# Characterization of monocot and dicot plant *S*-adenosyl-L-methionine decarboxylase gene families including identification in the mRNA of a highly conserved pair of upstream overlapping open reading frames

Marina FRANCESCHETTI*[1], Colin HANFREY*, Sonia SCARAMAGLI*[1], Patrizia TORRIGIANI†, Nello BAGNI†, Daniel BURTIN*[2] and Anthony J. MICHAEL*[3]

*Division of Food Safety Science, Institute of Food Research, Norwich Research Park, Colney, Norwich NR4 7UA, U.K., and †Dipartimento di Biologia e.s., Università di Bologna, Via Irnerio 42, 40126 Bologna, Italy

*S*-Adenosyl-L-methionine decarboxylase (AdoMetDC; EC 4.1.1.50) is one of the key regulatory enzymes in the biosynthesis of polyamines. Isolation of genomic and cDNA sequences from rice and *Arabidopsis* had indicated that this enzyme is encoded by a small multigene family in monocot and dicot plants. Analysis of rice, maize and *Arabidopsis* AdoMetDC cDNA species revealed that the monocot enzyme possesses an extended C-terminus relative to dicot and human enzymes. Interestingly, we discovered that all expressed plant AdoMetDC mRNA 5′ leader sequences contain a highly conserved pair of overlapping upstream open reading frames (uORFs) that overlap by one base. The 5′ tiny uORF consists of two or three codons and the 3′ small uORF encodes 50–54 residues. Sequences of the small uORFs are highly conserved between monocot, dicot and gymnosperm AdoMetDC mRNA species and the C-terminus of the plant small uORFs is conserved with the C-terminus of nematode AdoMetDC uORFs; such a conserved arrangement is strongly suggestive of a translational regulatory mechanism. No introns were found in the main AdoMetDC proenzyme ORF from any of the plant genes encoding AdoMetDC, whereas introns were found in conserved positions flanking the overlapping uORFs. The absence of the furthest 3′ intron from the *Arabidopsis* gene encoding AdoMetDC2 suggests that this intron was lost recently. Reverse-transcriptase-mediated PCR analysis of the two *Arabidopsis* genes for AdoMetDC indicated that AdoMetDC1 is abundant and ubiquitous, whereas the gene for AdoMetDC2 is expressed preferentially in leaves and inflorescences. Investigation of recently released *Arabidopsis* genome sequences has revealed that in addition to the two genes encoding AdoMetDC isolated as part of the present work, four additional genes are present in *Arabidopsis* but they are probably not expressed.

Key words: polyamine, translational regulation.

## INTRODUCTION

*S*-Adenosyl-L-methionine decarboxylase (AdoMetDC) is a key enzyme in the biosynthesis of the polyamines spermidine and spermine. Polyamines are small polycations found in all cells and are essential for basic cellular processes such as protein translation and cell proliferation [1]. An indication of the essential role of polyamines has been demonstrated in the minimal genome of *Mycoplasma genitalium*. All recognizable polyamine biosynthetic genes are absent from this minimal genome [2] but global transposon mutagenesis [3] has shown that the three open reading frames (ORFs) corresponding to the *potA*, *potB* and *potC* genes of the polyamine transporter [4] are indispensible for cell viability, indicating that polyamines are part of the minimal metabolic repertoire required for independent life.

The enzymic function of AdoMetDC is to remove the carboxylate group of *S*-adenosyl-L-methionine to produce decarboxylated *S*-adenosylmethionine. An aminopropyl group from decarboxylated *S*-adenosylmethionine is transferred by spermidine synthase to putrescine (1,4-diaminobutane) to form spermidine. Similarly, spermine synthase transfers another aminopropyl group to spermidine to form spermine. Mammalian AdoMetDC activity is negatively regulated in response to increasing polyamine levels by means of a translational mechanism involving a small upstream ORF (uORF) in the 5′ leader sequence of the mRNA [5–11]. The small uORF encodes the hexapeptide MAGDIS (single-letter amino acid codes) located towards the 5′ end of the 330-base 5′ leader sequence. Polyamine-induced translational repression of AdoMetDC is thought to be mediated by the peptide product of the MAGDIS uORF. The MAGDIS peptide might interact with polyamines to regulate the translational efficiency of the AdoMetDC message through the polyamine-dependent binding of the MAGDIS peptide to a target in the translational machinery. Polyamines seem to cause ribosomes to stall after translating the MAGDIS peptide, thereby greatly decreasing the efficiency of ribosome loading on the message. Recently, small uORFs potentially encoding 17-residue peptides were found in the AdoMetDC 5′ leader sequences of the nematodes *Onchocerca volvulus* [12] and *Caenorhabditis elegans* [13]. The 5′ leader sequence of plant AdoMetDCs from

Abbreviations used: AdoMetDC, *S*-Adenosyl-L-methionine decarboxylase (EC 4.1.1.50); BAC, bacterial artificial chromosome; EST, expressed sequence tag; ORF, open reading frame; uORF, upstream ORF; RT–PCR, reverse-transcriptase-mediated PCR.

[1] Present address: Dipartimento di Biologia e.s., Universitá di Bologna, Via Irnerio 42, 40126 Bologna, Italy.

[2] Present address: John Innes Centre, Norwich Research Park, Colney, Norwich NR4 7UH, U.K.

[3] To whom correspondence should be addressed (e-mail tony.michael@bbsrc.ac.uk).

The nucleotide sequence data reported will appear in DDBJ, EMBL and GenBank Nucleotide Sequence Databases under the accession numbers Y07766, AJ251899, Y07767, AJ251915, AJ252214, AJ252213, Y07765 and AJ252212.

*Catharanthus roseus* [14] and carnation [15] were found to have conserved small uORFs encoding a highly conserved peptide of approx. 50–54 residues. We show here that all recognizable expressed monocot, dicot and gymnosperm AdoMetDC 5′ leader sequences contain two uORFs of 2–3 and 50–54 amino acids that overlap by one nucleotide, suggesting a novel form of translational regulation.

