An open reading frame in intron seven of the sea urchin DNA-methyltransferase gene codes for a functional AP1 endonuclease

Anna Valentina CIOFFI^{*1}, Diana FERRARA^{†1}, Maria Vittoria CUBELLIS[‡], Francesco ANIELLO[§], Marcella CORRADO^{*}, Francesca LIGUORI^{*}, Alessandro AMOROSO^{*}, Laura FUCCI[§] and Margherita BRANNO^{*2}

^{*}Biochemistry and Molecular Biology Laboratory, Stazione Zoologica 'A. Dohrn', Villa Comunale 80121 Naples, Italy, [†]Dipartimento Medicina Sperimentale, Sezione di Fisiologia Umana e Funzioni Biologiche, Integrate 'F. Bottazzi', Second University of Naples, Via Costantinopoli 16, 80138 Naples, Italy, [‡]Department of Organic and Biological Chemistry, University of Naples 'Federico II', Via Mezzocannone 16, 80138 Naples, Italy, and [§]Department of Genetics, General and Molecular Biology, University of Naples 'Federico II', Via Mezzocannone 8, 80138 Naples, Italy

Analysis of the genome structure of the *Paracentrotus lividus* (sea urchin) DNA methyltransferase (DNA MTase) gene showed the presence of an open reading frame, named METEX, in intron 7 of the gene. METEX expression is developmentally regulated, showing no correlation with DNA MTase expression. In fact, DNA MTase transcripts are present at high concentrations in the early developmental stages, while METEX is expressed at late stages of development. Two METEX cDNA clones (Met1 and Met2) that are different in the 3' end have been isolated in a cDNA library screening. The putative translated protein from Met2 cDNA clone showed similarity with

Escherichia coli endonuclease III on the basis of sequence and predictive three-dimensional structure. The protein, over-expressed in *E. coli* and purified, had functional properties similar to the endonuclease specific for apurinic/aprimidinic (AP) sites on the basis of the lyase activity. Therefore the open reading frame, present in intron 7 of the *P. lividus* DNA MTase gene, codes for a functional AP endonuclease designated SuAP1.

Key words: development, DNA repair, glycosylase, *Paracentrotus lividus*, splicing.

INTRODUCTION

All types of eukaryotic genomes contain interrupted genes. The proportion of interrupted genes is low in yeast and increases in the lower eukaryotes; few genes are uninterrupted in higher eukaryotes. The lengths of the introns, which can extend from 0.1–100 kb in vertebrates, determine the gene size.

Introns usually have no coding functions, but intronic open reading frames (ORFs) have been reported first in fungal mitochondria and in chloroplasts [1,2].

Sometimes introns of *Saccharomyces cerevisiae* mitochondrial genes include long ORFs that are thought to be involved in splicing of the gene transcript [3]. Another intronic gene product appears to catalyse the integration of the intron itself into intronless copies of the gene [4]. The large rRNA gene of *Neurospora crassa* mitochondria contains an intronic ORF thought to encode a ribosomal protein [1]. Recently it has been found in the ascidian *Molgula oculata* that the first intron of the *Manx* gene contains an ORF that encodes a DEAD-box RNA helicase [5] and that the *Manx-bobcat* gene complex has a role in the development of chordate feature in ascidian tadpole larvae.

In all reported cases the intronic ORF is encoded on the same strand as the exonic gene. These ORFs do not appear to be transcribed independently from the exonic genes, so that the extent to which they are examples of genes in other genes is debatable.

There are few examples of independent expression of intronic ORFs. Henikoff and co-workers [6] reported that a fully functional *Drosophila melanogaster* pupal cuticle protein gene is

located within an intron of the Gart locus encoding three purine pathway enzyme activities [7]. The cuticle protein gene and the purine pathway gene have no common sequences, as they are derived from opposite DNA strands; moreover, the patterns of their expression are very different during development.

The presence of an ORF within an intron might derive from the insertion of a DNA sequence through the action of mobile genetic elements (transposons and retrotransposons).

Within retrotransposons the long interspersed nuclear elements (LINEs) are a very large class of transposable elements that are present in most eukaryotes [8]. They possess an ORF with coding capacities for a protein with endonuclease [9,10], reverse transcriptase [8] and sometimes RNase H activities [11]. The chronology of the acquisition of the various enzymatic domains suggests that the first LINE retrotransposon contained an endonuclease activity related to restriction enzymes and that the LINEs that have a nuclease structurally related to apurinic/aprimidinic (AP) endonuclease [9,10] derived from this ancestral group. The RNase H domain would have been acquired later [8].

Here, we identify for the first time an AP endonuclease embedded in an intron of the DNA methyltransferase (MTase) gene. We isolated two cDNA clones that are expressed and regulated during sea urchin development. Their expression is independent from the host gene, and this is the first example of an independent ORF in an intron that is transcribed on the same strand as the exonic gene. We have expressed and purified the coded protein and demonstrated that its endonuclease activity belongs to the class AP1.

