1890–1899 *Nucleic Acids Research, 1999, Vol. 27, No. 8*

Cloning and characterisation of mtDBP, a DNA-binding protein which binds two distinct regions of sea urchin mitochondrial DNA

Paola Loguercio Polosa, Marina Roberti, Clara Musicco, Maria Nicola Gadaleta, Ernesto Quagliariello and Palmiro Cantatore*

Dipartimento di Biochimica e Biologia Molecolare, Università di Bari and the Centro Studi sui Mitocondri e Metabolismo Energetico, CNR, Via Orabona 4, 70125 Bari, Italy

Received December 21, 1998; Revised and Accepted February 24, 1999

DDBJ/EMBL/GenBank accession no. AJ011076

ABSTRACT

The cDNA for the sea urchin mitochondrial D-loop-binding protein (mtDBP), a 40 kDa protein which binds two homologous regions of mitochondrial DNA (the D-loop region and the boundary between the oppositely transcribed ND5 and ND6 genes), has been cloned. Four different 3′-untranslated regions have been detected that are related to each other in pairs and do not contain the canonical polyadenylation signal. The *in vitro* synthesised mature protein (348 amino acids), deprived of the putative signal sequence, binds specifically to its DNA target sequence and produces a DNase I footprint identical to that given by the natural protein. mtDBP contains two leucine zippers, one of which is bipartite, and two small N- and C-terminal basic domains. A deletion mutation analysis of the recombinant protein has shown that the N-terminal region and the two leucine zippers are necessary for the binding. Furthermore, evidence was provided that mtDBP binds DNA as a monomer. This rules out a dimerization role for the leucine zippers and rather suggests that intramolecular interactions between leucine zippers take place. A database search has revealed as the most significative homology a match with the human mitochondrial transcription termination factor (mTERF), a protein that also binds DNA as a monomer and contains three leucine zippers forming intramolecular interactions. These similarities, and the observation that mtDBP-binding sites contain the 3′-ends of mtRNAs coded by opposite strands and the 3′-end of the D-loop structure, point to a dual function of the protein in modulating sea urchin mitochondrial DNA transcription and replication.

INTRODUCTION

Despite the large amount of knowledge on biogenesis of mammalian mitochondria, little is known of mitochondrial (mt)DNA metabolism of invertebrates. Sea urchin is an excellent subject for investigation as it is one of the most developed invertebrates and is used as a model system for studying mitochondrial biogenesis during development. Sea urchin mtDNA is a circular, double-stranded molecule of ~15.7 kb whose sequence has been determined in several species (1–3). It contains the same genes found in vertebrate mitochondrial genomes (4,5), but the gene order and the distribution between the two strands are strikingly different. Major differences concern the separation of the two rRNA genes, the clustering of 15 tRNA genes and the reduced size of the main non-coding region (NCR), which is ~130 bp long. An H-strand replication origin was mapped in the NCR of *Strongylocentrotus purpuratus* mtDNA (6) that was associated with a D-loop triplex of ~80 nt, with the nascent H-strand consisting mostly of RNA. The L-strand replication origin has not formally been mapped; the lagging strand probably initiates from multiple points, one of which appears to be located near the main H-strand replication pause site, at the junction between the genes for ATPase 6 and COIII (7). The sea urchin mitochondrial transcription mechanism appears to differ from that existing in vertebrates: studies in *Paracentrotus lividus* and *S.purpuratus* supported a mechanism based on multiple and probably overlapping transcription units in which post-transcriptional processing events play a relevant role (8,9). In particular, it is intriguing that the 3′-ends of 12S and 16S rRNAs were both located a few bases inside the adjacent downstream gene. Nevertheless, contrary to what occurs in mammals, the termination of the two rRNAs seems not to depend on a protein factor (10). These peculiar mechanisms suggest the likely involvement of regulatory factors different from those described in vertebrates. Two sequence-specific DNA binding proteins, mtPBP1 and mtPBP2, were identified in *S.purpuratus*, that interact with two sequences located in the region of the major H-strand replication pause site, between the ATPase 6 and COIII genes (11,12). By fractionating a mitochondrial lysate from *P.lividus* eggs we identified and purified a 40 kDa protein which binds tightly and specifically a sequence of ~25 bp located in the NCR corresponding to the 3′-end of the D-loop structure (13,14). This suggested that the protein, which was named mtDBP for mitochondrial D-loop binding protein, may serve as a regulatory element in the mtDNA replication process. The same protein also

*To whom correspondence should be addressed at: Dipartimento di Biochimica e Biologia Molecolare, Università di Bari, Via Orabona 4, 70125 Bari, Italy. Tel: +39 080 5443378; Fax: +39 080 5443317; Email: p.cantatore@biologia.uniba.it

recognises, though with lower affinity, another sequence highly homologous to the NCR binding site and encompassing the adjacent 3′-ends of the oppositely transcribed ND5 and ND6 genes. Recent dimethylsulphate footprinting studies confirmed that the binding to both regions has a physiological significance, as it also occurs *in vivo* in unfertilised eggs and in embryos (10).

Here we describe the cloning, sequencing and characterisation of the cDNA encoding for mtDBP. Structure–function analysis of the recombinant protein shows that it exhibits the expected specific DNA-binding capacity, binds DNA as a monomer and contains two leucine zipper domains that probably act by promoting intramolecular interactions. Moreover, mtDBP displays a significant sequence homology with the human mitochondrial transcription termination factor mTERF (15), a protein that also binds DNA as a monomer and contains three leucine zippers (16). These data, together with the above reported features of the sea urchin mitochondrial genetic system (1–3,6–14), point to a dual role of mtDBP in regulating both mtDNA replication and transcription.

MATERIALS AND METHODS

Purification of mtDBP and protein sequence analysis

mtDBP was purified from ~900 g of *P.lividus* eggs (2.5–3.0 g of mitochondrial proteins) as already reported (14). The DNA affinity chromatography eluates (0.9–1.2 mg of proteins) were combined, TCA precipitated, and the pellet was electrophoresed on a 12% SDS–polyacrylamide minigel according to Laemmli (17). The proteins were electrotransferred to polyvynilidene difluoride (PVDF) membrane and stained with 0.1% Amido Black. The protein-containing membrane was excised, rinsed with HPLC grade water and *in situ* digested with trypsin (W. Lane, Harvard Microchemistry Facility). The resulting peptides were fractionated by reverse phase HPLC. Five tryptic peptides were sequenced, yielding five continuous stretches of 6–15 amino acids. Peptide sequences were: EAAFLR (peptide 1); EFGYHGEDL(V + I) (peptide 2); SVYELVEYLK (peptide 3); FFSTPETVMDNI (peptide 4); SLGLENADIINIIYK (peptide 5).