Although the human, rat, mouse and *Drosophila AMD* (*AdoMetDC*) genes are interrupted by introns throughout the main ORF, there are no introns in the 5′ leader sequence [16–19]. In contrast, the potato *SAMDC* (*AdoMetDC*) gene is devoid of introns in the main ORF but has a single intron in the 5′ leader sequence [20]. By isolating genomic sequences for AdoMetDC from *Arabidopsis* and rice we show that AdoMetDC is encoded by a small gene family and each member contains introns flanking the two overlapping uORFs; however, there are no introns in the main proenzyme ORF.

A differential accumulation of steady-state mRNA levels for the *Arabidopsis* AdoMetDC1 and AdoMetDC2 in different organs of *Arabidopsis* plants is demonstrated, suggesting diverged roles for the two enzymes. Four other *Arabidopsis AdoMetDC* genes were detected in genome sequences but each of these genes seems to be defective in some respect and none appears to be expressed.

## EXPERIMENTAL

### Isolation and sequencing of cDNA and genomic clones

Two *Arabidopsis* expressed sequence tags (ESTs) potentially encoding different AdoMetDCs were identified by similarity with published AdoMetDC sequences (clones 52B6BT7 and 178M17T7 for *Arabidopsis* AdoMetDC1 and AdoMetDC2 respectively). The cDNA clones were obtained from Dr Doreen Ware (*Arabidopsis* Biological Resource Center, Ohio State University, Columbus, OH, U.S.A.). Two rice ESTs were similarly identified (clones RICS1785A and E2795_1A for rice AdoMetDC1 and AdoMetDC2 respectively) and the cDNA species were obtained from Dr Takuji Sasaki at the Rice Genome Research Program, Tsukuba, Japan. A maize EST (clone csuh00006) potentially encoding AdoMetDC was obtained from Dr Chris Baysdorfer (The Maize cDNA Program, California State University, Hayward, CA, U.S.A.). All cDNA double-stranded plasmid templates were prepared for sequencing with a Plasmid Mini Kit (Qiagen). Sequencing of templates was achieved by primer walking and sequencing was performed with the Big Dye Terminator Cycle Sequencing Kit (Applied Biosystems, Warrington, Cheshire, U.K.). Sequencing gels were run on an ABI 373 Stretch Sequencer.

Genomic clones of the *Arabidopsis AdoMetDC1* and *AdoMetDC2* genes were obtained by high-stringency screening of a Columbia ecotype lambda ZAPII genomic library (Stratagene, La Jolla, CA, U.S.A.) with the use of the corresponding cDNA species as probes. The cDNA probes were $^{32}$P-labelled by random priming as described previously [21]. After individual plaques had been isolated, the genomic insert in pBluescript was excised with the ExAssist (Stratagene) *in vivo* excision protocol of the manufacturer. The genomic inserts were sequenced with the same protocol as for the cDNA species.

Genomic clones of the two rice AdoMetDCs were obtained by PCR, with rice genomic DNA as the PCR template. For the PCR, 200 ng of rice leaf genomic DNA was employed as the template and the AdoMetDC1 sequence was amplified with the sense oligonucleotide A684 (5′-GTC TCG TGG TCG AGA GAA ACC GAG-3′) and the anti-sense oligonucleotide A655 (5′-CTC GAC AGC ATC GTC TTC-3′). The AdoMetDC2 sequence was amplified with the sense oligonucleotide A574 (5′-GAG AAA TCC GTA GGC TCA GAG-3′) and with the anti-sense oligonucleotide A382 (5′-CTC ATA ACT GGC GTA GCT GA-3′). PCRs were performed with 200 ng of genomic DNA, 0.25 $\mu$M of each primer, 0.2 mM dNTPs and 2.5 units of *Pfu* DNA polymerase (Stratagene), with 1 × *Pfu* buffer in a total volume of 100 $\mu$l. An initial denaturation stage of 94 °C for 4 min was followed by a 5 min step at 75 °C during which the polymerase was added, followed by 35 cycles of 1 min at 94 °C, 2 min at 55 °C and 2 min at 72 °C; PCR was concluded with one step for 10 min at 72 °C. Products were ligated into the *Eco*RV site of pBluescript and transformed into *Escherischia coli* XL-1 Blue competent cells (Stratagene). Genomic inserts were sequenced from plasmids as described above for the cDNA inserts.

### RNA isolation and reverse-transcriptase-mediated PCR (RT–PCR)

Total RNA was isolated from different organs of *Arabidopsis* plants grown at 22 °C with 16 h daylight in a greenhouse. After the plant material had been frozen in liquid nitrogen, the material was ground to a fine powder and total RNA was extracted as described previously [22]. First-strand cDNA was synthesized as described [21] in a total volume of 50 $\mu$l. Three reactions were performed, containing 1.1 $\mu$g and 3.3 $\mu$g of total RNA for each sample of leaf, stem, inflorescence and fruit (silique) RNA. In addition, a control reaction was performed with 3.3 $\mu$g of RNA and without reverse transcriptase. The reaction products were diluted with sterile distilled water to 1.0 ml.

The RT–PCR used the following primers: AdoMetDC1, sense primer A480 (5′-ATC TCT ATC TGG TTT GAG G-3′) and anti-sense primer A481 (5′-ATT ATG AAG ATC GGT GCA CG-3′); AdoMetDC2, sense primer A482 (5′-TTA GGG TTT CGA GTT TTC TCA-3′) and anti-sense primer A483 (5′-AAA GCT ATG GAG GGA ACT G-3′). The primers for AdoMetDC1 were designed to cross intron positions and therefore would not amplify genomic DNA, whereas the primers for AdoMetDC2 were designed to avoid intron positions and be capable of amplifying genomic DNA. Primers for AdoMetDC2 can therefore indicate the presence of contaminating genomic DNA in the first-strand cDNA samples. The RT–PCR used 5 $\mu$l of the first-strand cDNA, 200 $\mu$M dNTPs, 0.2 $\mu$M of each primer, *Pfu* polymerase buffer supplied by the manufacturer (Stratagene) and 2.5 units of *Pfu* DNA polymerase in a total volume of 100 $\mu$l; *Pfu* DNA polymerase was added at 75 °C. The PCR reaction was started with a denaturation step of 94 °C for 4 min, a 5 min step at 75 °C when the polymerase was added, and 32 cycles of 94 °C for 1 min, 55 °C for 2 min and 72 °C for 2 min followed by a single step of 72 °C for 10 min. To determine the relative abundances of the two AdoMetDC messages, 10 $\mu$l of the PCR products was size-fractionated on a 1.0 % (w/v) agarose gel.

