



www.elsevier.com/locate/gene

Molecular cloning of an Atlantic salmon nucleoside diphosphate kinase cDNA and its pattern of expression during embryogenesis

Gene 257 (2000) 139-148

Madeline Murphy^{a,1}, Tina Harte^{a,2}, James McInerney^b, Terry J. Smith^{a,*}

^a BioResearch Ireland, National Diagnostics Centre, National University of Ireland, Galway, Ireland ^b National University of Ireland, Maynooth Co. Kildare, Ireland

Received 26 February 2000; received in revised form 26 July 2000; accepted 14 August 2000 Received by M. Schartl

Abstract

To gain insight into the process of development in Atlantic salmon (*Salmo salar*), we sought to identify genes that were differentially expressed at gastrulation. A polymerase chain reaction-based differential screening strategy allowed for the isolation of an Atlantic salmon nucleoside diphosphate kinase cDNA (nm23). Structural characterisation showed a high degree of homology with a large number of previously isolated nucleoside diphosphate kinases (NM23s), both prokaryote and eukaryote, though it represents the first teleost nucleoside diphosphate kinase identified. Highest similarities were found with the type 1 and type 2 NM23 isoforms of mammals. Phylogenetic analysis indicates that the duplication event that gave rise to these isoforms occurred after the splitting of tetrapods and fish, suggesting that the salmon NM23 represents a more ancestral isoform. The position of the salmon sequence on the phylogenetic tree indicates that the salmon genome is expected to have at least three copies of genes from the nm23 gene family. Northern blot analysis showed a single transcript of approximately 0.7 kb in both embryonic and adult tissues. Examination of the temporal pattern of expression of salmon nucleoside diphosphate kinases are thought to have a vital role in regulatory processes such as signal transduction, proliferation and differentiation. Taken together, these results suggest that nucleoside diphosphate kinases have an important role to play in early embryogenic development in vertebrates. © 2000 Elsevier Science B.V. All rights reserved.

Keywords: awd; cDNA; Development; Differential expression; NDPK; NM23; Salmo salar; Teleost

1. Introduction

Development of an organism proceeds from a fertilised egg through a predetermined sequence in which

^{*} The nucleotide sequence of salmon nucleoside diphosphate kinase reported in this paper has been submitted to the GenBank database at NCBI and assigned the Accession No. AF045187.

fax: +353-91-5865701.

specific genes are activated and inactivated. These changes in gene expression lead to pattern formation, differentiation and changes in form. In order to elucidate the molecular regulation of developmental processes, genes differentially expressed in this regulatory cascade must be identified, characterised and studied in detail. Vertebrate gastrulation is a dynamic period of development characterised by extensive cell migrations. This stage of development is likely to require the expression of a new genetic repertoire to initiate and direct these dramatic changes.

An important aspect of understanding developmental processes is to examine and compare embryogenesis in different species. As the most primitive group among vertebrates, fish provide an excellent model for evolutionary comparison (Colbert, 1969; Powers, 1991).

Here, we used differential hybridisation to identify genes that are differentially regulated at the time of gastrulation in Atlantic salmon (*Salmo salar*).

Abbreviations: A, adenosine; aa, amino acid(s); bp, base pair(s); cDNA, DNA complimentary to RNA; cpm, counts per minute; dNTP, deoxyribonucleoside triphosphate; ds, double-stranded; kb, kilobase(s) or 1000 bp; Leu, leucine; NDPK, nucleoside diphosphate kinase; oligo, oligodeoxyribonucleotide; ORF, open reading frame; PCR, polymerase chain reaction; pfu, plaque-forming unit(s); rRNA, ribosomal RNA; SDS, sodium dodecyl sulfate; *UTR*, untranslated region(s).

^{*} Corresponding author. Tel.: + 353-91-586559;

E-mail address: terry.smith@nuigalway.ie (T. Smith)

¹ Present address: Biotechnology Centre, University College Dublin, Belfield, Dublin 4, Ireland.

² Present address: Forensic Science Laboratory, Garda Headquarters, Phoenix Park, Dublin 8, Ireland.

Differential hybridisation of cDNA libraries has been widely used to identify genes that are differentially expressed (Sambrook et al., 1989), and recently, this method has been adapted to a more rapid two-step PCR based technique (Luo et al., 1994).

In this paper, we report the isolation of a cDNA encoding an Atlantic salmon nucleoside diphosphate kinase (*Ss*NM23) in a differential screen for genes induced at gastrulation. Nucleoside diphosphate kinases (EC 2.7.4.6) comprise a ubiquitous class of enzymes that exist in multiple isoforms and are highly conserved throughout evolution. These enzymes regulate growth and development and have been implicated in the pathogenesis and metastasis of tumors (Postel, 1998). Analysis of the gene expression pattern of Atlantic salmon *nm23* at different stages of early embryo development suggests a role for NM23 in the developmental process.

2. Materials and methods

2.1. Sampling of Atlantic salmon embryos and adult tissues

Salmon embryos were obtained from Cong Fish Hatchery, Cong Co. Mayo. Salmon eggs were collected from trapped wild salmon, fertilised and placed in water trays. Embryos were collected at regular intervals following fertilisation, a sample of each collection was immediately fixed in 4% paraformaldehyde, and the remainder was frozen directly in liquid nitrogen. The stage of embryonic development was determined following removal of shell and microscopic examination. Tissues were collected from adult Atlantic salmon and frozen directly in liquid Nitrogen prior to RNA isolation.

