: A review. *J Therm Biol* 29: 663-667
- BACHELLERIE JP, NICOLOSO M, QU LH, MICHOT B, CAIZERGUES-FERRER M, CAVAILLE J, RENALIER MH (1995) Novel intron-encoded small nucleolar RNAs with long sequence complementarities to mature rRNAs involved in ribosome biogenesis. *Biochem Cell Biol* 73: 835-843
- BACHELLERIE JP, CAVAILLE J (1997) Guiding ribose methylation of rRNA. *Trends Biochem Sci* 22: 257-261
- BOURBON, HM, LAPEYRE B, AMALRIC F (1988a) Structure of the mouse nucleolin gene. The complete sequence reveals that each RNA binding domain is encoded by two independent exons. *J Mol Biol* 200: 627-638
- BOURBON, HM, PRUDHOMME M, AMALRIC F (1988b) Sequence and structure of the nucleolin promoter in rodents: Characterization of a strikingly conserved CpG island. *Gene* 68: 73-84F
- BOURBON HM, AMALRIC F (1990) Nucleolin gene organization in rodents: Highly conserved sequences within three of the 13 introns. *Gene* 88: 187-196
- BOUVET P, DÍAZ JJ, KINDBEITER K, MADJAR JJ, AMALRIC F (1998) Nucleolin interacts with several ribosomal proteins through its RGG domain. *J Biol Chem* 273: 19025-19029
- ERARD MS, BELENGUER P, CAIZERGUES-FERRER M, PANTALONI A, AMALRIC F (1988) A major nucleolar protein, nucleolin, induces chromatin decondensation by binding to histone H1. *Eur J Biochem* 175: 525-530
- FROMONT-RACINE M, SENGHER B, SAVEANU C, FASIOLO F (2003) Ribosome assembly in eukaryotes. *Gene* 14: 17-42

- GINISTY H, AMALRIC F, BOUVET P (1998) Nucleolin functions in the first step of ribosomal RNA processing. *EMBO J* 17: 1476-1486
- GINISTY H, SICARD B, ROGER P, BOUVET P (1999) Structure and functions of nucleolin. *J Cell Sci* 112: 761-772
- KHARRAT A, DERANCOURT J, DOREE M, AMALRIC F, ERARD M (1991) Synergistic effect of histone H1 and nucleolin on chromatin condensation in mitosis: Role of a phosphorylated heteromer. *Biochemistry* 30: 10329-10336
- KISS T (2001) Small nucleolar RNA-guided post-transcriptional modification of cellular RNAs. *EMBO J* 20: 3617-3622
- KITO S, MORIMOTO Y, TANAKA T, HANEJI T, OHBA T (2005) Cleavage of nucleolin and AgNOR proteins during apoptosis induced by anticancer drugs in human salivary gland cells. *J Oral Pathol Med.* 34: 478-485
- LAFONTAINE DL, TOLLERVEY D (1998) Birth of the snoRNPs: The evolution of the modification-guide snoRNAs. *Trends Biochem Sci* 23: 383-388
- LAPEYRE B, CAIZERGUES-FERRER M, BOUCHE G, AMALRIC F (1985) Cloning of cDNA encoding a 100 Kda nucleolar protein (nucleolin) of Chinese. *Nucleic Acids Res* 13: 5805-5816
- LARHAMMAR D, RISINGER C (1994) Molecular genetic aspects of tetraploidy in the common carp *Cyprinus carpio*. *Mol Phylogenet Evol* 3: 59-68
- MERTZ LM, RASHTCHIAN A (1994) PCR radioactive labeling system: A rapid method for synthesis of high specific activity DNA probes. *Focus* 16: 45-48
- MESSMER B, DREYER C (1993) Requirements for nuclear translocation and nucleolar accumulation of nucleolin of *Xenopus laevis*. *Eur J Cell Biol* 61: 369-382
- MOLINA A, CORTA A, MARTÍN RS, ÁLVAREZ M, BURZIO LO, KRAUSKOPF M, VERA MI (2002) Gene structure of the carp fish ribosomal protein L41: Seasonally regulated expression. *Biochem Biophys Res Commun* 295: 582-586
- MOUNT SM (1982) A catalogue of splice junction sequences. *Nucleic Acids Res* 10: 459-72
- NICOLOSO M, CAIZERGUES-FERRER M, MICHOT B, AZUM MC, BACHELLERIE JP (1994) U20, a novel small nucleolar RNA, is encoded in an intron of the nucleolin gene in mammals. *Mol Cell Biol* 14: 5766-5776