## RESULTS

### Characterization of rice, maize and *Arabidopsis* AdoMetDC cDNA species

Similarity searching of DNA databases identified two *Arabidopsis* ESTs, two rice ESTs and one maize EST potentially encoding AdoMetDC. Clones were obtained from the relevant sequencing programmes and each cDNA was fully sequenced. The two rice cDNA species encoded AdoMetDC proenzymes of 398 residues (AdoMetDC1) and 395 residues (AdoMetDC2) and the two *Arabidopsis* cDNA species encoded proenzymes of 366 residues (AdoMetDC1) and 362 residues (AdoMetDC2); a proenzyme of

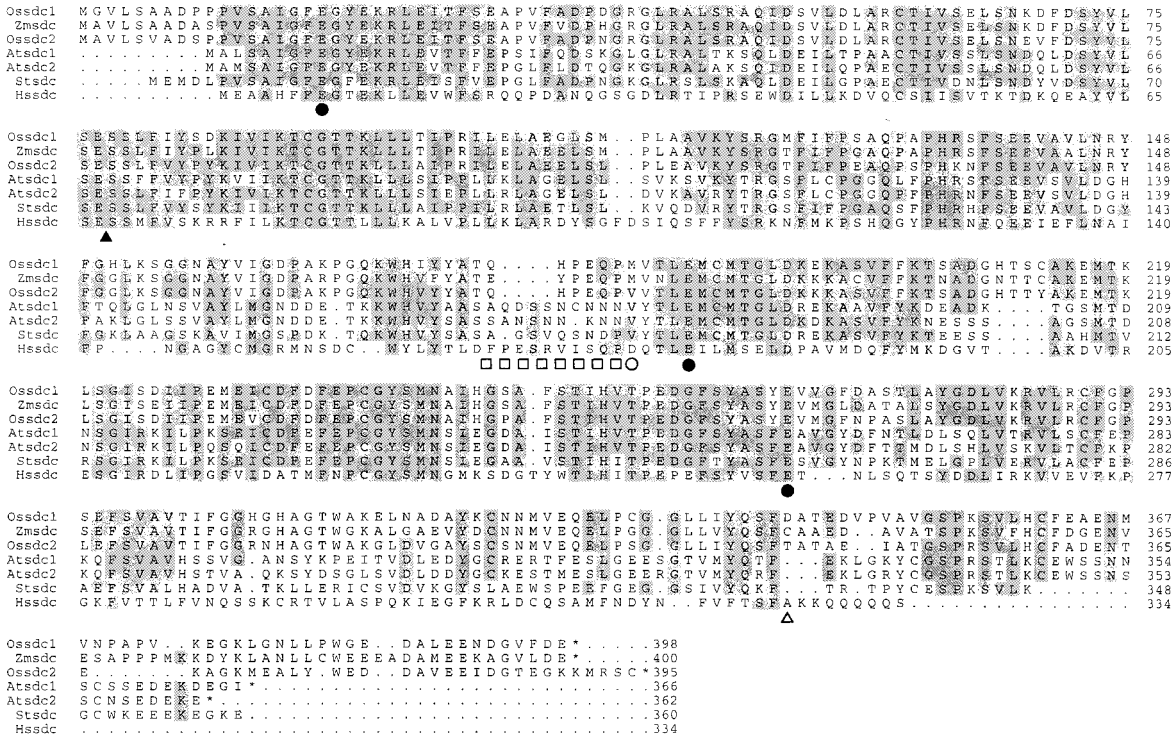

**Figure 1 Alignment of the deduced amino acid sequences of rice, maize, *Arabidopsis*, potato and human AdoMetDC proenzymes**

Sequences of the rice AdoMetDC1 [Ossdc1 (Y07766)], rice AdoMetDC2 [Ossdc2 (AJ251899)], maize AdoMetDC [Zmsdc (Y07767)], *Arabidopsis* AdoMetDC1 [Atsdc1 (U63633)], *Arabidopsis* AdoMetDC2 [Atsdc2 (AJ251915)], potato AdoMetDC [Stsdc (Z11680)] and human AdoMetDC [Hssdc (M21154)] are aligned; accession numbers are given in parentheses. Conserved glutamate residues necessary for the stimulation of proenzyme processing of the human AdoMetDC by putrescine are indicated by filled circles. The aspartic residue necessary for the stimulation of processing in the human enzyme by putrescine is indicated by an open circle [23]. Location of the proenzyme-processing site is indicated by the filled triangle. The position at which the human and potato proteins can be truncated without affecting processing is indicated by an open triangle [23]. A disordered loop that connects the two structurally similar halves of the human enzyme is indicated by open squares [25]. Shading represents the conservation of identical amino acids.