Abbreviations used: ORF, open reading frame; LINE, long interspersed nuclear element; DNA MTase, DNA methyltransferase; AP, apurinic/aprimidinic; SSC, sodium saline citrate; TBE, Tris-borate-EDTA; LB, Luria-Bertoni (broth); 3'UTR, 3' untranslated region; SuAP1, sea urchin endonuclease III homologue 1; Exo, exonuclease; GST, glutathione S-transferase.

¹ These authors contributed equally to this work.

² To whom correspondence should be addressed (e-mail maggy@alpha.szn.it).

MATERIALS AND METHODS

Library screening, sequencing of genomic and cDNA clones, and computer analysis

DNA was isolated from the sperm of individual *Paracentrotus lividus* sea urchins using the procedure of Sambrook et al. [12] and a genomic library was constructed in lambda EMBL4 vector (Stratagene, La Jolla, CA, U.S.A.). Poly(A)⁺ RNA was isolated from male gonads of sea urchin and a cDNA library was prepared in the Uni-Zap vector (Stratagene) as described previously [13]. Library screening was carried out according to the protocols for each vector. The genomic subclone and the isolated cDNA clones were sequenced on both strands using [³⁵S]ATP (1000 Ci/mmol, Amersham) by the dideoxy chain termination method [14]. The deduced protein sequences were compared with the EMBL gene bank database and protein sequence alignments were generated using the Homstrad database [15] and the FUGUE program [16].

DNA and RNA probes

For Northern-blot analysis and cDNA-library screening we prepared a DNA probe by PCR using a genomic DNA subclone as template. Part of METEX was amplified with the following forward and reverse primers: P1 (5'-GGCGGGAGTCCACAGGAGCTCTG-3') and P2 (5'-ACTGTGCAGAAGGTGTTGTATGG-3'). The PCR product, named 0.85 kb METEX DNA, was labelled with [γ -³²P]dATP and [γ -³²P]dCTP (3000 Ci/mmol, Amersham) using the random priming labelling kit (Amersham) and it was used as probe. The RNA probes for whole-mount-*in-situ* hybridization were prepared from the pBluescript clone containing the complete cDNA insert using the Digoxigenin RNA labelling kit (Boehringer Mannheim, Mannheim, Germany).

Northern-blot analysis

P. lividus sea urchin embryos were grown as described previously [17]. Total RNAs were prepared from male and female gonads, from embryos at different stages of development following the method described by Chomczynski and Sacchi [18]. Poly(A)⁺ RNAs were prepared from total RNA by oligo(dT)-cellulose chromatography [12]. Poly(A)⁺ RNA (10 μ g) from the indicated source was loaded in each lane, electrophoresed on a 1% agarose/formaldehyde gel and transferred to Hybond-N⁺ filters (Amersham) by overnight capillary blotting. The filter was incubated in prehybridization solution (5 $\times$ SSC, 5 $\times$ Denhardt's, 100 μ g/ml sonicated salmon sperm DNA and 50 mM sodium phosphate at pH 7.0) for 1 h at 65 °C and hybridized overnight at 65 °C with 2 $\times$ 10⁶ c.p.m. of radiolabelled probe/ml of hybridization solution (1 $\times$ SSC is 0.15 M NaCl/0.015 M sodium citrate). The filter was washed twice for 30 min at 65 °C in 0.2 $\times$ SSC and 0.1% SDS and finally was exposed to X-ray film (HR-H, Fuji Photo Film Co., Ltd., Tokyo, Japan).

Southern analysis

HindIII- and HaeIII-digested DNA (10 μ g) from a single sea urchin individual were fractionated on a 0.7% agarose gel in 1 $\times$ Tris/borate/EDTA (TBE) and transferred to Hybond-N⁺ filters (Amersham) by overnight capillary blotting. The hybridization of the membrane was carried out with the same procedure used for the Northern-blot analysis.

Whole mount *in-situ* hybridization

Embryos at a selected stage were fixed in 2.5% glutaraldehyde, 0.14 M NaCl, 0.2 M phosphate buffer, pH 7.4, washed, and dehydrated in a 30, 50 and 70% ethanol series and then kept at -20 °C. All pre-hybridization and hybridization steps were carried out as described by Ransick and co-workers [19] and modified by Kirchhamer and Davidson [20], using 0.01 μ g/ml probe.

Construction of a glutathione S-transferase (GST)-fusion protein in pGEX-2T

The open reading frame was amplified by PCR from the pBluescript clone containing the complete cDNA insert using the following forward and reverse primers: Su1 (5'-CGATGACCTGCATCACAGCATAGTAAAG-3') and Su2 (5'-GTATATGTATCTTGTAATTAGGATTC-3'). These primers were designed to incorporate *Sma*I and *Eco*RI restriction sites into the 5' and 3' ends of the sense strand, respectively. PCR was conducted as follows: 1 cycle of 94 °C for 5 min and 30 cycles of 94 °C for 1.5 min, 58 °C for 1.5 min and 72 °C for 2 min, followed by a final cycle of 10 min at 72 °C. The resulting PCR product was digested with *Sma*I and *Eco*RI, gel purified, ligated into gel purified pGEX-2T vector (Pharmacia, Uppsala, Sweden) that had previously been digested with the same enzymes, and electroporated into the *Escherichia coli* strain BL21. Colonies were selected via growth on Luria-Bertani (LB) agar/ampicillin (50 μ g/ml) plates and the presence of the insert was verified via colony PCR, using appropriate primers.