Isolation of RNA and cDNA synthesis

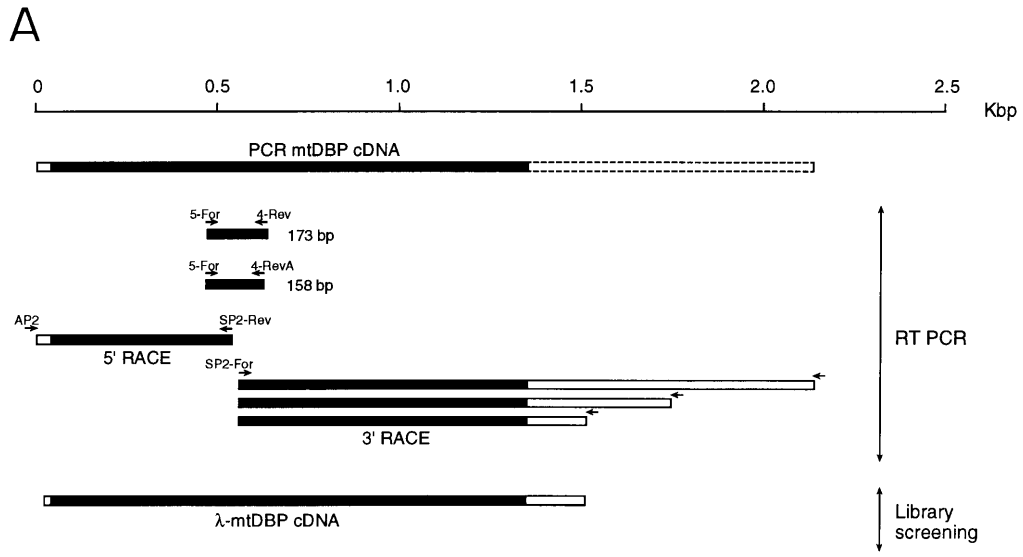
Total RNA was extracted from sea urchin eggs (18) and poly(A)$^+$ RNA was prepared from it using Dynabeads (Dynal AS, Oslo), following the manufacturer's instructions. Single-stranded (ss) cDNA was synthesised by reverse transcription of ~1 µg of poly(A)$^+$ RNA with oligo(dT) primer using a cDNA synthesis kit from Amersham. The reaction mixture, containing ~500 ng of ss cDNA, was stored at +4°C. Prior to use, ss cDNA was heated for 5 min at 95°C and quenched on ice.

Amplification, cloning and sequencing of mtDBP cDNA

Two degenerate primers, 5-For (underlined sequence) derived from peptide 5 (SLGLENADIINIIYK) and 4-Rev (underlined sequence) derived from peptide 4 (**FFSTPETVMDNI**), were designed and used to amplify by PCR ~10 ng of the ss cDNA pool. The reaction was carried out in a 100 µl volume, in the presence of 200 µM dNTPs and 1 µM each primer. After heating at 94°C for 90 s, 3 U of *Taq* DNA polymerase (Boehringer) were added, then the reaction was cycled 30 times (95°C for 1 min, 45°C for 2 min, 72°C for 2 min), with a final incubation at 72°C for 7 min. A 15 µl sample of the PCR mixture was used as template in a second amplification step performed in the same conditions as above except that primer concentration was raised to 1.4 µM. A series of bands including a fragment of 173 bp were obtained (Fig. 1A). A 15 µl sample of the above PCR mixture was then subjected to a third round of PCR in the presence of the same sense primer 5-For and of the antisense primer 4-Rev A, selected from peptide 4 (bold sequence above). A single 158 bp band was obtained, thus confirming the specificity of the 173 bp fragment as expected on the basis of the antisense priming site (Fig. 1A). The second round PCR products were electrophoresed through a 2% agarose gel; the 173 bp fragment was purified, inserted into vector pCRII (Invitrogen) and sequenced according to Loguercio Polosa and Cantatore (19). Rapid amplification of cDNA ends (RACE) was performed to obtain the 5′- and 3′-ends of the mtDBP cDNA (20). The 5′-end of the clone was obtained by using a Marathon cDNA amplification kit (Clontech) following the supplier's recommendations. The 5′-end double-stranded cDNA was made by priming egg poly(A)$^+$ RNA with the specific primer SP1-Rev (nt 584–564, Fig. 1B). The products were ligated to Marathon cDNA adaptors and then subjected to PCR amplification with primers SP1-Rev and AP1 (Clontech). One two-hundredth of the reaction was further amplified with nested primers SP2-Rev (nt 540–518, Fig. 1B) and AP2 (Clontech). The product (Fig. 1A) was cloned into vector pCRII and sequenced. To obtain the 3′-end of the clone, 1 µg of egg poly(A)$^+$ RNA was reverse transcribed with a 36mer oligonucleotide containing $(dT)_{15}$ and a 21-base linker. The ss cDNA pool was used as template in a first PCR reaction with primers SP1-For (nt 518–540, Fig. 1B) and $(dT)_{15}$-linker. One-thousandth of the PCR mixture was subjected to further PCR with nested primer SP2-For (nt 564–584, Fig. 1B) and primer $(dT)_{15}$-linker. 3′-RACE products were shown to be specific by Southern blot hybridisation using an open reading frame (ORF)-containing PCR fragment. The specific 3′-RACE products were gel eluted, cloned into vector pCRII and sequenced. Finally, two ORF-containing PCR products (nt 15–540 and 564–1075, Fig. 1B) were labelled by random priming with [α-^{32}P]dATP (21) and used as probes to screen, according to standard procedure (22), 1×10^6 plaques from a λ Uni-Zap cDNA library prepared from *P.lividus* embryos at the stage of four blastomeres (23). One positive plaque was obtained that was converted into recombinant plasmid pBluescript according to the manufacturer's protocol (Stratagene) and was sequenced on both strands (Fig. 1B).

Plasmid constructs and *in vitro* translation of wild-type and mutated versions of mtDBP

Plasmid constructs of mtDBP suitable for *in vitro* translation were obtained by amplifying the appropriate fragments from the library-derived cDNA clone (λ mtDBP). To generate constructs mtDBPr1, mtDBPr2 and ΔN, the forward primers were DBPr1-For (5′-ACACGAATTCACCATGGTGTCCTCGGAA-TTAACATG-3′), DBPr2-For (5′-ACACGAATTCACCATGGC-AAACTTCCGGGAATGCTCGACT-3′) and ΔN-For (5′-ACAC-GAATTCACCATGGCACCTACGGTCCTGAAACAGAAC-3′), containing an *Eco*RI restriction site and the initiation codon (underlined). As reverse primer, the M13 20mer was used. To generate ΔC, oligonucleotides DBPr2-For and ΔC-Rev (5′-GTGT-CTCGAGCTATAGCACTATCAGTTCATGTTT-3′), containing the stop codon (underlined) and a *Xho*I restriction site, were used.



To obtain ΔL1, primers DBPr2-For and ΔL1-Rev (5′-GTGTGGATCCACGAGTTTTCTGCAGCCTGTC-3′), containing a *Bam*HI restriction site, and ΔL1-For (5′-ACACGGATCCGGTCTGAAGGATGGTGAGGTA-3′), containing a *Bam*HI restriction site, and M13 20mer were used. To generate ΔL2, primers were DBPr2-For and ΔL2-Rev (5′-GTGTGGATCCACAGTACCTGATGATTGATAT-3′), containing a *Bam*HI restriction site, and ΔL2-For (5′-ACACGGATCCGGTTTTACGAAAGAGGAGATG-3′), containing a *Bam*HI restriction site, and M13 20mer. PCR products mtDBPr1, mtDBPr2, ΔN and ΔC were digested with *Eco*RI and *Xho*I, purified and cloned into vector pBluescript II (SK$^+$) (Stratagene). PCR fragments for generating ΔL1 and ΔL2 were *Bam*HI digested, purified and ligated. The constructs were then gel eluted, digested with *Xho*I and *Eco*RI and cloned into pBluescript. The correct nucleotide sequence of all constructs was verified. Proteins were synthesised in reticulocyte lysate by using the coupled transcription–translation system (TNT) from Promega. A 0.8 or 1.5 μg amount of recombinant plasmid was added to a 25 or 50 μl reaction volume and reactions were performed according to the manufacturer's protocols.