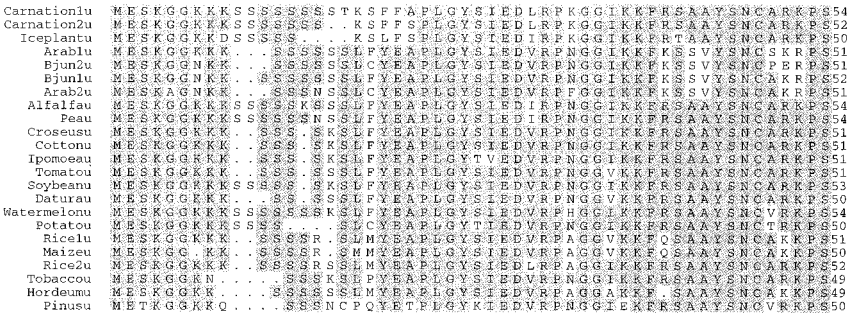

**Figure 2 Alignment of the deduced amino acid sequences of the small uORF in the plant AdoMetDC mRNA 5′ leader sequences**

Alignment of the small uORF amino acid sequences deduced from mRNA and EST sequences (accession numbers in parentheses): Carnation1u, *Dianthus caryophyllus* AdoMetDC1 (U38526); Carnation2u, *Dianthus caryophyllus* AdoMetDC2 (U38527); Iceplantu, *Mesembryanthemum crystallinum* (EST-AI822872); Arab1u, *Arabidopsis thaliana* AdoMetDC1 (U63633); Arab2u, *Arabidopsis thaliana* AdoMetDC2 (AJ251899); Bjun1u, *Brassica juncea* AdoMetDC1 (U80916); Bjun2u, *Brassica juncea* AdoMetDC2 (X95729); Rice1u, *Oryza sativa* AdoMetDC1 (Y07766); Rice2u, *Oryza sativa* AdoMetDC2 (AJ251899); Maizeu, *Zea mays* (Y07767); Hordeumu, *Hordeum chilense* × *Tritordeum turgidum* conv. Durum (X83881); Potatou, *Solanum sogarandinum* (AA076673); Daturau, *Datura stramonium* (Y07768); Croseusu, *Catharanthus roseus* (U12573); Cottonu, *Gossypium hirsutum* (EST-AI728571); Ipomoeau, *Ipomoea nil* (U64927); Tomatou, *Lycopersicon esculentum* (EST-AI483261); Soybeanu, *Glycine max* (EST-AI442381); Alfalfau, *Medicago trunculata* (EST-AA661003); Peau, *Pisum sativum* (U60592); Watermelonu, *Citrullus lanatus* (EST-AI563097); Tobaccou, *Nicotiana tabacum* (AF033100); Pinusu, *Pinus taeda* (EST-AI725223). Shading represents conservation of identical amino acids.

400 residues was encoded by the maize cDNA. Each monocot proenzyme possessed an extended C-terminus relative to the *Arabidopsis* AdoMetDCs (Figure 1). The monocot proenzymes were approximately 20 residues longer than the dicot enzymes and the maize protein was 51 residues longer than that from human. Both the monocot and dicot C-termini were particularly rich in acidic amino acids. An aspartic residue in the human AdoMetDC (Asp-174), identified by Xiong et al. [23] as being necessary for stimulation of the processing of the human proenzyme by putrescine, is absent from all of the plant sequences,

```
Arab1/2          gcg tga atg aga tta tgATG GAG T
B.jun1/2         gcg tga atg aga tta tgATG GAA T
rice1            aag tgg atg tac taATG GAG TCA A
rice2            aaa ttg atg cac taATG GAG TCC A
tobacco            g tga atg atc taATG GAG TCG A
maize            aat ttg atg tac taATG GAG TCT A
Pinus taeda      ggc tga atg ttc tgATG GAG ACC A
Ipomoea nil      ggt gta atg aac tgATG GAG TCT A
Hordeum          tag tga atg ttc taATG GAG TCA A
Ice plant        ttg tga atg atc taATG GAG TCT A
carnation1       tat tga atg agc taATG GAA TCA A
carnation2       tat tga atg agc taATG GAG TCA A
C.roseus         gag tga atg atc taATG GAG TCT A
Pea              gac tga atg agc taATG GAG TCT A
Datura           ttg tga atg atc taATG GAG TCA A
Alfalfa          gac tga atg agc taATG GAG TCT A
Cotton           gac tga atg atc taATG GAG TCT A
Watermelon       ttg cga atg att taATG GAG TCA A
Tomato           ttg tga atg atc taATG GAG TCA A
Potato           gaa tcc atg aac taATG GAG TCT A
Soybean          gac tga atg aac taATG GAG TCT A
```

**Figure 3 cDNA sequences of the junction between the tiny and small uORFs in the plant AdoMetDC mRNA 5′ leaders**

The sequence of the tiny uORF is indicated in lower-case bold letters and that of the beginning of the small uORF in capital letters with the initiating ATG of the small uORF in bold. For an explanation of the genes and accession numbers see the legend to Figure 2. The junction sequences of the two *Arabidopsis* AdoMetDCs and of the two *Brassica juncea* AdoMetDCs are identical.

whereas the three glutamic residues (Glu-11, Glu-178 and Glu-256 [24]) necessary for stimulation of processing of the human preprotein by putrescine are conserved. The crystal structure of the human enzyme indicates that there are two similar domains linked by a disordered loop structure [25]. A region of the plant proenzymes corresponding to the disordered loop is notable for the lack of conservation between monocot and dicot sequences and between plant and human sequences.

## Conservation of tiny and small overlapping uORFs

Each of the plant AdoMetDC cDNA species contains a long 5′ leader sequence: *Arabidopsis* AdoMetDC1, 505 bp; *Arabidopsis* AdoMetDC2, 481 bp; rice AdoMetDC1, 539 bp; rice AdoMetDC2, 536 bp; maize AdoMetDC, 366 bp. It was noted previously [14] that plant AdoMetDC 5′ leader sequences contained a conserved small uORF encoding approx. 50 amino acid residues [14]. We have detected the small uORF in 23 plant AdoMetDC mRNA species and ESTs that include four monocot sequences and one gymnosperm sequence (Figure 2). The amino acid sequence of the encoded peptide is highly conserved, with only one region of variation towards the N-terminus where the variation is due primarily to varying numbers of successive serine residues. Although there is no similarity between the plant and mammalian AdoMetDC small uORFs, the last four amino acids of the plant small uORFs (Lys/Arg,Lys/Arg,Pro,Ser) resemble the C-terminal sequence of the two nematode AdoMetDC 5′ leader sequence uORFs of 17 residues each (Arg,Arg,Pro,Pro) [12,13]. The plant small uORF is positioned towards the middle of the long 5′ leader sequence.