Protein expression, purification and gel electrophoresis

LB broth (600 ml) containing ampicillin (50 μ g/ml) was inoculated with 60 ml of overnight single-colony cultures. The expression of the fusion protein was obtained as described by Hilbert and co-workers [21] with minor modifications. In particular, 10% (v/v) glycerol, 0.5 mg/ml leupeptin, 0.5 mM Pefabloc and 0.7 mg/ml pepstatin (Roche, Mannheim, Germany) were added to the sonication buffer. Protein purification and gel electrophoresis of the fusion protein were performed as described by Hilbert and co-workers [21]. The purified fusion protein was used for endonuclease nicking and AP-endonuclease-activity assays.

Nicking and AP-endonuclease assays

Supercoiled DNA in pBluescript KS (pBS) was prepared by double banding in caesium chloride/ethidium bromide [12]. The AP-endonuclease activity was determined by measuring the conversion of partially depurinated supercoiled plasmid into an open circle form as described by Feng and co-workers [10] and using a synthetic oligonucleotide (AP28) containing an AP site according to the method described by Haltiwanger and co-workers [22]. The AP28 oligonucleotide (5'-TGGCGAAAGGGGUTGTGCTGCAAGGCG-3') and its complementary strand used as substrate were obtained from PRIMM (Milan, Italy).

RESULTS

The METEX gene is embedded in the DNA MTase gene

To determine the structure of the DNA MTase gene, a sea urchin genomic library was screened with the *Pl-MTase* cDNA probe [13] and several positive clones were isolated. We succeeded in determining the genomic structure (F. Aniello, G. Villano,

Figure 1 METEX gene structure and Southern-blot analysis

(A) Schematic representation of a region of the DNA MTase gene. The METEX region is depicted in the oblique-striped box, the exon 7 of the DNA MTase in the vertical-striped box and thin black lines indicate the introns. The restriction sites *Hae*III (H) and *Hind*III (Hi) are also indicated. (B) Genomic DNA was digested with *Hind*III (lane 1) and *Hae*III (lane 2). The blot was hybridized with the Met1 cDNA probe. (C) The same blot as in (B) stripped and hybridized with the DNA MTase 560 probe. Bar corresponds to 100 bp.

M. Corrado, A. Locascio, M. T. Russo, S. D'Aniello, L. Fucci and M. Branno, unpublished work) and a long open reading frame was found (Figure 1A) inside the intron 7 (GenBank Accession Number AJ418712), which is lacking in *Pl-MTase* cDNA clones. This coding gene was designated METEX. Southern blots were done to determine the copy number of the METEX gene and to confirm its linkage to *Pl-MTase* gene. Figure 1(B) shows a blot containing sea urchin genomic DNA digested with *Hind*III (lane 1) and *Hae*III (lane 2) restriction enzymes and hybridized with the Met1 cDNA probe. Each lane contains two labelled bands, as expected from the genomic structure (Figure 1A), implying that METEX is a single-copy gene. Figure 1(C) shows the same blot hybridized with the probe 560 corresponding to exon 7 and part of intron 7 of DNA MTase gene (Figure 1A). The results show that identical bands were labelled in the two blots, indicating that METEX is closely linked to the *Pl-MTase* gene.

Temporal expression of METEX

METEX expression was determined in sea urchin embryos and in different adult tissues by Northern-blot analysis. Using the 0.85 kb METEX fragment as probe, different signals were detected in male gonads and in late stages of development (Figure 2A). No signals were present in female gonads, in four blastomeres and in the morula stages. Figure 2(A) shows that at least four different mRNAs, ranging from 1.2–2.8 kb in length, were synthesized and were regulated during development. The same blot, was hybridized with the DNA MTase cDNA probe [13], to analyze the correlation of METEX gene expression with the expression of DNA MTase (Figure 2B). The time course of this accumulation was as described previously [13] and clearly different from that of METEX. As a control for RNA loading, the same blot was hybridized with an H3.3 histone cDNA probe [23] (Figure 2C).

Figure 2 Expression of METEX analysed by Northern blot

Each lane contains 10 µg of poly(A)⁺ RNA from female (fg) and male (mg) gonads (lanes 1 and 2), from 4 blastomeres (4b), morula (m), blastula (bl), gastrula (ga) and pluteus (pl) stages (lanes 3–7). The blot was hybridized with the ³²P-labelled 0.85 kb METEX DNA fragment obtained by PCR (A), with the DNA MTase cDNA probe (B), and with an H3.3 histone cDNA probe as control (C).