2.2. RNA preparation and cDNA library construction

Total cellular RNA was isolated by the single-step acid guanidium thiocyanate-phenol-chloroform extraction (Chomcynski and Sacchi, 1987) with an additional LiCl precipitation to remove egg proteins (Sambrook et al., 1989). Poly(A)⁺ mRNA was isolated by affinity chromatography to oligo (dT) (Aviv and Leder, 1972). Total cDNA was synthesized from 5 µg of mRNA isolated from stage 14–15 embryos and stage 20–21 embryos (Battle, 1944) using the Pharmacia cDNA synthesis kit as directed by the manufacturer. The cDNAs were used to create libraries in the λ Zap II vector (Stratagene) according to the manufacturer's instructions.

2.3. Differential hybridisation and PCR screening

The differential hybridisation was performed essentially as described by Luo et al. (1994). Approximately

 10^5 plaque-forming units (pfu) of the stage 20–21 library were plated at low density (10⁴ pfu/150 mm plate) and transferred in duplicate to Nytran membranes. One filter set was probed with 1×10^6 cpm/ml of ³²P-labelled cDNAs prepared from mRNAs isolated from stage 14-15 salmon embryos, while the duplicate set was probed with 1×10^6 cpm/ml of ³²P -labelled cDNAs prepared from mRNA isolated from stage 20-21 salmon embryos. The filters were hybridised at 65°C for 14 h and washed to $0.1 \times$ SSC and 0.1% SDS at 65°C. Following exposure to X-ray film at -70° C, plaques that appeared differentially expressed were picked with sterile toothpicks into 100 µl of SM (100 mM NaCl, 10 mM MgSO₄, 50 mM Tris-Cl, pH 7.5, 0.001% gelatin) buffer (Sambrook et al., 1989). These were stored at 4°C in 1 ml of phage buffer (Sambrook et al., 1989). Three microlitres of phage plaques from each of 20 primary positives were used as template for the PCR reaction. Tag polymerase was obtained from Promega. The primers used for the PCR reaction were T3 and T7, cycling conditions were as follows; 70°C for 5 min, 4°C for 5 min, 94°C for 5 min followed by 30 cycles of 94°C for 1 min, 45°C for 1 min and 70°C for 1 min. Equal amounts (20 µl per well) of each of the PCR products were run on duplicate 1% agarose gels and blotted onto Hybond N filters. The filters were hybridised with two cDNA probes as described for the primary screen. PCR products showing differential hybridisation patterns were further characterised by sequencing.

2.4. DNA sequence determination

Plasmid was excised from bacteriophage using ExAssist helper phage (Stratagene). DNA sequence analysis was performed manually using T7 polymerase (Pharmacia) as directed by the manufacturer.

2.5. Northern blot analysis

Northern blots were performed as described previously (Ausubel et al., 1994). The probes used were the salmon *nm23* cDNA and the salmon L18a cDNA. cDNA probes were labelled with ³²P-dCTP using the HighPrime labelling kit (Boehringer).

2.6. Nomenclature used for nucleoside diphosphate kinase genes and proteins

At present, there is no consensus terminology for the genes encoding nucleoside diphosphate kinases, (NM23s, NDPKs, NMEs). For clarity, throughout this paper, we have adopted a single uniform nomenclature for nucleoside diphosphate kinase genes and their encoded proteins. All genes are indicated in the text and figures as *nm23*. Genes encoding known mammalian isoforms of NM23 also have a letter code, indicating

whether they are human (H), mouse (M) or rat (R) followed by a number representing the type where known (e.g. *nm23-H1*, *nm23-M1* and *nm23-R1* for the human, mouse and rat type 1 genes respectively). Throughout the text, the corresponding nucleoside diphosphate kinase deduced amino acid sequences and proteins are denoted as NM23, followed by the same letter and numerical code for the various mammalian protein isoforms (e.g. NM23-H1, NM23-M1 and NM23-R1 for the human, mouse and rat type 1 proteins, respectively). The numerical codes used for the mammalian isoforms of NM23 correspond with the NME nomenclature system used by the National Centre for BioInformatics, NIH. The *Salmo salar* NM23 protein is abbreviated as *Ss*NM23.

3. Results and discussion

3.1. Determination of developmental stage of salmon embryos

The developmental stage for five randomly selected embryos from each sample collection was determined by microscopic analysis of embryos. Developmental stage was assigned as described previously (Battle, 1944; see Table 1).

3.2. Isolation of a cDNA clone for a differentially expressed embryonic salmon gene

A differential screen of 10⁴ pfu of the stage 20–21 salmon cDNA library identified 18 cDNA clones to be more abundant in stage 20–21 salmon embryos than in stage 14–15 salmon embryos. The positive pools of phage candidates following the primary screen were

Table 1

Developmental stage of Atlanic salmon embryonic collections^a

Days post- fertilisation	Description of embryo	Developmental stage ^b
2	Many large cells	10-11
5	Blastodisk (1.4–1.5 mm)	11-12
7	Blastodisk (1.5 mm)	12-13
9	Blastodisk (1.6-1.7 mm)	13-14
11	Blastodisk (1.6-2.0 mm)	14–15
14	Germ ring and shield	15-17
16	Blastoderm, differentiation at anterior of shield	17–18
18	Blastoderm overgrown 50% of yolk	18-19
21	Blastoderm overgrown 75% of yolk	20-21
23	Yolk sac complete, blastopore closed	22
90	Yolk sac fry	34

^a Atlantic salmon embryos collected at fixed time points following fertilisation were assigned a developmental stage according to Battle (1944).