We noticed that in all the plant AdoMetDC 5′ leader sequences there is a tiny uORF upstream of and overlapping the small uORF (Figure 3). The tiny uORF encodes only two amino acid residues, the second of which is not conserved; the two *Arabidopsis* and two *Brassica juncea* tiny uORFs encode three residues and are evolutionarily close species. In all cases, the last A nucleotide of the tiny uORF stop codon is the first nucleotide of the initiating ATG of the small uORF, suggesting that if the ATG of the tiny uORF is identified by the translational machinery, it is unlikely that the ATG of the small uORF will be recognized. For the tiny uORF the translational initiation sequence context of the ATG is always very poor (Figure 3), whereas that of the small uORF is invariably in a better context [26]. The +4 position relative to the A of the ATG is never the optimal G for the tiny uORF, whereas the same nucleotide position is always a G for the small uORF. A tiny uORF is conserved in the sequence of the *Pinus taeda* EST sequence, indicating that the overlapping arrangement of uORFs in the 5′ leader sequence of the plant AdoMetDC predated the origin of the gymnosperm and angiosperm divergence. Outside the overlapping uORFs in the 5′ leader sequence there is relatively little conservation of sequence between AdoMetDC family members in the same species or between species.

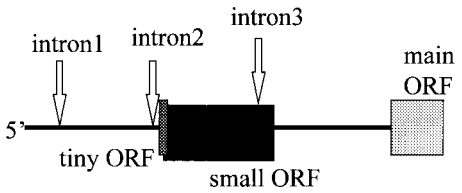


**Figure 4 Schematic representation of the genomic structure of the plant AdoMetDC 5′ leader**

The relative positions of the introns and overlapping tiny and small uORFs are depicted. Only the beginning of the main AdoMetDC proenzyme ORF is shown. In the *Arabidopsis AdoMetDC2* gene, intron 3 is absent. The presence of intron 1 was determined only for the two genes from *Arabidopsis* and the gene from *Ipomoea nil* (accession number U64927).

**Table 1 Sizes and positions of introns in plant *AdoMetDC* gene 5′ leader sequences**

The accession numbers for the other genomic sequences are: tobacco, AF033100; *S. tuberosum*, S74514; *Ipomoea nil*, U64927; carnation1, U94786. The inter-ORF figure is the distance from the end of the small uORF to the beginning of the main proenzyme ORF.

| | Sizes of the *AdoMetDC* gene 5′ leader components (bp) | | | | | |
|---|---|---|---|---|---|---|
| Gene | Intron1 | Exon2 | Intron2 | Between intron2 and small uORF | Intron3 | Inter-ORF |
| *Arabidopsis*1 | 103 | 95 | 397 | 18 | 95 | 179 |
| *Arabidopsis*2 | 113 | 70 | 256 | 19 | 0 | 157 |
| Rice1 | – | – | 940 | 15 | 83 | 154 |
| Rice2 | – | – | 961 | 19 | 91 | 118 |
| *Ipomoea nil* | 123 | 79 | 660 | 13 | 108 | 200 |
| Carnation1 | – | – | 598 | 13 | 108 | 156 |
| *S. tuberosum* | – | – | – | 12 | 95 | 196 |
| Tobacco | – | – | – | 12 | 99 | 213 |

## Genomic structure of the *Arabidopsis* and rice *AdoMetDC* genes

The corresponding genomic clones for the *Arabidopsis* AdoMetDC1 and AdoMetDC2 cDNA species were isolated from a lambda ZAP genomic library. PCR with rice genomic DNA was used to isolate the genomic region corresponding to the cDNA species of the two rice *AdoMetDC* genes. None of the genes contained any introns in the 3′ UTR or in the main proenzyme ORF; however, a small intron was found in a highly conserved position (Figure 4) in the C-terminus of the small uORF (except for *Arabidopsis* AdoMetDC2). A longer intron was present just upstream of the tiny uORF, being larger in the two rice genes than in the two *Arabidopsis* genes or other dicot AdoMetDC genomic sequences found in the database (Table 1).

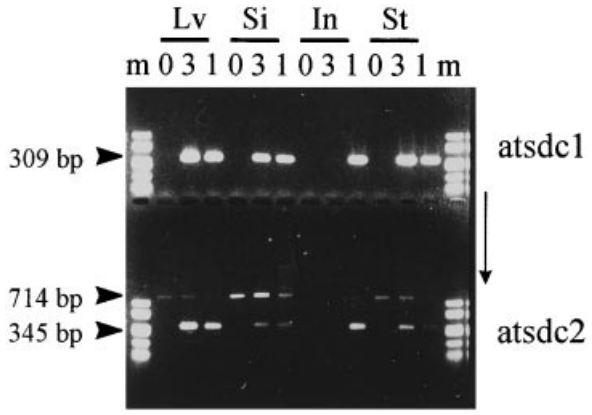


**Figure 5 Differential accumulation of the *Arabidopsis* AdoMetDC1 and AdoMetDC2 transcripts detected by RT–PCR**

First-strand cDNA was prepared with greenhouse-grown *Arabidopsis* total RNA from rosette leaves (Lv), siliques (Si), inflorescences (In) and stems (St) by using either 3 or 1 $\mu$g of template RNA as shown (lanes labelled 3 and 1 respectively); in addition, a control reaction containing 3 $\mu$g of template RNA but without added reverse transcriptase was included (lanes 0). Lanes m, lanes containing molecular mass markers. Primers specific for AdoMetDC1 or AdoMetDC2 were used to amplify gene-specific products. The AdoMetDC2 primers were designed to avoid intron junctions to show the mRNA-specific product and the contaminating genomic DNA-derived products (the mRNA product is 345 bp and the genomic DNA product is 714 bp). The product from the AdoMetDC1 (309 bp) is mRNA-specific because primers cross intron junctions.