DNA binding assays

Gel mobility shift assays were carried out in a 20 μl reaction mixture containing 20 mM HEPES pH 7.9, 5 mM $MgCl_2$, 75 mM KCl, 0.1 mM EDTA, 1 mM DTT, 2 μg of poly (dI·dC), 2 μg of BSA, 20 fmol of the appropriate labelled probe, and the protein fraction as specified in figure legends. After incubation at room temperature for 30 min, samples were loaded on a native 6 or 10% polyacrylamide gel and run at 4°C in 0.5× TBE at 300 V. To quantify the DNA-binding activity of mtDBP mutants relative to the wild-type version, differences in the efficiency of translation of the various constructs were determined. A densitometric analysis of the [^{35}S]methionine-labelled products electrophoresed on a SDS–polyacrylamide gel was performed. Densitometric values were then corrected for the number of methionines of each mutant version.

DNase I footprinting

DNase I footprinting was carried out as reported previously (13) with some modifications. The probes used were as already reported (13); binding reactions were as for the gel shift assay except that the sample was 3.5 times the sample for gel shift assay. After incubation at room temperature for 30 min, the mixture was added to an equal volume of 5 mM $CaCl_2$, 10 mM $MgCl_2$, followed by the addition of 5–50 ng/ml of DNase I (Boehringer), and incubated at room temperature for 60 s. Reactions were stopped by the addition of 20 mM EGTA and processed as described elsewhere (13).

RESULTS

Cloning and sequencing of mtDBP

In order to clone the cDNA for *P.lividus* mtDBP a RT–PCR strategy was used (Fig. 1A). Approximately 1 μg of affinity-purified mtDBP was separated on a SDS–polyacrylamide gel and transferred to PVDF membrane as detailed in Materials and Methods. Following HPLC fractionation of the tryptic peptides, amino acid information was obtained for five of them (Materials and Methods). Because the relative position of the peptides within mtDBP was not known, two pairs of sense and antisense oligonucleotide guessmers were selected from the two longest peptides (peptides 4 and 5) and both primer combinations were tested in PCRs. A particular combination of oligonucleotide guessmers (5-For and 4-Rev) produced, after two rounds of amplification, a series of bands including a fragment of 173 bp (Fig. 1A). This product was inserted into vector pCRII and sequenced. Its deduced amino acid sequence was compared with that of peptides 4 and 5 and shown to encode the known amino acid sequence. Furthermore, inspection of the DNA sequence revealed that it contained peptides 1 and 3 (Fig. 1B). This confirmed that a short unique cDNA sequence of mtDBP was produced. The sequence information was then used to design two specific primers that were employed for RACE using egg poly(A)$^+$ RNA as template for reverse transcription (20). In

B

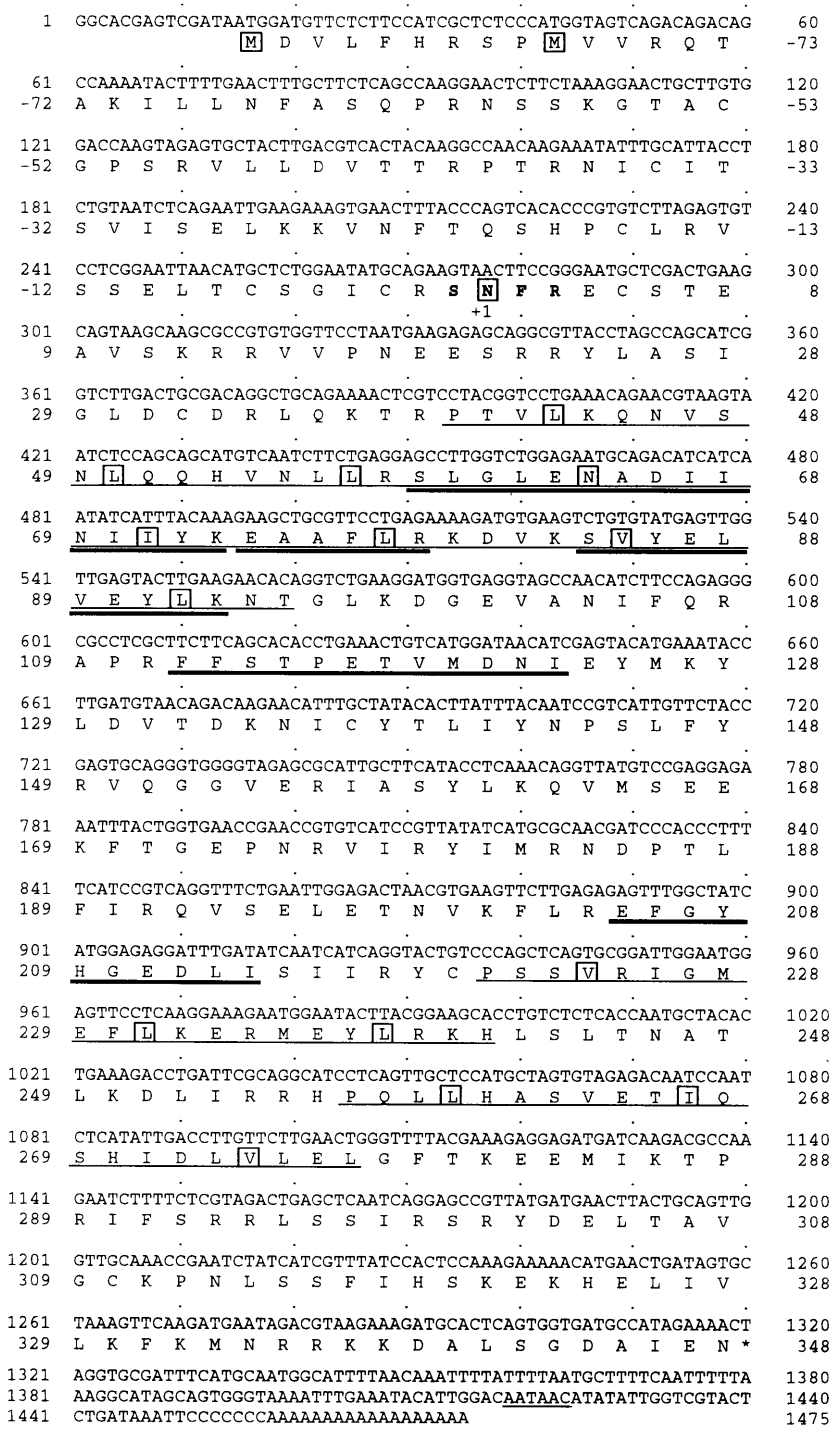


Figure 1. (Above and opposite) Cloning and sequence of mtDBP cDNA. (**A**) Schematic representation of cDNA clones isolated by RT–PCR and library screening. In the upper part of the figure, the entire mtDBP clone obtained by combining the PCR clones is represented. Black bars indicate the protein coding region, open bars the 5′- and 3′-UTR regions. The dashed bar denotes the heterogeneity of the 3′-UTR region. Arrows pointing left or right show the position and direction of synthetic oligonucleotide primers used in PCR. (**B**) Nucleic acid and deduced protein sequences of λ mtDBP cDNA. The two potential starting methionine residues are boxed. The putative N-terminal presequence is indicated with negative numbering. The cleavage sequence (in bold) and the boxed first amino acid of the mature protein were suggested from sequence analysis of mtDBP protein with the PSORTII algorithm. Thick underlines indicate tryptic peptides derived from the purified mtDBP. The positions of the putative leucine zippers are shown by a thin underline with the residue at the *d* position boxed. The asterisk represents the first stop codon. A potential polyadenylation signal is underlined.