^b Battle (1944).

subjected to PCR using T3 and T7 primers, which flank the cloning site in the λ ZAPII vector. The PCR products were then equally loaded and run on duplicate gels. DNA in the duplicate gels were blotted onto two Hybond-N membranes and hybridised with ³²P-labelled cDNA probes prepared from stage 14–15 and 20-21 salmon embryos. Eleven of 18 primary positive clones contained cDNA inserts, which hybridised more strongly to cDNA from stage 20-21 embryos than to cDNA from stage 14-15 embryos (data not shown). Five of these were approximately 0.7 kb in size; these were partially sequenced and were found to contain the same cDNA insert. Sequence analysis of the remaining six clones revealed that three encoded novel genes, two encoded the mitochondrial enzyme cytochrome C-oxidase III, and one encoded ribosomal protein L18a.

3.3. Analysis of salmon nm23 cDNA sequence

The entire sequence of the Atlantic salmon nm23 cDNA was determined and found to consist of 720 nucleotides. The clones contained a 5' untranslated region (UTR) of 24 bases followed by an open reading frame coding for 151 amino acids, a 3'UTR of 240 nucleotides in length followed by a tract of adenosine (A) residues (Fig. 1). Comparison of the salmon cDNA sequence with sequences held in GenBank/EMBL revealed a significant homology with a large number of NM23 sequences at the amino acid level (Table 2). The best homologies were with vertebrate sequences. The deduced amino acid sequence of the SsNM23 was highly similar to NM23-H1 (82% identity) and NM23-H2 (80% identity). Both NM23-H1 and NM23-H2 can function as transcription factors (Postel et al., 1993; Chae et al., 1998), suggesting that SsNM23 may also play a role in transcriptional regulation. A high homology was also found with all three NM23 isoforms of Xenopus (77% identity) and with the Drosophila NM23-awd (71% identity) (Table 2).

PROSITE domain analysis of the deduced amino acid sequence of the salmon nm23 cDNA sequence identified six protein sequence motifs (Fig. 2). These were two protein kinase C phosphorylation sites, a casein kinase II (CK2) phosphorylation site, a tyrosine kinase phosphorylation site, an N-myristoylation site, a cell attachment sequence and the nucleoside kinase active site. Apart from the CK2 phosphorylation site, these motifs are conserved in NM23-H1 and NM23-H2, in Xenopus NM23 and in Drosophila NM23-awd (figure 2). These motifs suggest that SsNM23 is involved in the mediation of extracellular signals to the nucleus. PKCs and CK2 are serine threonine protein kinases. CK2 is required for cell-cycle progression (Blanquet, 2000), and PKCs play key roles in extracellular signal transduction (Murphy et al., 1998). In addition, many growth factors (i.e. PDGF, EGF) signal via tyrosine kinase phosphorylation from cell surface to nucleus.

1	CGC	GGC(CGC'.	FAG'. +	ГСА'.	FCA/	AGC(GCCI	ATG	ГСА/ +	AACO	GAG(GAG(CGG# +	4CTI	ГТС <i>і</i>	4TΤ(-+-	GCC <i>I</i>	ATCA	\AG +
									М	s	N	Е	Е	R	т	F	I	А	I	K
61	CCA	GAT	GGA	.GTT	CAG	AGA	AGG	CTT	GTC	GGC	GAT	ATC	ATC	AAG	AGA	TTT	GAG	CTG	AAG	GGC
01	Ρ	D	G	v	Q	R	R	L	v	G	D	I	I	к	R	F	Ē	L	ĸ	G
121	TTC	AAA	ATG	GTG +	GGG	ATG	AAA -+-	.TTC	ATC	'AAG	GCC	CCA	.GAG	тст +	CTG	CTG	AAG -+-	GAG	CAC	ТАТ +
	F	к	М	v	G	М	к	F	I	K	A	Ρ	Е	s	\mathbf{L}	L	К	Е	н	Y
181	GCC	GAC	CTG	AAG +	GAC	AGA	.ccc	TTC	TTC	сст	GGT	стс	GTC	AGC +	TAC	ATG	ACC -+-	ТСА 	.GGC	CCA +
	A	D	L	к	D	R	Ρ	F	F	Ρ	G	L	V	s	Y	М	т	S	G	Ρ
241	GTG	GTG	GCT	ATG +	GTG	TGG	GAA -+-	.GGG	TTC	ТАА +	GTG	GTG	AAG	ACA +	GGC	CGA	GTG -+-	ATG	стс	GGA +
	v	v	A	М	v	W	Е	G	F	N	v	v	K	т	G	R	v	М	\mathbf{L}	G
301	GAG	ACC	AAC	ССТ +	GCT	GAC	TCC	AAA	.ccc	GGC	ACC	АТС 	CGA	GGA +	GAC	'TTC	TGC -+-	АТС 	CAA	GTG +
	Е	т	N	Ρ	A	D	S	K	Ρ	G	т	I	R	G	D	F	С	I	Q	v
361	GGCAGGAACATCATCCATGGCAGTGACTCAGTAGAGAGTGCTAACACAGAGATCAATCTG																			
501	G	R	N	I	Ι	Н	G	S	D	S	V	Е	S	A	Ν	т	Ē	I	N	г.
421	TGG	TTC	AAA 	.ССТ +	GAG	GAG	СТG -+-	TGC	AGT	ТАС +	ACT	AGC	TGC	тсс +	AGC	CAG	TGG -+-	СТС	TAC	TGA +
	W	F	K	Ρ	E	Е	L	С	S	Y	т	S	С	S	S	Q	W	L	Y	*
481	GGA	GCT	AAA 	.GGC +	TAC	AAG	TTG -+-		CCA	.GGC +	ACA	GGT	ста	GGA +	CCA	GTT 	TAC -+-	CAT 	CCA	CAA +
541	ACC	ATT 	AGC	ATT +	AGA	AGG	GGG -+-	ААТ 	'АТА	ААТ +	CTC	TTC 	CTG	GCC +	ACT 	CTG	GAC -+-	CAC	ACC	ACT +
601	TCC	CGT	TCA 	TCT +	TCT	GTA	.ccc	АТС 	TGT	'AA'I +	'TGG	TCC	атс 	GCT +	GAT	TCC	ATT -+-	CAT	TCC	тст
661	GCA	.CGT	ACA	.GTT +	GTC	АТА	.GTG	AAT	ATG	TTT +	'TTT	ATT	GTC	ATT +	TGT 	ТСА 	АСТ -+-	TTC	CAA	TAA +