The two *Arabidopsis* genes and the previously reported *Ipomoea nil* gene contained a third, small intron at the 5′ end of the leader sequence, upstream of the other two introns. So far, the *Arabidopsis AdoMetDC2* gene seems to be unique in lacking the small uORF C-terminal intron. Conservation of the small uORF flanking introns in monocot and dicot *AdoMetDC* genes suggests that the *Arabidopsis AdoMetDC2* gene has lost the small uORF C-terminal intron. The *Arabidopsis AdoMetDC1* has recently been sequenced as part of the *Arabidopsis* genome project and is located on chromosome 3 [bacterial artificial chromosome (BAC) clone F16B3] and the *AdoMetDC2* gene is located on a BAC clone from chromosome 5 (BAC clone F1N13).

### RT–PCR analysis of the *Arabidopsis AdoMetDC* genes

RT–PCR was used to distinguish between the expression of the *Arabidopsis AdoMetDC1* and *AdoMetDC2* genes in rosette leaves, stems, inflorescences and fruit (siliques). The primers used to amplify AdoMetDC1 were designed to cross intron–exon junctions and would therefore amplify only the cDNA. Primers for AdoMetDC2 were designed to avoid introns and could therefore amplify both genomic and cDNA sequences. The AdoMetDC2 genomic PCR product provided a control for further distinguishing genomic and mRNA products. No RT–PCR product corresponding to AdoMetDC1 or AdoMetDC2 was detected when the reverse transcriptase was omitted but a genomic PCR product of the expected size (714 bp) was detected for AdoMetDC2 (Figure 5). The RT–PCR indicated that the AdoMetDC1 message was abundant and ubiquitous, whereas AdoMetDC2 transcript was found at a much lower level with higher accumulation in leaves and inflorescences than in siliques and stems. The relative transcript levels of the two genes in *Arabidopsis* organs reflects the relative abundance of each message in the *Arabidopsis* EST database (results not shown).

**Table 2 Location of *Arabidopsis* genomic sequences for the genes encoding AdoMetDC3, AdoMetDC4, AdoMetDC5 and AdoMetDC6**

The nucleotide coordinates for the proenzyme ORF are given except for AdoMetDC6, which was too degraded for assignment of the whole ORF; the conserved region is therefore given instead. Genomic sequences are derived from unannotated BAC clones.

| Isoform | Clone and accession no. | Chromosome | Proenzyme ORF nucleotide position |
|---|---|---|---|
| AdoMetDC3 | P1 clone MWL2 (AB025639) | 3 | 75231–74206 |
| AdoMetDC4 | BAC clone F17K4 (AC068655) | 5 | 23072–24115 |
| AdoMetDC5 | P1 clone MKP6 (AB022219) | 3 | 81707–82854 |
| AdoMetDC6 | P1 clone MTC11 (AB024038) | 3 | 51682–51923 |

### Other *AdoMetDC* genes present in the genome of *Arabidopsis*

The *Arabidopsis* genome sequence is now almost complete and recently we have detected new members of the *Arabidopsis* AdoMetDC family. Four other *AdoMetDC* genes are present in *Arabidopsis* on unannotated BAC clones, although we could not find any corresponding ESTs. This would indicate that these genes are not expressed or are expressed at a low level in a small subset of cells. Table 2 lists the genes and their corresponding BAC clones, nucleotide positions and chromosome assignments. AdoMetDC3, AdoMetDC5 and AdoMetDC6 are found on chromosome 3, whereas AdoMetDC4 is found on chromosome 5.

The gene encoding AdoMetDC3 (proenzyme N-terminus MAVSATGFEGFEKE) contains tiny and small uORFs but the tiny uORF encodes five amino acid residues and the stop codon does not overlap with the start codon of the small uORF. An intron is found in the conserved position of the small uORF C-terminus; however, the peptide sequence of the small uORF is rather diverged. The gene for AdoMetDC4 (proenzyme N-terminus MAVSGFEGFEKE) is on chromosome 5 and it does not possess a small uORF. Both AdoMetDC3 and AdoMetDC4 encode intact proenzyme ORFs of 349 and 348 residues respectively. The genes for AdoMetDC5 (N-terminus MDSQVQFFEGVEKR) and AdoMetDC6 (N-terminus not detectable) are located on chromosome 3; neither possesses a small uORF. There are three in-frame stop codons in the proenzyme sequence of AdoMetDC5, whereas AdoMetDC6 is highly diverged with a 32-residue deletion eliminating conserved motifs; it is unlikely that AdoMetDC5 or AdoMetDC6 proenzymes are enzymically active. None of the six *Arabidopsis AdoMetDC* genes are found on the same BAC clone, indicating that they are not recent adjacent duplications.

## DISCUSSION

The outstanding feature of plant *AdoMetDC* genes is the presence in the message 5′ leader sequence of overlapping tiny and small uORFs. Amino acid sequences of the small uORFs are highly conserved between monocot and dicot genes and even between angiosperms and the gymnosperm *Pinus taeda*. Mammalian AdoMetDC is translationally regulated in a cell-specific manner [5,6,9] and in response to intracellular polyamine levels [7,8,10,11]. Translational regulation of the mammalian AdoMetDC is mediated through a small uORF encoding the peptide MAGDIS [6,9]. The 5′ end of the mRNA, including the MAGDIS uORF, is also responsible for translational regulation of AdoMetDC in response to polyamine levels [8,27]. It was subsequently concluded that the MAGDIS uORF was responsible solely for translational regulation in response to

polyamine levels [10]. The amino acid sequence of the uORF and not the DNA sequence is the essential factor involved in the translational repression of mammalian AdoMetDC activity. There is a strict amino acid sequence specificity in critical positions of the peptide for inhibitory activity of the uORF [11].