- ORRICK LR, OLSON MO, BUSCH H (1973) Comparison of nucleolar proteins of normal rat liver and Novikoff hepatoma ascites cells by two-dimensional polyacrylamide gel electrophoresis. *Proc Nat Acad Sci USA* 70: 1316-1320
- OTAKE Y, SENGUPTA TK, BANDYOPADHYAY S, SPICER EK, FERNANDES DJ (2005) Retinoid-induced apoptosis in HL-60 cells is associated with nucleolin down-regulation and destabilization of Bcl-2 mRNA. *Mol Pharmacol* 67: 319-326
- PINTO R, IVALDI C, REYES M, DOYEN C, MIETTON F, MONGELARD F, ÁLVAREZ M, MOLINA A, DIMITROV S, KRAUSKOPF M, VERA MI, BOUVET P (2005) Seasonal environmental changes regulate the expression of the histone variant macroH2A in an eurythermal fish. *FEBS Lett* 579: 5553-5558
- PLATANI M, LAMOND AI (2003) Nuclear organization and subnuclear bodies. *Prog Mol Sub cell Biol* 35: 1-22
- RANKIN ML, HEINE MA, XIAO S, LEBLANC MD, NELSON JW, DIMARIO PJ (1993) A complete nucleolin cDNA sequence from *Xenopus laevis*. *Nucleic Acids Res* 21: 169
- RASKA I, KOBERNA K, MALINSKY J, FIDLEROVA H, MASATA M (2004) The nucleolus and transcription of ribosomal genes. *Biol Cell* 96: 579-594
- REBANE A, METSPALU A (1999) U82, a novel snoRNA identified from the fifth intron of human and mouse nucleolin gene. *Biochim Biophys Acta.* 1446: 426-430
- ROGER B, MOISAND A, AMALRIC F, BOUVET P (2002) Repression of RNA polymerase I transcription by nucleolin is independent of the RNA sequence that is transcribed. *J Biol Chem* 277: 10209-10219
- ROGER B, MOISAND A, AMALRIC F, BOUVET P (2003) Nucleolin provides a link between RNA polymerase I transcription and pre-ribosome assembly. *Chromosoma* 111: 399-407
- ROGERS J (1985) Exon shuffling and intron insertion in serine protease genes. *Nature* 315: 458-459
- RUSKIN B, GREEN MR (1985) Role of the 3' splice site consensus sequence in mammalian pre-mRNA splicing. *Nature* 317: 732-734
- SÁEZ L, ZUVIC T, AMTHAUER R, RODRÍGUEZ E, KRAUSKOPF M (1984) Fish liver protein synthesis during cold acclimatization: Seasonal changes of the ultrastructure of the carp hepatocyte. *J Exp Zool* 230: 175-186
- SAITOU N, NEI M (1987) The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4: 406-425
- SAN MARTÍN R, CÁCERES P, AZÓCAR R, ÁLVAREZ M, MOLINA A, VERA MI, KRAUSKOPF M (2004) Seasonal environmental changes modulate the prolactin receptor expression in an eurythermal fish. *J Cell Biochem* 92: 42-52
- SRIVASTAVA M, FLEMING PJ, POLLARD HB, BURNS AL (1989). Cloning and sequencing of the human nucleolin cDNA. *FEBS Lett* 250: 99-105
- SRIVASTAVA M, MCBRIDE OW, FLEMING PJ, POLLARD HB, BURNS AL (1990) Genomic organization and chromosomal localization of the human nucleolin gene. *J Biol Chem* 265: 14922-14931
- SRIVASTAVA M, POLLARD HB (1999). Molecular dissection of nucleolin's role in growth and cell proliferation: New insights. *FASEB J* 13: 1911-2200
- TAYLOR JS, VAN DE PEER Y, BRAASCH I, MEYER A (2001) Comparative genomics provides evidence for an ancient genome duplication event in fish. *Philos Trans R Soc Lond B Biol Sci* 356: 1661-1679
- THOMPSON JD, GIBSON TJ, PLEWNIK F, JEANMOUGIN F, HIGGINS DG (1997) The Clustal X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 24: 4876-4882
- VERA MI, KAUSEL G, BARRERA R, LEAL S, FIGUEROA J, QUEZADA C (2000) Seasonal adaptation modulates the expression of the protein kinase CK2 beta subunit gene in the carp. *Biochem Biophys Res Commun* 271: 735-740
- VERA MI, MOLINA A, PINTO R, REYES M, ÁLVAREZ M, KRAUSKOPF E, QUEZADA C, TORRES J, KRAUSKOPF M (2003) Genomic organization of the rDNA cistron of the teleost fish *Cyprinus carpio*. *Biol Res* 36: 241-51