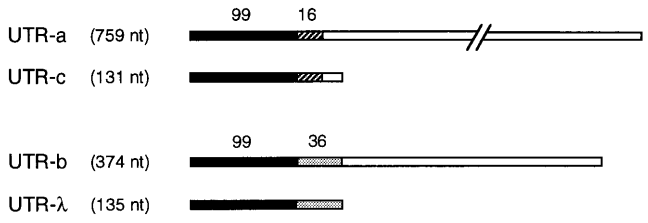


Figure 2. Schematic representation of mtDBP 3′-UTRs. Numbers in parentheses indicate the length of 3′-UTRs. Black bars indicate nucleotide sequence common to the four 3′-UTRs; striped bars indicate the 16 nt stretch common to UTR-a and UTR-c; spotted bars represent the 36 nt stretch common to UTR-b and UTR-λ. Open bars indicate the remaining stretches which are unique.

further experiments two PCR fragments containing part of the ORF were used as probes to screen a λ Uni-Zap cDNA library of *P.lividus* embryos at the 4 cell stage (23). The screening of about 1×10^6 plaques yielded one positive clone (λ mtDBP) having an insert of ~1.5 kb. The λ clone was converted to plasmid DNA by phage rescue excision and then sequenced on both strands. The nucleotide sequence of mtDBP cDNA and the deduced amino acid sequence are displayed in Figure 1B. The cDNA clone, which is 1475 bp long, begins with a short 5′-untranslated region (5′-UTR) of 14 bases. There follows an ORF of 1308 bases encoding 435 amino acids and a 3′-untranslated region (3′-UTR) of 135 bp (UTR-λ). In the 3′-UTR a non-canonical polyadenylation sequence AAUAAC (24) is located 96 bp downstream of the TAG termination codon, followed after 33 more bases by an 18 nt poly(A) tail. The ORF includes all the five peptide sequences obtained by tryptic digestion of the purified *P.lividus* mtDBP. Whereas the 5′-RACE generated one single product of 569 bp, the 3′-RACE produced three cDNA fragments of 1541, 1158 and 915 bp (Fig. 1A). They shared the coding region and only part of the 3′-UTR. Therefore four 3′-UTRs named UTR-a, UTR-b, UTR-c and UTR-λ were detected, having lengths of 759, 374, 131 and 135 nt, respectively (Fig. 2). The four 3′-UTRs ended with an 18–25 nt poly(A) tail; however, a canonical poly(A) addition signal could not be detected.

Structural features of recombinant mtDBP

The ORF of mtDBP cDNA was expected to specify the precursor of mtDBP, including the mitochondrial targeting sequence. The first 100 amino acids encoded in the ORF contain two methionine residues, one of which could be the initiator amino acid of the mtDBP precursor. When a construct containing the entire ORF of 435 amino acids was used in a coupled transcription–translation system a product of ~43 kDa was obtained, having the size of a polypeptide starting at the first methionine (Fig. 3A and B). Use of this AUG as the initiator codon is supported by the observation that the adjacent sequence has a better match to a Kozak consensus initiation sequence than does the second methionine (25). According to an analysis with the PSORT II algorithm, the mtDBP polypeptide sequence is characterised by an N-terminal portion that has the typical features of a mitochondrial targeting sequence (26). The potential cleavage site should be placed between residues S(–1) and N(+1) according to the 'R(–2) rule' (27). Based on this assignment the mature mtDBP should be 348 amino acids long and have a calculated molecular mass of 40.67 kDa, which is actually consistent with the apparent size of 40 kDa estimated by SDS–polyacrylamide gel electrophoresis.

In order to analyse the functional capacity of the recombinant mtDBP (mtDBPr), the wild-type cDNA construct and two other constructs, mtDBPr1 and mtDBPr2, having N-terminal deletions of different sizes, were *in vitro* translated (Fig. 3A and B). The MtDBPr1 construct lacked part of the putative signal peptide (deleted residues were from –87 to –17) whereas mtDBPr2 lacked the whole signal peptide and started from amino acid +1. The binding capacity of mtDBPr was initially determined by gel electrophoretic mobility shift assay. As shown in Figure 3C, the precursor version of the protein was not able to form a complex with the labelled double-stranded oligonucleotide containing the NCR binding site. This feature seems to be common to all the precursors of the mtDNA-binding proteins identified to date, as it has been reported also for the human transcription factors mTERF (16) and mtTFA (28). A tentative explanation for this phenomenon is that the structural fold of the signal sequence could somehow mask the DNA-binding domain and prevent the protein from contacting the corresponding site. Partial (mtDBPr1) and complete (mtDBPr2) removal of the N-terminus from the precursor relieved the apparent inhibition of DNA binding, producing a single, specific retarded band. The binding capacity displayed by mtDBPr2 was much stronger than that of mtDBPr1, which retains part of the presequence. Moreover, the complex formed by mtDBPr2 had the same mobility as that produced by the affinity-purified protein. These data let us assume that mtDBPr2 represents the active and probably mature version of the protein. Increasing the amount of the protein resulted in an increase in the intensity of the retarded band without the appearance of any more slowly moving secondary band. This finding argues against the formation of different complexes of mtDBP with the probe. Finally, no complex formation was detected using a probe with a deletion in the NCR-binding site (data not shown), indicating that the interaction was specific. The specificity of mtDBPr interaction with the DNA was further confirmed by DNase I footprinting analysis. As shown in Figure 4, the pattern of DNase I protection at the NCR (nt 1098–1126) and at the ND5/ND6 boundary (nt 14 028–14 053) binding site produced by mtDBPr2 is essentially the same as that obtained with the affinity-purified mtDBP. This result conclusively confirms that the cDNA clone we isolated codes for *P.lividus* mtDBP.

A BLASTP analysis of the amino acid sequence of mtDBP with the available protein databases revealed as statistically significant a match with the human mitochondrial transcription termination factor mTERF (accession no. Y09615; $P = 1.7 \times 10^{-14}$) and a match with an unknown protein from *Arabidopsis thaliana* (accession no. AC000375; $P = 1 \times 10^{-10}$). This is a 462 amino acid long polypeptide (J.Schwartz, personal communication) that is predicted to be a mitochondrial protein by PSORTII analysis. When the comparison was performed with the mature versions of the proteins, 22% amino acid identity and 61% amino acid similarity were obtained for the pair mtDBP/mTERF (Fig. 5); a comparison between mtDBP and the unknown *A.thaliana* protein (not shown) gave 18% amino acid identity and 59.8% amino acid similarity. Concerning the pair mtDBP/mTERF, the homology seems to be uniformly distributed along the molecule; however, regions of higher similarity (residues 99–113, 202–221 and 310–325) can be detected. Almost 40% of the common residues between mtDBP and mTERF were conserved among the three proteins.