If, by analogy with the mammalian AdoMetDC 5′ leader sequence, the plant small uORF is responsible for translational repression in response to polyamines, the mechanism of response to polyamines and translational repression must be different from that of the mammalian gene. The plant 5′ leader sequence contains the tiny uORF overlapping with the small uORF and it is unlikely that the ATG of the small uORF will be recognized if the tiny uORF is translated. Backwards scanning is unusual and inefficient; if the tiny uORF is translated, it is more likely that the ribosome will reinitiate at the downstream proenzyme ATG, which is sufficiently far downstream to permit efficient reinitiation [28]. We can speculate that the small uORF might be responsible for translational repression of the plant AdoMeDC and it is worth noting that the C-terminal residues of the small uORFs of the two nematode AdoMetDCs are Arg,Arg,Pro,Pro and those of the plant small uORFs are Lys/Arg,Lys/Arg,Pro,Ser. Use of the PROFILESEARCH program of the Wisconsin GCG package [29] and searching the protein database with a consensus profile of 23 small uORF sequences resulted in an identification of low similarity with numerous bacterial, chloroplastic and nuclear diverse ribosomal proteins from both subunits; chloroplastic and bacterial L1 proteins had the highest similarity scores (A. J. Michael, unpublished work). Also prominent were several diverse ribonucleases. This low-level similarity suggests a possible RNA-binding function for the small uORF product. Although the low similarity with very diverse ribosomal proteins makes it difficult to draw any conclusions about the significance of the similarity, it is salient to note that ribosomal protein genes were previously postulated to be the origin of bacterial leader sequences specifying regulatory peptides [30].

The genomic structure of the plant *AdoMetDC* genes is unusual; there are no introns in the proenzyme ORF, whereas the overlapping uORFs are flanked by introns. The furthest 3′ intron is in the C-terminus of the small uORF, in-frame, preceding the final six residues and its position is exactly conserved between monocot and dicot genes. This suggests that the intron was in position before the divergence of the monocot and dicot plants. Absence of the 3′ intron from the *Arabidopsis AdoMetDC2* gene is therefore likely to be due to loss of a pre-existing intron. The flanking arrangement of the introns around the overlapping uORFs is reminiscent of the exon capture structure postulated [31] to explain the origin of plant homeodomain–leucine-zipper transcription factors. The exon containing the plant AdoMetDC uORFs might have been captured from another transcript, thereby providing a regulatory module. All of the expressed plant genes sequenced so far contain an intron immediately 5′ to the tiny uORF; close inspection of the potato *AdoMetDC* gene sequence [20] suggests that the region annotated as the promoter is in fact the 5′ intron whose position is four nucleotides upstream of the tiny uORF ATG.

There is a considerable difference in the abundance of the *Arabidopsis* AdoMetDC1 and AdoMetDC2 mRNA species as well as in their organ distribution. AdoMetDC1 is ubiquitous and abundant compared with AdoMetDC2. The carnation CSDC16 (AdoMetDC2) transcript was more abundant and ubiquitous than the CSDC9 (AdoMetDC1) transcript [15]. One possible explanation, which probably applies to most gene families, is that one gene might respond more to environmental signals and in this respect it is interesting to note that the *AdoMetDC* gene of *Tritordeum* [32] and *Pharbitis nil* [33] is regulated by circadian rhythms and also, for *Pharbitis nil*, by light [33].

The discovery of four additional *Arabidopsis AdoMetDC* genes in the genome sequence has interesting implications for the evolution of pathways and genes. Each of the *AdoMetDC3*, *AdoMetDC4*, *AdoMetDC5* and *AdoMetDC6* genes is defective in some way. Certainly it is unlikely that AdoMetDC5 and AdoMetDC6 could be enzymically active and it is also unlikely that AdoMetDC3 and AdoMetDC4 could be regulated in the same way as AdoMetDC1 and AdoMetDC2 because the overlapping uORF arrangement in AdoMetDC3 is degraded and in AdoMetDC4 it is absent. Because there are no corresponding EST sequences for any of the four genes, it is reasonable to assume that they are probably not expressed, although this remains to be proven.

We thank John Payne for running the automated sequencer.