Fig. 1. Nucleotide and deduced amino acid sequence of the salmon nm23 cDNA clone. The GenBank/EMBL Accession No. is AF045187.

N-myristoylation is an acylation process in which a protein associates with the membrane (Boutin, 1997). These motifs suggest that *Ss*NM23 is attached at the cell/nuclear membrane and is activated by a number of kinases. They indicate that *Ss*NM23 functions as a mediator of cell signalling in addition to its role as an active enzyme. This is consistent with the proposed role of mammalian NM23s as regulators of cell growth (Postel, 1998). The high degree of conservation of these protein sequence motives in both vertebrates and invertebrates suggests that they are important in NM23 function.

NM23s have been cloned from a wide variety of prokaryotes and eukaryotes; they perform a 'housekeeping' role in maintaining the intracellular levels of dNTPs, which are used for the synthesis of nucleic acids as constituents, polysaccharides and lipids as cofactors and protein as regulators (Parks and Agawal., 1973). NM23s are proteins consisting of four to six identically folded subunits of approximately 16–20 kDa (Postel, 1998). They act via a ping-pong mechanism in which a histidine residue is phosphorylated by transfer of the terminal phosphate group from ATP. In the presence of magnesium, the phospho enzyme can transfer its phosphate group to any NDP, to produce an NTP. It is thought that multiple forms of NM23 are targeted to different regions of the cell where they fulfil distinct metabolic and regulatory roles (Postel, 1998).

To date, seven isoforms of NM23 have been identified in humans. NM23-H1 is a suppressor of metastasis in some tumor types (Delarosa et al., 1995) and may function as a transcription factor (Chae et al., 1998). NM23-H2 has been shown to function in vitro in the transcriptional regulation of c-myc expression (Postel et al., 1993). NM23-H1 and NM23-H2 show 88% identity (Stahl et al., 1991). Dr-NM23 shares 70%

Table 2 Sequence identity between SsNM23 and NM23 homologues from other species^a

Sequence	Accession	Percentage similarity amino acid identities				
NM23-H1	NM_000269	82				
NM23-H2	NM_002512	80				
Gallus gallus	AF043542	80				
Xenopus laevis A1	X97900	77				
Xenopus laevis B2	X97901	77				
Xenopus laevis A2	X97902	77				
Ginglymostoma cirratum	M63964	74				
Drosophila melanogaster	X13107	71				
Dictyostelium discoideum	P22887	60				
DR-nm23	U80813	62				
NM23-H4	NM_005009	58				
NM23-H5	NM_003551	28				
NM23-H6	NM_005793	28				
NM23-H7	AF153191	29				

^a Identity of the deduced amino acid sequence of the salmon cDNA was compared with Human, *X. laevis*, *D. melanogaster*, *G. cirratum* and *D. discoideum* NM23s.

identity with NM23-H1 and NM23-H2 (Venturelli et al., 1995). Its overexpression inhibits granulocyte differentiation and induces apoptosis (Venturelli et al., 1995). NM23-H4 shows between 55 and 60% identity to the other isoforms and encodes an NM23 with an apparent presequence that has some characteristics consistent with import into mitochondria (Milon et al., 1997). NM23-H5 is highly expressed in testis and shares between 27 and 31% identity to the other isoforms. NM23-H5 is thought to be involved in the early stages of spermatogenesis (Munier et al., 1998). NM23-H6 shares 34–61% identity with the other human isoforms and is expressed at a moderately low level in many human tissues (Mehus et al., 1999). A seventh NM23 isoform (NM23-H7) has recently been identified (Accession No. AF153191).

The Atlantic salmon *nm23* cDNA deduced amino acid sequence shows a significant similarity to many previously isolated NM23 enzymes, with the highest similarities to the type 1 and 2 isoforms in mammals. The high level of evolutionary conservation between NM23 sequences indicates that this protein is involved in essential functions in the cell.