Isolation of cDNA clones and their spatial expression

To identify METEX cDNA clones, a cDNA library prepared from male gonads was screened using a METEX fragment as a probe. Two positive clones were isolated; the sequence analysis showed that the first, named Met1 (GenBank Accession Number AJ418711), was 966 bp in length, while the second, named Met2 (GenBank Accession Number AJ418710), was 1208 bp in length. Met2 is identical with METEX, while Met1 differs at the 3' end. Specifically it is lacking the last 393 nucleotides present in METEX and has 24 nucleotides different in the ORF and 133 nucleotides different in the 3' untranslated region (UTR).

The longest ORF from the putative initiation codon was 681 bp (nucleotides 155–835) in the Met1 clone, whilst the longest in Met2 was 993 bp (nucleotides 38–1030). The predicted product of the METEX gene, corresponding to Met2 cDNA, is 330 amino acids in length, with a calculated molecular mass of 38 kDa. The predicted product of Met1 is 226 amino acids in length; in comparison with Met2 it shows eight different amino acids at the C-terminus and it is 104 amino acids shorter.

In-situ-hybridization assays were then carried out to determine the localization of METEX message at the pluteus stage (Figure 3). No stained cells were detected in control embryos hybridized with a sense probe (results not shown). With an anti-sense probe of Met2 the transcripts were detected in the midgut and hindgut regions of the archenteron, in the cells along the anus and in the cells of the oral arms (Figures 3C and 3D).

Figure 3 *In situ* localization of the Met2 transcript

Whole-mount pluteus embryos were hybridized with a digoxigenin-labelled Met2 antisense RNA probe and stained with alkaline phosphatase. (A) Schematic representation of a *P. lividus* embryo at pluteus stage. (B) Embryos at pluteus stage observed with Nomarski optics. (C) Aboral and (D) oral view of a pluteus embryo showing the presence of Met2 transcripts in the midgut and hindgut regions, in the cells along the anus and in the cells of the oral arms. g, gut; a, anus; oa, oral arms. 1 cm = 25 micron.

Structural analysis of the protein

Comparison of the predicted product of the METEX with protein in the database (Swissprot) revealed a very low level of sequence similarity with some lyase proteins. The Homstrad database [15] of homologous protein families was searched using the program FUGUE. The predicted protein was assigned to the family of AP1 endonucleases and then named SuAP1 (sea urchin endonuclease III homologue 1). The alignment provided by FUGUE [16] with minor manual modifications is shown in Figure 4.

The family of AP1 endonucleases includes two proteins whose structure was solved, the *Escherichia coli* AP-DNA-repair endonuclease [24], indicated as ExoIII, and the human DNA repair endonuclease [25], named HAP1. Besides these, the family of AP1 endonucleases includes several other protein sequences, but only those whose pair-wise identical residues were lower than 30%, were included in Figure 4.

ExoIII and HAP1 belong to the class of 4 layer α/β -sandwiches which also includes bovine DNase I [26]. Therefore the sequence of DNaseI was structurally aligned to those of ExoIII and HAP1 using the program COMPARE (www.cryst.bioc.cam.ac.uk) and included in the alignment shown in Figure 4. The active site residues of ExoIII and HAP1, which are conserved in AP1 endonucleases, are shaded. As can be seen, although the overall number of identical residues is very low, most of the residues belonging to the active site of AP1 endonuclease, as well as those of the PROSITE signatures, are conserved in SuAP1.

In fact, in SuAP1, the conserved residues are those which directly participate in catalysis, i.e. Asp-274 and His-302

(alignment positions 332 and 372) or indirectly, in binding a metal ion, i.e. Glu-100 (alignment position 46). Most of the polar residues: Asn-71, Arg-102, Arg-157, Asp-211, Asn-213 (alignment positions 8, 48, 158, 224, 226), which directly interact by hydrogen or ionic bond with the substrate, are conserved as well as those which form a hydrophobic pocket: Val-133, Tyr-175, Tyr-262, Leu-273 (alignment positions 112, 180, 314, 331).

Once a reliable alignment with sequences of structurally characterized protein was obtained, we built a model for SuAP1. The best model obtained is shown in Figure 5, using the program Modeller [29] which allows the usage of multiple template structures. We used the structures of *E. coli* and human AP1 endonuclease, alone or together, to build the model and the structure of bovine DNase I as a control. The quality of the four models was judged using the program ProsaII [30]. It was found that the best model (Figure 5) could be obtained using both the *E. coli* and human AP1 endonucleases.

As was expected, the model for SuAP1 had two central layers of β strands (dark and light grey arrows respectively), surrounded by two layers of α helix (dark and light grey cylinders respectively). To clarify the correlation between the protein sequence and secondary structure elements, dark and light grey arrows and cylinders were drawn on the sequence of SuAP1 in Figure 4.

β strands are conserved, and this finding supports the idea that they form the inner core of the structure. α helices are more variable and, in correlation with these secondary elements, deletions and insertions are frequently encountered not only in SuAP1 but also in the other AP1 endonucleases.