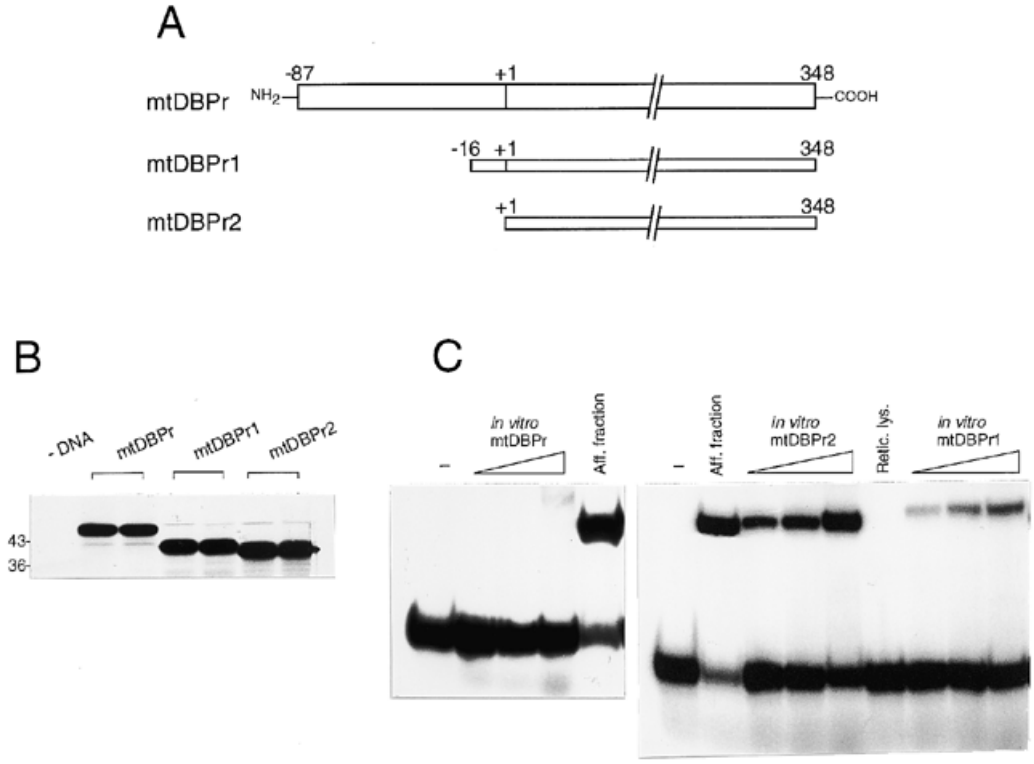


Figure 3. Analysis of mtDBP recombinant forms. (**A**) Schematic representation of the precursor and shortened versions of mtDBP. The numbers represent amino acid positions according to the numbering system used in Figure 1B. (**B**) SDS–polyacrylamide gel analysis. Precursor and shortened versions of mtDBP shown in (A) were *in vitro* expressed in the presence of [^{35}S]methionine, separated on a SDS–polyacrylamide gel and subjected to autoradiography. The position of molecular mass markers (in kDa) are shown to the left. (**C**) Mobility shift assays. Three increasing amounts of *in vitro* transcription–translation reaction mixtures (2, 4 and 8 µl) programmed with the templates shown in (A) were incubated with [α-^{32}P]-labelled double-stranded 44mer oligonucleotide probe containing the NCR binding site (14). The protein–DNA complexes were resolved on a native polyacrylamide gel. The affinity-purified mtDBP (Aff. fraction), a minus protein (–) and a non-programmed reticulocyte lysate reaction (Retic. lys.) were used as controls.

Sequence analysis of mtDBP cDNA revealed, as the most evident feature, the presence of multiple leucine zipper (LZ) motifs (underlined in Fig. 1B). The LZ motif consists of a repetition of leucines, or similar hydrophobic amino acids, which are spaced seven residues apart (29,30). The most typical LZ motif found in mtDBP is LZ1, situated near the N-terminus of the mature protein between residues 40 and 95. It is composed of eight repeats of the heptad X_3LX_3 with asparagine, isoleucine and valine substituting for leucine at the *d* position in the fourth, fifth and seventh heptads, respectively (the residues in the heptad being designed *a*–*g*). The repeats exhibit the expected preponderance of hydrophobic residues at the *a* (7/8) and *d* (7/8) positions of the heptads. A second potential leucine zipper can be localised near the C-terminus, between positions 221 and 277. In this case, the motif is formed by two repeats of three heptads located between residues 221–241 and 257–277 and separated by a 15 amino acid loop. These heptads also have a high preponderance of hydrophobic residues at *a* (4/6) and *d* (6/6) positions, with two valines and one isoleucine substituting for leucine at the *d* position of the first, sixth and fifth heptads, respectively.

The roles of the different regions of mtDBP in its DNA-binding activity were investigated in gel shift experiments employing deletion mutants. Four mutated versions of the protein were derived from mtDBPr2, since this has been shown to have the same binding activity as the natural mtDBP. Two deletion constructs were designed to produce N- and C-terminal truncated versions of the protein (ΔN and ΔC) lacking 39 and 19 amino acids, respectively. The LZ1 region was deleted in the construct ΔL1 eliminating amino acids from 40 to 95. To test the effect of specifically disrupting the bipartite leucine zipper domain LZ2, another construct, ΔL2 (lacking amino acids 221–277), was made (Fig. 6A). Of these mutated versions, only ΔC (Fig. 6A) retains some binding activity (~50%). When the N-terminus or the two leucine zipper motifs were deleted separately, no binding activity was observed. This result could be explained either by the removal of a domain that interacts with DNA or by a deletion-induced change in the protein conformation which prevents the protein from binding DNA.

Band shift assays using mtDBPr2 and a DNA probe containing the specific binding site revealed a single retarded band even when large amounts of the protein were used (Fig. 3C). This result, together with the absence of a dyad symmetry in the mtDBP binding site suggests that the protein binds DNA as a monomer. In order to obtain conclusive evidence about this point the heterodimerisation assay was used. This assay requires that two versions of mtDBPr differing in size form DNA–protein complexes with different mobility in a gel shift assay. If dimers of mtDBPr bind the DNA target, then using both protein forms in the assay, three retarded bands should appear (homodimers of the large form, homodimers of the small form and heterodimers of the

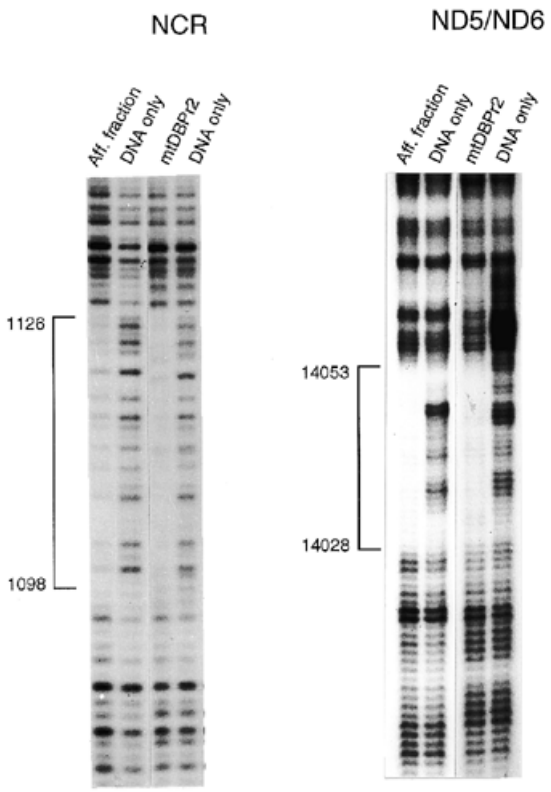


Figure 4. DNase I footprinting analysis. (**A**) Protected regions on the L-strand of the binding site in the NCR and on the H-strand of the binding site at the junction of the ND5/ND6 genes (ND5/ND6) are indicated by brackets. Numbers denote the nucleotide position (1). The recombinant mtDBP (mtDBPr2) and the affinity-purified fraction (Aff. fraction) were incubated with DNA fragments and treated with DNase I as described in Materials and Methods. 'DNA only' refers to sample containing unbound DNA.