3.4. Phylogenetic analysis of Atlantic salmon NM23

A similarity search of GenBank was carried out using the position-specific Iterative BLAST algorithm (Ψ -BLAST). The algorithm was iterated until no new entries were found following subsequent runs — the algorithm had converged. A total of 131 GenBank entries were identified. Unfortunately, not all of these entries were full-length, and it was decided that phylogenetic analysis of these sequences would not be a profitable exercise. A dataset of 83 entries were selected for further analysis.

SsNM23 NM23-H1 NM23-H2 Xenopus A2 Drosophila	1 1 1 1	MS NEERTFIAIK PDGVQRRLVG cceprgsrar fgcwrlqpef kpkqlegtMA CAI. MA .LAI. MI
SSNM23 NM23-H1	23 51 23	DIIKRFELKG FKMVGMKFIK APESLLKEHY ADLKDRPFFP GLVSYMTSGP E
Xenopus A2 Drosophila	23 22 24	D
SSNM23 NM23-H1 NM23-H2 Xenopus A2 Drosophila	73 101 73 72 74	VVAMVWEGFN VVKTGRVMLG ETNPADSKPG TIRGDFCIQV GRNIHGSDS .LAVK S S .LAVK S .LAVK S .LAVK S .LAVK S .LAVK S .LAVK S
SSNM23 NM23-H1 NM23-H2 Xenopus A2 Drosophila	123 151 123 122 124	VESANTEINL WFKPEELCSY TSCSSQQWLY EKGHPEVDY TSCAQN.I EKSKPEVDY KSCAHD.V NKAKDEVEN KSCAYE.VN E. ANEK.VTW TPDAKD.T -

Fig. 2. Alignment of the salmon *nm23* cDNA deduced amino acid sequences with *nm23* sequences held in the database Sequences are from *nm23-H1* (Accession No. X17620) and *nm23-H2* (Accession No. L16785). *Xenopus laevis nm23* A2 (Accession No. X97902) and *Drosophila melanogaster nm23-awd* (Accession No. X13107). Sequences were aligned using the Dialign program (Morgenstern et al., 1996). Amino acids given in lower case show no conservation between the five sequences examined. Protein motifs for protein kinase C phosphorylation site (86–88, 103–106), CK2 phosphorylation site (2–5), tyrosine kinase phosphorylation site (135–142), N-myristoylation site (102–107), cell attachment sequence (105–107) and the nucleoside kinase active site (115–123) are underlined.

Alignment of the 83-sequence dataset was carried out using the ClustalX 1.8 software for progressive alignment (available from ftp://ftp.ebi.ac.uk/). The default options were employed, and upon inspection, the alignment of most positions appeared to be reasonably robust. There were only three sites in the alignments that were constant across all sequences. The dataset was further manipulated using the se-al software for the Apple Macintosh, which facilitated the removal of those positions whose positional homology could not be reasonably well guaranteed. For the purposes of phylogenetic reconstruction, it is necessary to use only those characters for which a reasonably well-supported hypothesis of positional homology can be inferred. The final alignment was 156 positions in length.

A preliminary analysis of the 83-sequence dataset indicated that a number of taxa were present whose analysis was not relevant for the present study. In such cases, most of the sequences were removed from the analysis, and a single exemplar sequence was retained. These taxa included many fungi, prokaryotes and plant sequences. The final alignment consisted of 43 sequences from a diverse group of organisms.

The PUZZLE software program (version 4.0.2) (Strimmer and VonHaeseler, 1996) was used in order to calculate the maximum likelihood distance matrix under the JTT substitution model (Jones et al., 1992). Phylogenetic hypotheses based upon these distances were inferred using the neighbor-joining method, as implemented in the NEIGHBOR program of the PHYLIP package (Felsenstein, 1993). Trees were visualised using the Treeview program (Page, 1996). The robustness of the inferred relationships was evaluated using the bootstrap method (Felsenstein, 1985) and was carried out using the puzzleboot shell script and the SEQBOOT and CONSENSE programs of the PHYLIP package.

Fig. 3 shows the phylogenetic tree recovered by neighbor-joining. This tree places the salmon sequence as an early-branching member of a clade that includes other vertebrate sequences. Bootstrap support for this clade is low (18), but this topology is recovered using the original data matrix and in a majority-rule consensus of the bootstrapped data. In Fig. 3, this is designated 'Clade A'. Within this clade are sequences from reptiles, birds and mammals. There are two groups of human sequences within Clade A, and within each of these groups are two separate sequences. It is not possible from sequence annotation to say whether these are simply allelic variants or genuine duplications, although the former is more likely.

There are a total of seven human sequences in the dataset, and using these data combined with a phylogenetic analysis, it is possible to predict the minimum number of paralogs that might be present in the salmon genome.

The two groups of human sequences in clade A (nm23-HI from one group, nm23-H2 and nm23 (NDK6) from the other group) appear to have arisen by duplication after the speciation events that gave rise to modern teleost fish and tetrapods. The common ancestor of these four sequences appears to have existed after the splitting of tetrapods and fishes. This duplication event would not be expected to be found in the salmon genome. It is most parsimonious to assume that the salmon genome would have one copy of this particular paralog.