Figure 4 Structure-based amino acid sequence alignment for AP endonuclease

Alignment of the amino acid sequence of SuAP1 endonuclease with those of putatively homologous proteins. Residues in the AP1 endonuclease sequences, which obey the consensus patterns of the family, are in bold. The active site residues are shadowed. The helices belonging to different layers are indicated as light and dark cylinders and the strands belonging to different layers are indicated as light and dark arrows, respectively. Numbers in parenthesis refer to the first and the last amino acid residues in each line of the respective protein sequences. The sequences are SuAP1 (sea urchin, GenBank Accession Number A418710), ExoIII (*Escherichia coli*, Swiss-Prot #P09030), HAP1 (*Homo sapiens*, Swiss-Prot #P27695), Pacrc (*Pseudomonas aeruginosa*, GenBank Accession Number Q51380), ScAPN2 (*Saccharomyces cerevisiae*, [27]), ARP (*Arabidopsis thaliana*, GenBank Accession Number O65642), L1 (*Homo sapiens*, [28]), OIRT (*Oryzias latipes*, GenBank Accession Number O93520), CeRT (*Caenorhabditis elegans*, GenBank Accession Number O44767), DNaseI (*Bos taurus*, [25]).

Figure 5 Three-dimensional model of AP endonuclease

The best three-dimensional model for SuAP1 protein was obtained using the Modeller program. Two central layers of β strands are indicated as dark and light arrows, surrounded by two layers of dark and light α helices.

The high quality of the model was assessed using the program Prosa II and represents an independent proof of the quality of the alignment and consequently of the assignment of SuAP1 to the family of AP1 endonucleases.

Expression of SuAP1, purification of the recombinant protein, and endonuclease activity

To demonstrate that the isolated cDNA Met2 encoded a functional homologue of endonuclease III, a GST fusion protein was constructed consisting of amino acids residues of the complete ORF fused to the C-terminus of the 30 kDa GST protein. SDS/PAGE analysis of the isopropyl 1-thio- β -D-galactopyranoside-induced affinity-purified fusion protein (Figure 6A) revealed a predominant 70 kDa full-length protein. Two additional lower-molecular-mass protein species were present in the purified preparation. We believe these to be fragments of the 68 kDa protein that arose through abortive synthesis of the protein or proteolysis occurring before, during and possibly after cell lysis and affinity purification, due to the action of contaminating cellular protease.

Endonuclease activities were assayed by measuring the ability of the SuAP1 to convert supercoiled plasmid into open circles and to cut AP oligonucleotides. In Figure 6(B) the nicking activity is illustrated as a function of protein concentration. Supercoiled substrate DNA (0.2 μ g) was incubated with increasing amounts of purified fusion protein ranging from 0.5–1.5 μ g (lanes 2–4). Almost all the supercoiled DNA was converted into a mixture of open-circle DNA as well as linear plasmid DNA by the fusion-protein addition (Figure 6B, lane 4). The nicking activity was linked to SuAP1, because no activity

Figure 6 *In-vitro* expression and nuclease activity of the recombinant SuAP1

(A) SDS/PAGE analysis of the GST fusion protein is shown. Lane 1, total lysate from uninduced *E. coli* transformed by the pGEX-2T vector containing the SuAP1 coding sequence; lane 2, total lysate from induced *E. coli* transformed by the recombinant pGEX-2T vector; lane 3, purified GST fusion protein after elution from affinity medium; M, molecular mass markers (Gibco BRL Bench Mark™ Prestained Protein Ladder). (B) Nicking activity. Lane 1, substrate pBS DNA partially depurinated, no protein added; lane 2, addition of 0.5 μ g of purified GST fusion protein; lane 3, addition of 1.0 μ g of purified GST fusion protein; lane 4, addition of 1.5 μ g of purified GST fusion protein; lane 5, addition of 1.5 μ g of purified GST protein; M, molecular mass markers (1 kb, Boehringer). sc, supercoiled plasmid; oc, open (nicked) circular plasmid; l, linear plasmid. (C) AP endonuclease activity. 5'-end-labelled AP28 oligonucleotide with a natural AP site, untreated, treated with 1.5 units of Endo III and Endo IV (Trevigen, Gaithersburg, U.S.A.) or with 2 μ g of purified GST fusion protein for 30 min at 37 °C. Reactions were analysed directly on a 12% (w/v) denaturing polyacrylamide gel. The positions of the substrate and products are also indicated.

was detected in the presence of the GST purified protein used as control, expressed in the same system (Figure 6B, lane 5).

To demonstrate that the expressed protein has AP endonuclease activity and to identify the class it belongs to, an assay using as control Endo III and Endo IV was carried out. A double-stranded oligonucleotide substrate, containing a single AP site in a defined position, was prepared by uracil glycosylase excision as described by Haltiwanger and co-workers [22]. The AP endonuclease belonging to class I cleaves on the 3' side of the AP site by β -elimination, while that belonging to class II cleaves on the 5' side of the AP site. To confirm that SuAP1 belongs to the class I AP endonuclease, products generated by SuAP1, by Endo III and Endo IV activities were compared on an acrylamide gel. Figure 6(C) shows that the products of oligonucleotide

cleavage by SuAP1 and Endo III were identical and different from that generated by Endo IV, consistent with SuAP1 incision on the 3' side of the AP site; SuAP1, therefore, belongs to class I of the AP endonuclease.