two forms). As the deletion mutant ΔC binds the DNA probe producing a retarded complex that runs faster than the wild-type protein-containing complex (Fig. 6A), a mixture of the polypeptides mtDBPr2 and ΔC in different proportions was employed to perform mobility shift experiments. As reported in Figure 6B, only two retarded bands corresponding to those produced by the two proteins alone were obtained. No third intermediate band was observed, as would be expected in the case of a dimer. Since the recombinant mtDBP produced by the clone mtDBPr2 gave a single retarded band with the same mobility of that shown by the natural mtDBP (Fig. 3C) the conclusion that mtDBPr2 binds to DNA as a monomer likely also applies to the natural mtDBP.

DISCUSSION

In this paper we report the cloning and characterisation of the cDNA encoding for *P.lividus* mtDBP, a mitochondrial protein of 40 kDa which specifically binds two regions of sea urchin mtDNA. At present this is the third animal mitochondrial DNA-binding protein whose cDNA has been cloned (16,28). Band shift and DNase I footprinting analysis (Figs 3 and 4) showed that the recombinant protein, obtained by *in vitro* transcription–translation of its cDNA, binds with high specificity the same two regions contacted by the affinity chromatography purified sea urchin protein. Band shift experiments with N-terminal and C-terminal deletion mutants (Fig. 6A) indicated that removal of the N-terminus completely abolishes the binding capacity whereas the C-terminal deletion has a moderate effect on binding (which is reduced by ~50%). Since the deleted regions contain a small stretch of basic amino acids (residues 12–39 and 330–338), it is likely that these residues, particularly those located at the N-terminus, could be involved in contacting DNA.

The analysis of the amino acid sequence also showed the presence of two heptad repeats between residues 40–95 and 221–277 (the latter is bipartite). Since the *a* and *d* positions of the heptads are prevalently occupied by hydrophobic residues, with leucine being the prevalent amino acid in the *d* position, it can be assumed that these repeats, named LZ1 and LZ2, behave as leucine zippers. LZ1 and LZ2 appear to be relevant in determining mtDBP binding properties as experiments with deletion mutants (Fig. 6A) showed that the removal of each motif abolishes the binding to DNA. It is well known that leucine zipper motifs are required to form protein dimers (29,30); however, this does not seem the case for mtDBP. The heterodimerisation assay (Fig. 6B) showed that the protein binds DNA as a monomer, an observation that is consistent with the absence of a dyad symmetry in the contacted sequence. Therefore, the heptad motifs could establish intramolecular interactions so as to generate coiled-coil structures. Such conformations have been described for the heat shock factor from human and *Drosophila* cells (31,32), for seryl-tRNA synthetase (33), for a class of cytoskeleton proteins known as spectrins (34) and, more recently, for human mitochondrial transcription factor mTERF (16). This is a DNA-binding protein which binds DNA as a monomer downstream of the 3′-end of the 16S rRNA gene and terminates transcription of the ribosomal unit (15). mTERF exhibits three leucine zipper heptads which are probably used to form an intramolecular triple-stranded coiled-coil structure. This conformation would be needed to expose the binding domain, which appears to consist of two basic stretches located at the N- and C-termini of the protein, to DNA. This model could be easily adapted to sea urchin mtDBP in the sense that the LZ1 and LZ2 motifs would form a two-stranded coiled-coil structure so as to expose the N- and the C-terminal basic residues of the protein to DNA. The relationship between mtDBP and mTERF also concerns the primary structure as the two proteins show 22% amino acid identity (Fig. 5). Moreover, mtDBP is also related to a putative mitochondrial 462 amino acid long polypeptide of *A.thaliana* (18% amino acid identity) so that a common evolutionary origin for the three proteins can be suggested. The different positions of the leucine zipper motifs contained in mtDBP and mTERF implies that they arose independently after the separation of the two genes.

A further parallel between the mammalian and the sea urchin protein was provided by the observation that their DNA binding sites are located in correspondence with the 3′-ends of mtRNAs. It is known that the mTERF binding site is located downstream of the 3′-end of the 16S rRNA (15) and that both binding sites of mtDBP contain the 3′-ends of transcripts encoded by opposite strands. They are the RNA replication primer and a precursor of 12S rRNA in the NCR binding site (6) and the mRNAs for the ND5 and ND6 polypeptides in the other binding site (P.Cantatore *et al.*, unpublished results). These observations strongly point to a role of mtDBP as a transcription termination factor. By contacting its target site in the NCR, mtDBP would block the passage of RNA polymerase through the replication origin,

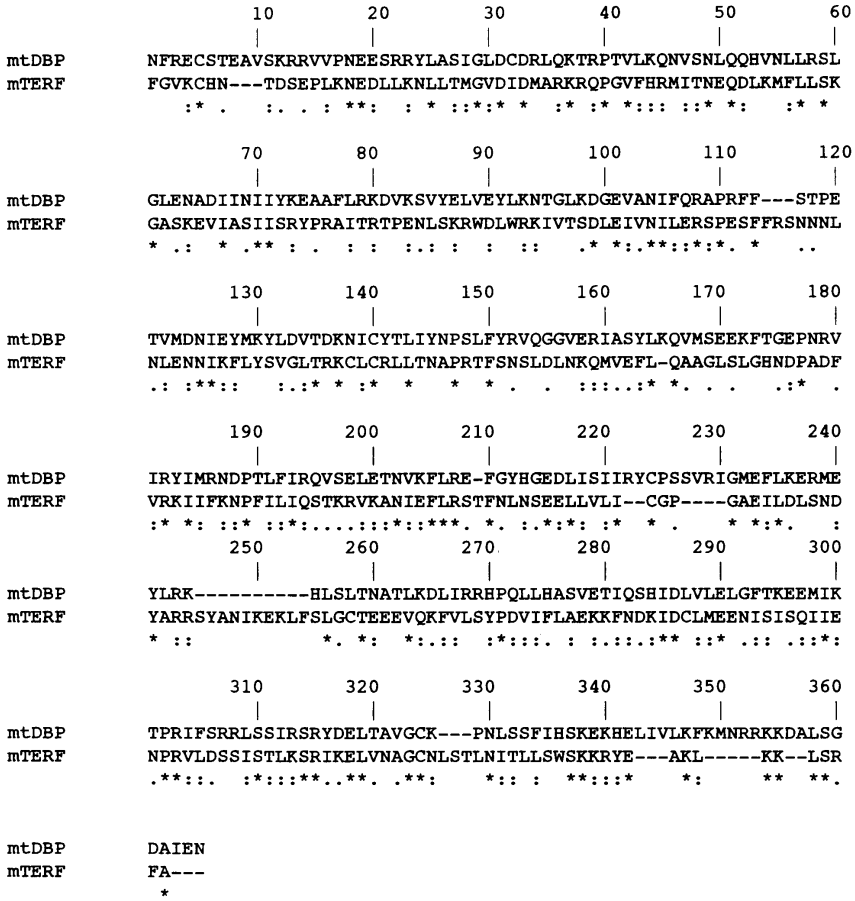


Figure 5. Sequence alignment of mtDBP and mTERF mature forms. Asterisks indicate identical nucleotides; colons and dots indicate very similar and similar residues according to Thompson *et al*. (43).