The human sequences designated numbers nm23-H4 and DR-nm23 in Fig. 3 share a common ancestor with some other vertebrate sequences. However, these particular paralogs share a recent evolutionary history that is independent of the other human paralogs. Within the human genome, these two sequences are each other's closest relatives, but the duplication event that gave rise to these two lineages predates the splitting of birds and mammals at the very least. The salmon genome would be expected to contain at least one paralog from this family. The sparse sampling of this evolutionary lineage means that an exact interpretation of this lineage is not possible. If the duplication of genes 2 and 3 postdates the separation of fish and tetrapods, then the salmon genome would be expected to only have a single copy of this 'family'. If, however, the duplication predates the separation of fish and tetrapods, then fish would only be expected to have two copies.

The seventh human sequence is very different to the rest of the paralogs. Sequence similarity is quite low, and almost all pairwise comparisons post more than one substitution per site since separation from the common ancestor. It is likely that this is the oldest duplication in this gene family (a supposition that relies on the existence of a molecular clock) and predates the separation of eukaryotes and prokaryotes. The salmon genome would be expected to contain an ortholog of this sequence.

We can postulate from the position of the salmon sequence on the phylogenetic tree that the salmon genome would be expected to have at least three copies of genes from the nm23 gene family. Deviations from this number would depend on polyploidisation of the genome or gene loss or, as mentioned above, whether or not duplication of human genes 2 and 3 pre- or postdated the speciation event that has led to modern tetrapods and fishes.

3.5. Pattern of expression of salmon NM23 during embryonic development

In order to examine the expression pattern of Atlantic salmon NM23 during early development, a Northern blot analysis was performed (Fig. 4). A transcript of approximately 0.7 kb was detected in stage 15–17 and



Fig. 3. Phylogenetic analysis of *nm23* cDNA deduced amino acid sequences. This tree was constructed using the neighbor-joining method based on maximum likelihood distances under the JTT model of sequence evolution. The salmon sequence, along with seven human paralogs, are indicated by the rounded boxes. The smallest clade that includes the salmon sequence is indicated as 'Clade A'. Numbers on internal branches are bootstrap proportions (BPs). To aid clarity, not all internal branches are labelled with BPs. Accession Nos for all sequences used in the alignments shown are available on request. *The sequence identified as Human *nm23 (NDK6)* was obtained from Swissprot (accession number 060361) and is likely to be a variant of human *nm23-H2*.

in all subsequent stages but not in stages 11-12 (Fig. 4A and B), stage 14–15 or unfertilised eggs (data not shown). The *nm23* gene transcript remained abundant in all embryonic stages examined following stage 15–17 and in yolk sac fry. Northern hybridisation did not detect *nm23* mRNA in fertilised eggs until after gastrulation, indicating that there is little, if any, de-novo synthesis of this message prior to this stage of development.

A similar pattern of expression has been described for *Xenopus laevis* NM23 during early development (Ouatas et al., 1998). In *X. laevis*, NM23s are expressed early and transiently during mesoderm and neural induction and are thought to be involved in cell fate establishment or totipotency maintenance. In addition, *Xenopus* NM23s are abundantly expressed in differentiated structures, especially neural derivatives. NM23 has been implicated in neurite outgrowth (Gervasi et al., 1996),



Fig. 4. Analysis of nm23 mRNA levels by Northern blot hybridisation. (A) Northern blot of RNA from (1) stage 15–17, (2) stage 18–19, (3) stage 20–21. (B) (1) stage 11–12, (2) stage 20–21, (3) stage 22, (4) stage 34. Shown below the autoradiographs are ethidium bromide stained agarose gels of RNA to account for total RNA levels. (C) Northern blot of RNA from (1) liver (2) kidney (3) brain (4) intestine from adult salmon. The blot was probed with salmon nm23 cDNA, followed by reprobing with ribosomal protein L18a to normalise for RNA levels.

and it is has been proposed that it may modulate neural cell proliferation and differentiation in development (Ouatas et al., 1998).

In mammals, NM23s have also been shown to have an important role in early development, it has been demonstrated that accumulation of NM23 protein, NM23-M1, is coincident with the functional differentiation of multiple epithelial tissues in the developing mouse fetus (Lakso et al., 1992). NM23-M1 was originally identified as a candidate tumor suppressor protein (Steeg et al., 1988). Reduced human NM23 (NM23-H1 and NM23-H2) expression has been correlated with high metastatic potential in breast, ovarian, cervical, gastric and hepatocellular carcinomas and melanoma (Delarosa et al., 1995). These results suggest that common regulatory elements underlie embryonic development and metastasing tumor cells and that NM23 may play an important regulatory role. The biochemical mechanisms by which NM23s regulate tumor potential and differentiation are unknown.

The type 1 and 2 mammalian isoforms and the other vertebrate isoforms show a high degree of homology with the *Drosophila awd* (abnormal wing discs) gene (Rosengard et al., 1989), shown to be essential for normal fly development (Dearolf et al., 1988a; Biggs et al., 1990). *Awd* is a microtubule-associated NM23 (Rosengard et al., 1989).