DISCUSSION

The present study shows that a class I AP endonuclease gene is located within intron 7 of the DNA MTase gene of the *P. lividus* sea urchin. The gene is present as a single copy in the genome and is transcribed during development, as demonstrated by the presence of the corresponding mRNA. The DNA MTase and SuAP1 nested genes are transcribed in the same direction but their expression profiles differ throughout development: DNA MTase is mainly expressed in embryos during the early stages of development, whilst SuAP1 is expressed during the later stages. Therefore, two genes that are functionally unrelated and differentially regulated exist in this novel arrangement. It could be that the METEX gene is regulated by its own promoter located upstream but inside intron 7. Such gene organization is not generated by a random event of genome rearrangement in a single sea urchin adult, but it is present in different individuals of the same species living in various Mediterranean areas (results not shown).

The gene is transcribed into various mRNAs, probably produced by alternative splicing events mainly in the 3'UTR. In fact, the two isolated cDNA clones differ only for a small part of their coding region and in the 3'UTR; it is still to be demonstrated that the Met1 putative translated protein is really functional.

The SuAP1 recombinant protein, expressed by Met2 in an *E. coli* system (Figure 6A), has AP endonuclease activity since it is capable of hydrolysing a 28-mer double-stranded DNA fragment containing an internal AP site (Figure 6C) and is capable of nicking supercoiled plasmids (Figure 6B).

When a pyrimidinic residue in cellular DNA becomes modified by oxidation, reduction or hydration of its 5,6-double bond, repair is initiated by a DNA glycosylase activity, that cleaves the N-glycosyl bond of the damaged residues, releasing the modified base and creating an abasic (AP) site in the DNA backbone. Such DNA glycosylase activities have been identified in bacteria, yeast and mammalian species [31–38]. The first such enzyme described was *E. coli* endonuclease III, which was identified not on the basis of its DNA glycosylase activity, but rather because it nicked UV-irradiated DNA [39]. It has been determined that the enzyme nicks DNA, catalysing β -elimination of the 3'-phosphate group at the AP site formed as a result of the enzyme's DNA glycosylase activity [40–42]. Enzymes with such activities are termed DNA glycosylase/AP lyases.

Our biochemical data confirmed the hypothesis that the SuAP1 was likely to be endowed with nuclease activity, based on the alignment built between its deduced amino acid sequence and the sequences of the family of AP1 endonucleases (Figure 4). It was interesting to note that several residues involved in the catalysis, by binding metal ions in interactions with hydrogen or ionic bonds, are conserved throughout the family. The structures of the ExoIII [24] and HAP1 [25] were solved and, in light of the high degree of conservation of critical residues, it is likely that the common core has a similar three-dimensional structure. In fact, our model (Figure 5) has similar folds to that of ExoIII and HAP1, despite its limited sequence similarities.

Relationships between the AP proteins and the SuAP1 come out also from the phylogenetic tree made using the conserved regions aligned in Figure 4. This analysis (Figure 7) showed the existence of two major branches, one corresponding to proteins with AP lyase activity where the SuAP1 is located, and the other

Figure 7 Phylogenetic analysis

Unrooted phylogenetic tree obtained by the comparison of the proteins used in the alignment (Figure 4). The tree was inferred using the Drawtree program from the PHILIP 3.52 computer package [43].

to the proteins with endonuclease and reverse transcriptase activities such DNase I and the LINE family. The L1 retrotransposon elements are close in evolution to proteins with AP activity, despite the fact that these proteins lack retrotransposition characteristics. It could be speculated that both descended from the same gene coding for protein that during evolution has acquired new enzymic activity.

ORFs, which are either freestanding within the intron or in frame with the upstream exon, have already been demonstrated in group I and group II introns. Some of these ORFs encode essential cellular proteins, but the majority either function as maturases, to promote splicing, or contribute to intron mobility. Endonucleases, that are specific for recognition sites with overall asymmetry, are encoded by group I introns of mitochondrial, chloroplast, protist nuclei and bacteriophage genomes. In the case of the gene described in this paper the endonuclease is encoded by an intron of the nuclear genome and it is more related to DNA glycosylase/AP lyase.

Until now no localization of the AP transcripts in sea urchin has been reported in the literature. We demonstrated that the SuAP1 is zygotic, is regulated during development and is also tissue specific. In fact, the transcript is present in the male gonads and appears during the late stages of development. It may be significant that expression of SuAP1 mRNA is elevated in intestine tissue, since digestive process can damage the genome. Reactive oxygen species are generated at a high rate by normal cellular processes, such as digestion. It seems likely that SuAP1 is an important component of the cellular defence machinery for counteracting cytotoxic and mutagenic damage to the genome.