avoiding the read-through of this region which might disturb primer–template base pairing; on the other hand, transcription arrest at the boundary between the ND5 and ND6 genes would prevent head-on collision between the H- and L-strand transcription machinery. The proposed role for mtDBP as a bidirectional transcription terminator provides new insights into the mechanism of mitochondrial transcription termination in sea urchins as compared to vertebrates. In mammals one termination event for the ribosomal transcription unit depending on mTERF has been described (15). In sea urchins transcription arrest would occur in correspondence with the two mtDBP binding sites, whereas the 3′-ends of the two rRNAs will be generated by post-transcriptional processing events as no protein-mediated termination event takes place in these regions (10).

Based on the observation that the mtDBP-binding site in the NCR contains the 3′-end of the D-loop structure, it was previously inferred that mtDBP might have a role in regulating mtDNA replication (14). By binding to its target site in the NCR, the protein could function as a negative regulator of H-strand elongation, thereby leading to D-loop formation. Relaxation of this interaction would favour H-strand extension thus resulting in productive replication of the mitochondrial genome. A protein of ~48 kDa, the TAS-binding factor, has been shown to serve this function in mammals (35). Therefore, mtDBP is likely to play a dual function in regulating both mitochondrial DNA replication and transcription. The use of the same protein to perform a role both in replication and transcription in sea urchins is justified by the compact organisation of the sea urchin D-loop (~130 nt as opposed to ~1000 nt in mammals) and by the observation that the 3′-end of the RNA primer is very close (20–30 bp) to the 3′-end of the newly synthesised DNA (6). Two other sequence-specific DNA-binding proteins (mtPBP-1 and mtPBP-2) from sea urchin mitochondria have been characterised (11,12). They bind to the main pause region of sea urchin mtDNA, which is located at the boundary between the COIII and ATPase 6 genes where the main origin for lagging strand replication was mapped. In this case an action through blocking of leading strand replication and progression of the RNA polymerase at this site has also been proposed. The observation that the same protein factor is able to arrest both replication and transcription has been reported in many prokaryotic and eukaryotic systems. In particular, in *Escherichia coli* and in *Bacillus subtilis* it was shown that the same protein factor (*ter* protein in *E.coli* and RTP in *B.subtilis*) is able to block both progression of the replication fork at specific sites and RNA chain elongation (36). In mammals it has recently been described that the RNA polymerase I transcription termination factor TTF-1 also causes polar arrest of rDNA replication, preventing head-on collision between the DNA replication apparatus and the transcription machinery (37).

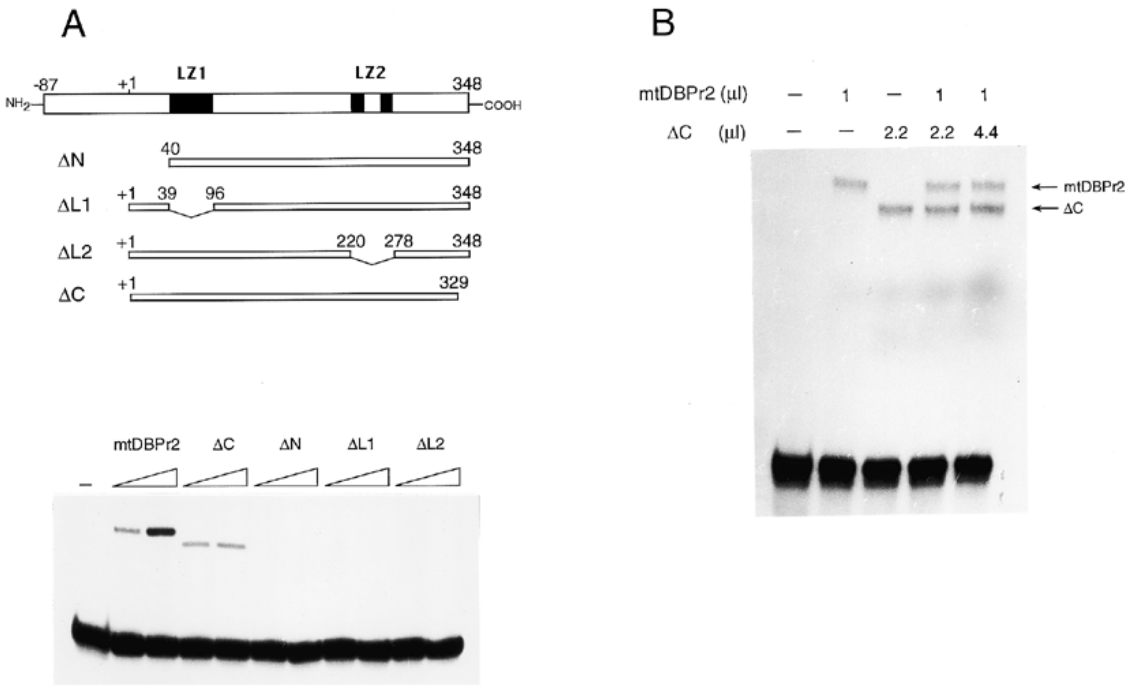


Figure 6. DNA-binding properties of recombinant mtDBP. (**A**) Mobility shift assay using different deletion mutants. (Upper) Schematic representation of wild-type and deleted constructs used as templates in the *in vitro* expression system. The putative leucine zipper domains (LZ1 and LZ2) are indicated by black boxes. The numbers represent the amino acid positions according to the numbering system used in Figure 1B. (Lower) Mobility shift assays using two different amounts (2 and 4 μl) of the expression reaction mixtures containing equivalent amounts of the constructs shown in the upper part as templates. The mature recombinant protein (mtDBPr2) was used as control. The probe was the 44mer double-stranded oligonucleotide. (**B**) mtDBP binds mtDNA as a monomer. Mobility shift analysis was performed by incubating the mature version (mtDBPr2) and the C-terminal truncated (ΔC) version of mtDBP with the labelled 44mer probe. Different amounts of the expression reaction mixtures were employed in the mobility shift assays, as reported at the top of the figure. The protein–DNA complexes were resolved on a 10% native polyacrylamide gel.

In light of all the observations reported here, it seems that mtDBP and mTERF are two proteins having a common evolutionary origin that diverged to accomplish different roles, according to the variation in gene organisation and expression between sea urchin and mammalian mitochondrial genomes.