Differential expression of *Drosophila awd* during early embryogenesis has been described (Dearolf et al., 1988a). Transcription of zygotic *awd* was detected during the second larval instar after mesoderm induction, while mutation or reduced *awd* expression resulted in aberrant differentiation after metamorphosis (Biggs et al., 1988; Dearolf et al., 1988a,b). Studies using Null mutants in *awd* indicated that the catalytic histidine is necessary, but not sufficient, for its biological function, and thus *awd* has other activities in vivo besides its nucleoside diphosphate kinase function (Xu et al., 1996). As the expression of NM23s during or closely following mesoderm induction is conserved in both vertebrates and invertebrates, it is likely that NM23 plays a critical role in these early developmental events.

NM23-H2 has been shown to function as the transcriptional regulator of c-myc (Postel et al., 1993). The temporal expression pattern of c-myc has been examined during embryogenesis in the zebrafish (*Danio rerio*) (Schreiber-Agus et al., 1993). As with NM23 in the salmon embryos, c-myc was not expressed in the earliest stages of development, was first expressed following gastrulation, and was expressed at highest levels in the later stage of development particularly during periods of active cellular proliferation. Taken together, these facts suggest that NM23 may be regulating growth and differentiation at least in part through control of c-myc expression.

The salmon *nm23* transcript was also detected in four adult tissues examined (Fig. 4C). The transcript was most abundant in RNA isolated from salmon brain. NM23s have been implicated in neural cell differentiation (Amendola et al., 1997; Ouatas et al., 1997), and it is possible that the salmon NM23 may have a similar function.

In summary, we report the isolation and analysis of

salmon *nm23* cDNA whose expression is elevated during early embryonic development. Significant homologies were found in the protein coding sequences with a number of NM23 sequences at the amino acid level. These proteins, which exist in a number of closely related isoforms, have been implicated in the metastatic potential of mammalian cancers and in normal tissue development and differentiation. The differential gene expression pattern identified in developing salmon embryos suggests a role for NM23 in early salmonid development.

4. Note

Since the submission of this manuscript, the isolation of a Zebrafish NM23-B has been reported (Lee and Lee, 2000).

Acknowledgements

The authors would like to thank Cong Fish Hatchery, Co. Mayo, Ireland, for supplying salmon embryos. This research was funded by the EU Biotechnology programme (Contact number BIO2 CT 93-0430) and by BioResearch Ireland.

References

- Amendola, R., Martinez, R., Negroni, A., Venturelli, D., Tanno, B., Calabretta, B., Raschella, G., 1997. Dr-NM23 gene expression in neuroblastoma cells: relationship to integrin expression, adhesion characteristics and differentiation. J. Natl. Cancer Inst. 89, 1300–1310.
- Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., Struhl, K., 1994. Current Protocols in Molecular Biology. Wiley, New York.
- Aviv, H., Leder, P., 1972. Purification of biological active globin message RNA by chromatography on oligothymidylic acid-cellulose. Proc. Natl. Acad. Sci. USA 69, 1408–1412.
- Battle, H., 1944. The embryology of the atlantic salmon (Salmo salar Linnaeus). Can. J. Res. 22, 105–125.
- Biggs, J., Tripoulas, N., Hersperger, E., Dearolf, C., Shearn, A., 1988. Analysis of the lethal interaction between the *prune* and *killler* of *prune* mutations of *Drosophila*. Genes Dev. 2, 1333–1343.
- Biggs, J., Hersperger, E., Steeg, P.S., Liotta, L.A., Shearn, A., 1990. A *Drosophila* gene that is homologous to a mammalian gene associated with tumour metastasis codes for a nucleoside diphosphate kinase. Cell 63, 933–940.
- Blanquet, P.R., 2000. Casein Kinase 2 as a potentially important enzyme in the nervous system. Prog. Neurobiol. 60, 211–246.
- Boutin, J.A., 1997. Myristoylation. Cell Signal 9, 15-35.
- Chae, S.K., Lee, N.S., Lee, K.J., Kim, E., 1998. Transactivation potential of the C-terminus of human NM23-H1. FEBS Lett. 423, 235–238.
- Chomcynski, P., Sacchi, N., 1987. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal. Biochem. 162, 156–159.
- Colbert, E.H., 1969. . In: Wiley, A. (Ed.), Evolution of the Vertebrates. Academic Press, New York, pp. 1–24.
- Dearolf, C.R., Hersperger, E., Shearn, A., 1988a. Developmental con-

sequences of *awd*b3, a cell autonomous lethal mutation of *Drosoph-ila* induced by hybrid dysgenesis. Dev. Biol. 129, 159–168.