REFERENCES

- Burke, J. M. and Rajbhandary, V. L. (1982) Intron within the large rRNA gene of *M. crassa* mitochondria: a long open reading frame and a consensus sequence possibly important in splicing. *Cell* (Cambridge, Mass.) **31**, 509–520
- Sugita, M., Shinozaki, K. and Sugiura, M. (1985) Tobacco chloroplast trnA^{Lys} (UUU) gene contains a 2.5 kilobase-pair intron: an open reading frame and a conserved boundary sequence in the intron. *Proc. Natl. Acad. Sci. U.S.A.* **82**, 3557–3561
- Larowska, J., Jacq, C. and Shonimski, P. P. (1980) Sequence of introns and flanking exons in wild-type and box3 mutants of cytochrome b reveals an interlaced splicing protein coded by an intron. *Cell* (Cambridge, Mass.) **22**, 333–348

- 4 Macreadie, I. G., Scott, R. M., Zinn, A. R. and Butow, R. A. (1985) Transposition of an intron in yeast mitochondria requires a protein encoded by that intron. *Cell* (Cambridge, Mass.) **41**, 395–402.
- 5 Swalla, B. J., Just, M. A., Pederson, E. L. and Jeffery, W. R. (1999) A multigene locus containing the *Manx* and *bobcat* genes is required for development of chordate features in the ascidian tadpole larva. *Development* (Cambridge, U.K.) **126**, 1643–1653.
- 6 Henikoff, S., Keene, M. A., Sloan, J. S., Bleskan, J., Hards, R. and Patterson, D. (1986) Multiple purine pathway enzyme activities are encoded at a single genetic locus in *Drosophila*. *Proc. Natl. Acad. Sci. U.S.A.* **83**, 720–724.
- 7 Henikoff, S., Keene, M. A., Fichtel, K. and Fristrom, J. W. (1986) Gene within a gene: nested *Drosophila* genes encode unrelated proteins on opposite DNA strand. *Cell* (Cambridge, Mass.) **44**, 33–42.
- 8 Malik, H. S., Burke, W. D. and Eickbush, T. H. (1999) The age and evolution of non-LTR retrotransposable elements. *Mol. Biol. Evol.* **16**, 793–805.
- 9 Martin, F., Maranon, C., Olivares, M., Alonso, C. and Lopez, M. C. (1995) Characterization of a non-long terminal repeat retrotransposon cDNA (L1Tc) from *Trypanosoma cruzi*: homology of the first ORF with the ape family of DNA repair enzyme. *J. Mol. Biol.* **247**, 49–59.
- 10 Feng, Q., Moran, J. V., Kazazian, H. H. and Boeke, J. D. (1996) Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. *Cell* (Cambridge, Mass.) **87**, 905–916.
- 11 Barzilay, G., Walker, L. J., Robson, C. N. and Kickson, I. D. (1995) Site-directed mutagenesis of the human DNA repair enzyme HAP1: identification of residues important for AP endonuclease and RNase H activity. *Nucleic Acids Res.* **23**, 1544–1550.
- 12 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.
- 13 Aniello, F., Locascio, A., Fucci, L., Geraci, G. and Branno, M. (1996) Isolation of cDNA clones encoding DNA methyltransferase of sea urchin *P. lividus*: expression during embryonic development. *Gene*, **178**, 57–61.
- 14 Sanger, F., Nicklen, S. and Coulson, A. R. (1975) A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J. Mol. Biol.* **94**, 441–448.
- 15 Mizuguchi, K., Deane, C. M., Blundell, T. L. and Overington, J. P. (1998) Homstrad: a database of protein structure alignments for homologous families. *Protein Sci.* **7**, 2469–2471.
- 16 Shi, J., Blundell, T. L. and Mizuguchi, K. (2001) FUGUE: sequence-structure homology recognition using environment-specific substitution tables and structure-dependent gap penalties. *J. Mol. Biol.* **310**, 243–257.
- 17 Branno, M., De Franciscis, V. and Tosi, L. (1983) *In vitro* methylation of histones in sea urchin nuclei during early embryogenesis. *Biochim. Biophys. Acta* **741**, 136–142.
- 18 Chomczynski, P. and Sacchi, N. (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.* **162**, 156–159.
- 19 Ransick, A., Ernst, S., Britten, R. J. and Davidson, E. H. (1993) Whole mount *in situ* hybridization shows Endo16 to be a marker for the vegetal plate territory in sea urchin embryos. *Mech. Develop.* **42**, 117–124.
- 20 Kirchhamer, C. V. and Davidson, E. H. (1996) Spatial and temporal information processing in the sea urchin embryo. Modular and intramodular organization of the *Cy11a* gene cis-regulatory system. *Development* (Cambridge, U.K.) **122**, 333–348.