In the course of mtDBP cDNA cloning, we identified four identical ORFs which display an unusual organisation of their 3′-UTRs (Fig. 2). All four UTRs share the first 99 bp sequence; however, while UTR-λ appears to be a shortened version of UTR-b, UTR-a and UTR-c overlap each other for only a total of 115 bp. Therefore, the four UTRs appear to be related to each other in pairs. No other sequence similarity, such as AU-rich sequence motifs (38), are exhibited by the four UTRs. Concerning the polyadenylation signal, no canonical AAUAAA sequence (39) was detected in any of the four UTRs. However, UTR-λ exhibits the signal AAUAAC, which has been found as a polyadenylation signal in the sea urchin CS H1 cDNA (24). UTR-a, UTR-b and UTR-c display other variants of the signal (AAGAAA, AGUAAA and AAUACA, respectively). The four 3′-UTRs detected so far are probably only some of the existing 3′-UTRs of mtDBP. A RT–PCR experiment (not shown) employing a set of primers designed to detect differences in the 3′-UTR region of RNAs indeed revealed the existence of a complex pattern of RNA species for mtDBP which appeared to be developmentally regulated. These multiple mRNA species might be involved in determining the translational efficiency and/or the stability of the message during development by promoting the rapid degradation and removal of the message or by conferring an increase in translation at critical developmental stages (40–42).

ACKNOWLEDGEMENTS

We are very grateful to R. Fiore for collaboration in the late stages of this investigation. The help of W. Lane (Harvard Microchemistry Facility, Cambridge, MA) in the protein sequencing analysis is gratefully acknowledged. We thank F. Aniello for kindly providing the λ Uni-Zap cDNA library of *P.lividus* embryos. The technical assistance of F. Milella and V. Cataldo is gratefully acknowledged. This work was supported in part by Ministero dell'Università e della Ricerca Scientifica, project 'Protein–nucleic acid interactions', from Consiglio Nazionale delle Ricerche, contract no. 96.03726.CT14, and from Telethon Italy (grant no. 863).

REFERENCES

1 Cantatore,P., Roberti,M., Rainaldi,G., Gadaleta,M.N. and Saccone,C. (1989) *J. Biol. Chem.*, **264**, 10965–10975.
2 Jacobs,H.T., Elliott,D.J., Math,V.B. and Farquharson,A. (1988) *J. Mol. Biol.*, **202**, 185–217.
3 De Giorgi,C., Martiradonna,A., Lanave,C. and Saccone,C. (1996) *Mol. Phylogenet. Evol.*, **5**, 323–332.
4 Cantatore,P. and Saccone,C. (1987) *Int. Rev. Cytol.*, **108**, 149–208.
5 Attardi,G. and Schatz,G. (1988) *Annu. Rev. Cell. Biol.*, **4**, 289–333.

6 Jacobs,H.T., Herbert,E.R. and Rankine,J. (1989) *Nucleic Acids Res*., **17**, 8949–8965.
7 Mayhook,A.G., Rinaldi,A.M. and Jacobs,H.T. (1992) *Proc. R. Soc. Lond. B*, **248**, 85–94.
8 Cantatore,P., Roberti,M., Loguercio Polosa,P., Mustich,A. and Gadaleta,M.N. (1990) *Curr. Genet.*, **17**, 235–245.
9 Elliott,D.J. and Jacobs,H.T. (1989) *Mol. Cell. Biol*., **9**, 1069–1082.
10 Roberti,M., Loguercio Polosa,P., Musicco,C., Milella,F., Qureshi,S., Gadaleta,M.N., Jacobs,H.T. and Cantatore,P. (1999) *Curr. Genet*., **34**, 449–458.
11 Qureshi,S.A. and Jacobs,H.T. (1993) *Nucleic Acids Res*., **21**, 811–816.
12 Qureshi,S.A. and Jacobs,H.T. (1993) *Nucleic Acids Res*., **21**, 2801–2808.
13 Roberti,M., Mustich,A., Gadaleta,M.N. and Cantatore,P. (1991) *Nucleic Acids Res*., **19**, 6249–6254.
14 Loguercio Polosa,P., Roberti,M., Mustich,A., Gadaleta,M.N. and Cantatore,P. (1994) *Curr. Genet*., **25**, 350–356.
15 Kruse,B., Narasimhan,N. and Attardi,G. (1989) *Cell*, **58**, 391–397.
16 Fernandez-Silva,P., Martinez-Azorin,F., Micol,V. and Attardi,G. (1997) *EMBO J*., **16**, 1066–1079.
17 Laemmli,U.K. (1970) *Nature*, **227**, 680–685.
18 MacDonald,R.J., Swift,G.H., Przybyla,A.E. and Chirgwin,J.M. (1987) *Methods Enzymol*., **152**, 223–224.
19 Loguercio Polosa,P. and Cantatore,P. (1997) *Comments Amers. Life Sci*., **23**, 10–11.
20 Frohman,M.A. (1994) *PCR Methods Applicat*., **4**, S40–S58.
21 Feinberg,A.P. and Vogelstein,B. (1984) *Anal. Biochem*., **137**, 266–267.
22 Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd Edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
23 Fucci,L., Piscopo,A., Aniello,F., Branno,M., Di Gregorio,A., Calogero,R. and Geraci,G. (1995) *Gene*, **152**, 205–208.
24 Mandl,B., Brandt,W.F., Superti-Furga,G., Graninger,P.G., Birnstiel,M.L. and Busslinger,M. (1997) *Mol. Cell. Biol*., **17**, 1189–1200.
25 Kozak,M. (1986) *Cell*, **44**, 283–292.
26 Nakai,K. and Kanehisa,M. (1992) *Genomics*, **14**, 897–911.
27 von Heijne,G., Steppuhn,J. and Herrmann,R.G. (1989) *Eur. J. Biochem*., **180**, 535–545.
28 Parisi,M. and Clayton,D.A. (1991) *Science*, **252**, 965–969.
29 Landshultz,W.H., Johnson,P.F. and McKnight,S.L. (1988) *Science*, **240**, 1759–1764.
30 Hurst,H. (1995) In Sheterline,P. (ed.), *Protein Profile*. Academic Press, London, UK, Vol. 2, pp. 105–168.
31 Zuo,J., Baler,R., Dahl,G. and Voellmy,R. (1994) *Mol. Cell. Biol*., **14**, 7557–7568.
32 Westwood,T. and Wu,C. (1993) *Mol. Cell. Biol*., **13**, 3481–3486.
33 Cusack,S., Berthet-Colominas,C., Härtlein,M., Nassar,N. and Leberman,R. (1990) *Nature*, **347**, 249–255.
34 Yan,Y., Winograd,E., Viel,A., Cronin,T., Harrison,S.C. and Branton,D. (1993) *Science*, **262**, 2027–2030.
35 Madsen,C.S., Ghivizzani,S.C. and Hauswirth,W.W. (1993) *Mol. Cell. Biol*., **13**, 2162–2171.
36 Mohanty,B.K., Sahoo,T. and Bastia,D. (1996) *EMBO J*., **15**, 2530–2539.
37 Gerber,J.K., Gögel,E., Berger,C., Wallisch,M., Müller,F., Grummt,I. and Grummt,F. (1997) *Cell*, **90**, 559–567.
38 Chen,C.-Y. and Shyu,A.-B. (1995) *Trends Biochem. Sci*., **20**, 465–470.
39 Humphrey,T. and Proudfoot,N.J. (1988) *Trends Genet*., **4**, 243–245.
40 Sheets,M.D., Ogg,S.C. and Wickens,M.P. (1990) *Nucleic Acids Res*., **18**, 5799–5805.
41 Jackson,R.J. (1993) *Cell*, **74**, 9–14.
42 Wilhelm,J.E. and Vale,R.D. (1993) *J. Cell Biol*., **123**, 269–274.
43 Thompson,J.D., Higgins,D.G. and Gibson,T.J. (1994) *Nucleic Acids Res*., **22**, 4673–4680.