- Dearolf, C.R., Tripoulas, N., Biggs, J., Shearn, A., 1988b. Molecular consequences of *awd*b3, a cell-autonomous lethal interaction mutation of *Drosophila* induced by hybrid dysgenesis. Dev. Biol. 129, 169–178.
- Delarosa, A., Williams, R.L., Steeg, P.S., 1995. NM23/nucleoside diphosphate kinase: toward a structural and biochemical understanding of its biological functions. Bioessays 17, 53–62.
- Felsenstein, J., 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39, 783–791.
- Felsenstein, J., 1993. PHYLIP Phylogenetic Inference Package 3.5c. 3.5c ed., University of Washington, Washington, DC. distributed by author.
- Gervasi, F., Dagnano, I., Vossio, S., Zupi, G., Sacchi, A., Lombardi, D., 1996. NM23 influences proliferation and differentiation of PC12 cells in response to nerve growth factor. Cell Growth Differ. 7, 689–1695.
- Jones, D.T., Taylor, W.R., Thornton, J.M., 1992. The rapid generation of mutation dot matrices from protein sequences. Comput. Appl. Biosci. 8, 275–282.
- Lakso, M., Steeg, P.S., Westphal, H., 1992. Embryonic expression of NM23 during mouse organogenesis. Cell Growth Differ. 3, 873–879.
- Lee, J.-S., Lee, S.-H., 2000. Cloning and characterization of cDNA encoding Zebrafish Danio regio NM23-B gene. Gene 245, 75–79.
- Luo, G., An, G., Wu, R., 1994. A PCR differential screening method for rapid isolation of clones from a cDNA library. Biotechniques 16, 670–675.
- Mehus, J.G., Deloukas, P., Lambeth, D.O., 1999. NME6: a new member of the NM23/nucleoside diphosphate kinase gene family located on the human chromosome 3p21.3. Hum. Genet. 104, 454–459.
- Milon, L., Rousseau-Merck, M.F., Munier, A., Erent, M., Lascu, I., Capeau, J., Lacombe, M.L., 1997. NM23-H4, a new member of the family of human NM23/nucleoside diphosphate kinase genes localised on chromosome 16p13. Hum. Genet. 9 (9), 550–557.
- Morgenstern, B., Dress, A., Werner, T., 1996. Multiple DNA and protein sequence alignment based on segment-to-segment comparison. Proc. Natl. Acad. Sci. USA 9 (3), 12098–12103.
- Munier, A., Feral, C., Milon, L., Pinon, V.P., Gyapay, G., Capeau, J., Guellaen, G., Lacombe, M.L., 1998. A new human NM23 homologue (NM23-H5) specifically expressed in testis germinal cells. FEBS Lett. 434, 289–294.
- Murphy, M., McGinty, A., Godson, G., 1998. Protein kinase C: potential targets for intervention in diabetic nephropathy. Curr. Opin. Nephrol. Hypertens. 7, 563–570.
- Ouatas, T., Abdallah, B., Gasmi, L., Bourdais, J., Postel, E., Mazabraud, A., 1997. Three different genes encode NM23/nucleoside diphosphate kinases in *Xenopus laevis*. Gene 194, 215–225.
- Ouatas, T., Selo, M., Sadji, Z., Hourdry, J., Denis, H., Mazabraud, A., 1998. Differential expression of nucleoside diphosphate kinases (NDPK/NM23) during *Xenopus* early development. Int. J. Dev. Biol. 42, 43–52.
- Page, R.D.M., 1996. TREEVIEW: An application to display phylogenetic trees on personal computers. Comput. Appl. Biosci. 12, 357–358.
- Parks, R.E., Agawal, R.P., 1973. Boyer, P.D. (Ed.), The Enzymes Vol. 8. Academic Press, New York, pp. 307–333.
- Postel, E.H., Berberich, S.J., Flint, S.J., Ferrone, C.A., 1993. Human c-myc transcription factor identified as NM23-H2 nucleoside diphosphate kinase, a candidate suppressor of tumour metastasis. Science 261, 478–479.
- Postel, E.H., 1998. NM23-NDP kinase. Int. J. Biochem. Cell Biol. 30, 1291–1295.
- Powers, D.A., 1991. Evolutionary genetics of fish. Adv. Genet. 29, 119–228.

- Rosengard, A.M., Krutzsch, H.C., Shearn, A., Biggs, J.R., Barker, E., Margulies, I.M.K., Richter King, C., Liotta, L.A., Steeg, P.S., 1989. Reduced NM23/Awd protein in tumor metastasis and aberrant *Drosophila* development. Nature 342, 177–180.
- Sambrook, J., Fritsch, E.F., Maniatis, T., 1989. Molecular Cloning: A Laboratory Manual. second ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Schreiber-Agus, N., Horner, J., Torres, R., Chiu, F.C., Dephinho, R.A., 1993. Zebrafish myc family and max genes: differential expression and oncogenic activity throughout vertebrate evolution. Mol. Cell. Biol. 13, 2765–2775.
- Stahl, J.A., Leone, A., Rosengard, A.M., Porter, L., King, C.R., Steeg, P.S., 1991. Identification of a second human nm23 gene, NM23-H2. Cancer Res. 51, 445–449.
- Steeg, P.S., Bevilacqua, G., Kopper, L., 1988. Evidence for a novel

gene associated with low tumor metastatic potential. J. Natl. Cancer Inst. 80, 200-204.

- Strimmer, K., Von Haeseler, A., 1996. Quartet puzzling: A quartet maximum-likelihood method for reconstructing tree topologies. Mol. Biol. Evol. 13, 964–969.
- Venturelli, D., Martinez, R., Melotti, P., Casella, I., Peschle, C., Cucco, C., Spampinato, G., Darzynkiewicz, Z., Calabretta, B., 1995. Overexpression of DR-NM23 a protein encoded by a member of the NM23 gene family inhibits granulocyte differentiation and induces apoptosis in 32Dc13 myeloid cells. Proc. Natl. Acad. Sci. USA 92, 7435–7439.
- Xu, J., Liu, L.Z., Deng, X.F., Timmons, L., Hersperger, E., Steeg, P.S., Veron, M., Shearn, A., 1996. The enzymatic activity of *Dro-sophila* AWD/NDP Kinase is necessary but not sufficient for its biological function. Dev. Biol. 177, 544–557.