- 21 Hilbert, T. P., Chung, W., Boorstein, R. J., Cunningham, R. P. and Teebor, G. W. (1997) Cloning and expression of the cDNA encoding the human homologue of the DNA repair enzyme, *Escherichia coli* endonuclease III. *J. Biol. Chem.* **272**, 6733–6740.
- 22 Haltiwanger, B. M., Karpnich, N. O. and Taraschi, T. F. (2000) Characterization of classII apurinic/aprimidinic endonuclease activities in the human malaria parasite *Plasmodium falciparum*. *Biochem. J.* **345**, 85–89.
- 23 Fucci, L., Aniello, F., Branno, M., Biffali, E. and Geraci, G. (1994) Isolation of a new H3.3 histone variant cDNA of *P. lividus* sea urchin: sequence and embryonic expression. *Biochim. Biophys. Acta* **1219**, 539–542.
- 24 Mol, C. D., Kuo, C.-F., Thayer, M. M., Cunningham, R. R. and Tainer, J. A. (1995) Structure and function of the multifunctional DNA-repair enzyme exonuclease III. *Nature* (London) **374**, 381–386.
- 25 Gorman, M. A., Morera, S., Rothwell, D. G., de La Fortelle, E., Mol, C. D., Tainer, J. A., Hickson, I. D. and Freemont, P. S. (1997) The crystal structure of the human DNA repair endonuclease HAP1 deoxyribose at DNA abasic sites. *EMBO J.* **16**, 6548–6558.
- 26 Suck, D. and Oefner, C. (1986) Structure of DNase I at 2.0 Å resolution suggests a mechanism for binding to and cutting DNA. *Nature* (London) **321**, 620–625.
- 27 Unk, I., Haracska, L., Johnson, R. E., Prakash, S. and Prakash, L. (2000) Apurinic endonuclease activity of yeast Apn2 protein. *J. Biol. Chem.* **275**, 22427–22434.
- 28 Hattori, M., Kuhara, S., Takenaka, O. and Sakaki, Y. (1986) L1 family of repetitive DNA sequence in primates may be derived from a sequence encoding a reverse transcriptase related protein. *Nature* (London) **321**, 625–628.
- 29 Sali, A. and Blundell, T. L. (1993) Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **234**, 779–815.
- 30 Sippl, M. J. (1993) Recognition of errors in three dimensional structures of proteins. *Proteins* **17**, 355–362.
- 31 Brent, T. P. (1973) A human endonuclease activity for gamma-irradiated DNA. *Biophys. J.* **13**, 399–401.
- 32 Bacchetti, S. and Benne, R. (1975) Purification and characterization of an endonuclease from calf thymus acting on irradiated DNA. *Biochim. Biophys. Acta* **390**, 285–297.
- 33 Duker, N. J. and Teebor, G. W. (1975) Different ultraviolet DNA endonuclease activity in human cells. *Nature* (London) **255**, 82–84.
- 34 Ness, I. F. and Nissen-Meyer, J. (1978) Endonuclease activities from a permanently established mouse cell line that act upon DNA damaged by ultraviolet light, acid and osmium tetroxide. *Biochim. Biophys. Acta* **520**, 111–121.
- 35 Demple, B. and Linn, S. (1980) DNA N-glycosylases and UV repair. *Nature* (London) **287**, 203–208.
- 36 Cunningham, R. P. and Weiss, B. (1985) Endonuclease III (nth) mutants of *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* **82**, 474–478.
- 37 Doetsch, P. W., Helland, D. E. and Haseltine, W. A. (1986) Mechanism of action of a mammalian DNA repair endonuclease. *Biochemistry* **25**, 2212–2220.
- 38 Boorstein, R. J., Hilbert, T. P., Cadet, J., Cunningham, R. P. and Teebor, G. W. (1989) Formation and stability of repairable pyrimidine photohydrates in DNA. *Biochemistry* **28**, 6164–6170.
- 39 Radman, M. (1976) An endonuclease from *Escherichia coli* that introduces single polynucleotide chain scission in ultraviolet-irradiated DNA. *J. Biol. Chem.* **251**, 1438–1445.
- 40 Bailly, V. and Verly, W. G. (1987) *Escherichia coli* endonuclease III is not an endonuclease but a beta-elimination catalyst. *Biochem. J.* **242**, 565–572.
- 41 Kim, J. and Linn, S. (1989) Purification and characterization of UV endonuclease I and II from murine plasmacytoma cells. *J. Biol. Chem.* **264**, 2739–2745.
- 42 Mazumder, A., Gerlt, J. A., Absolon, J. M., Stubbe, J., Cunningham, R. P., Withka, J. and Bolton, P.-H. (1991) Stereochemical studies of the beta-elimination reactions at aldehydic abasic sites in DNA: endonuclease III from *Escherichia coli*, sodium hydroxide, and Lys-Trp-Lys. *Biochemistry* **30**, 1119–1126.
- 43 Felsenstein, J. (1989) PHILIP: phylogeny interference package Ver. 3.4. *Cladistics* **5**, 164–166.

Received 19 December 2001/22 March 2002; accepted 15 April 2002

Published as BJ Immediate Publication 15 April 2002, DOI 10.1042/BJ20011857