## REFERENCES


1 Cohen, S. (1998) A Guide to the Polyamines, Oxford University Press, Oxford
2 Fraser, C. M., Gocayne, J. D., White, O., Adams, M. D., Clayton, R. A., Fleischmann, R. D., Bult, C. J., Kerlavage, A. R., Sutton, G., Kelley, J. M. et al. (1995). The minimal gene complement of *Mycoplasma genitalium*. Science **270**, 397–403
3 Hutchinson, III, C. A., Peterson, S. N., Gill, S. R., Cline, R. T., White, O., Fraser, C. M., Smith, H. O. and Venter, C. J. (1999) Global transposon mutagenesis and a minimal mycoplasma genome. Science **286**, 2165–2169
4 Igarashi, K. and Kashiwagi, K. (1999) Polyamine transport in bacteria and yeast. Biochem. J. **344**, 633–642
5 Hill, J. R. and Morris, D. R. (1992) Cell-specific translation of S-adenosylmethionine decarboxylase mRNA: dependence on translation and coding capacity of the *cis*-acting upstream open reading frame. J. Biol. Chem. **267**, 21886–21893
6 Hill, J. R. and Morris, D. R. (1993) Cell-specific translation of S-adenosylmethionine decarboxylase mRNA: regulation by the 5′ transcript leader. J. Biol. Chem. **268**, 726–731
7 Shantz, L. M., Holm, I., Jänne, O. A. and Pegg, A. E. (1992) Regulation of *S*-adenosylmethionine decarboxylase activity by alterations in the intracellular polyamine content. Biochem. J. **288**, 511–518
8 Shantz, L. M., Viswanath, R. and Pegg, A. E. (1994) Role of the 5′-untranslated region of mRNA in the synthesis of *S*-adenosylmethionine decarboxylase and its regulation by spermine. Biochem. J. **302**, 765–772
9 Ruan, H., Hill, J. R., Fatemie-Nainie, S. and Morris, D. R. (1994) Cell-specific translational regulation of S-adenosylmethionine decarboxylase mRNA: influence of the structure of the 5′ transcript leader on regulation by the upstream open reading frame. J. Biol. Chem. **269**, 17905–17910
10 Ruan, H., Shantz, L. M., Pegg, A. E. and Morris, D. R. (1996) The upstream open reading frame of the mRNA encoding S-adenosylmethionine decarboxylase is a polyamine-responsive translational control element. J. Biol. Chem. **271**, 29576–29582
11 Mize, G. J., Ruan, H., Low, J. J. and Morris, D. R. (1998) The inhibitory upstream open reading frame from mammalian S-adenosylmethionine decarboxylase mRNA has a strict sequence specificity in critical positions. J. Biol. Chem. **273**, 32500–32505
12 Da'Dara, A. A., Henkle-Dührsen, K. and Walter, R. D. (1996) A novel *trans*-spliced mRNA from *Onchocerca volvulus* encodes a functional *S*-adenosylmethionine decarboxylase. Biochem. J. **320**, 519–530
13 Da'Dara, A. A. and Walter, R. D. (1998) Molecular and biochemical characterization of *S*-adenosylmethionine decarboxylase from free-living *Caenorhabditis elegans*. Biochem. J. **336**, 545–550
14 Schröder, G. and Schröder, J. (1995) cDNAs for S-adenosylmethionine decarboxylase from *Catharanthus roseus*, heterologous expression, identification of the proenzyme-processing site, evidence for the presence of both subunits in the active enzyme, and a conserved region in the 5′ mRNA leader. Eur. J. Biochem. **228**, 74–78
15 Lee, M. M., Lee, S. H. and Park, K. Y. (1997) Characterization and expression of two members of the S-adenosylmethionine decarboxylase gene family in carnation flower. Plant Mol. Biol. **34**, 371–382
16 Pulkka, A., Ihalainen, R., Soursa, A., Riviere, M., Szpirer, J. and Pajunen, A. (1993) Structures and chromosomal localisations of two rat genes encoding S-adenosylmethione decarboxylase. Genomics **16**, 342–349
17 Nishimura, K., Kashiwagi, K., Matsuda, Y., Jänne, O. A. and Igarashi, K. (1999) Structure and activity of mouse S-adenosylmethionine decarboxylase gene promoters and properties of the encoded proteins. Gene **238**, 343–350
18 Maric, S. C., Crozat, A. and Jänne, O. A. (1992) Structure and organization of the human S-adenosylmethionine decarboxylase gene. J. Biol. Chem. **267**, 18915–18923

19 Larsson, J. and Rasmuson-Lestander, Å. (1997) Cloning, mapping and mutational analysis of the S-adenosylmethionine decarboxylase gene in *Drosophila melanogaster*. Mol. Gen. Genet. **256**, 652–660
20 Mad-Arif, S. A., Taylor, M. A., George, L. A., Butler, A. R., Burch, L. R., Davies, H. V., Stark, M. J. R. and Kumar, A. (1994) Characterisation of the S-adenosylmethionine decarboxylase (SAMDC) gene of potato. Plant Mol. Biol. **26**, 327–338
21 Michael, A. J., Hofer, J. M. I. and Ellis, T. H. N. (1996) Isolation by PCR of a cDNA clone from pea petals with similarity to petunia and wheat zinc finger proteins. Plant Mol. Biol. **30**, 1051–1058
22 Meissner, R. and Michael, A. J. (1997) Isolation and characterisation of a diverse family of *Arabidopsis* two and three-fingered C2H2 zinc finger protein genes and cDNAs. Plant Mol. Biol. **33**, 615–624
23 Xiong, H., Stanley, B. A., Tekwani, B. L. and Pegg, A. E. (1997) Processing of mammalian and plant S-adenosylmethionine decarboxylase proenzymes. J. Biol. Chem. **272**, 28342–28348
24 Stanley, B. A., Shantz, L. M. and Pegg, A. E. (1994) Expression of mammalian S-adenosylmethionine decarboxylase in *Escherichia coli*: determination of sites for putrescine activation of activity and processing. J. Biol. Chem. **269**, 7901–7907
25 Ekstrom, J. L., Mathews, I., Stanley, B. A., Pegg, A. E. and Ealick, S. E. (1999) The crystal structure of human S-adenosylmethionine decarboxylase at 2.25 Å resolution reveals a novel fold. Structure **7**, 583–595
26 Kozak, M. (1991) Structural features in eukaryotic mRNAs that modulate the initiation of translation. J. Biol. Chem. **266**, 19867–19870
27 Suzuki, T., Kashiwagi, K. and Igarashi, K. (1993) Polyamine regulation of S-adenosylmethionine decarboxylase synthesis through the 5′-untranslated region of its mRNA. Biochem. Biophys. Res. Commun. **192**, 627–634
28 Kozak, M. (1987) Effects of intercistronic length on the efficiency of reinitiation by eukaryotic ribosomes. Mol. Cell. Biol. **7**, 3438–3445
29 Devereux, J., Haeberli, P. and Smithies, O. (1984) A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12**, 387–395
30 Gu, Z., Harrod, R., Rogers, E. J. and Lovett, P. S. (1994) Anti-peptidyl transferase leader peptides of attenuation-regulated chloramphenicol-resistance genes. Proc. Natl. Acad. Sci. U.S.A. **91**, 5612–5616
31 Schena, M. and Davis, R. W. (1994) Structure of homeobox-leucine zipper genes suggests a model for the evolution of gene families. Proc. Natl. Acad. Sci. U.S.A. **91**, 8393–8397
32 Dresselhaus, T., Barcelo, P., Hagel, C., Lörz, H. and Humbeck, K. (1996) Isolation and characterization of a Tritordeum cDNA encoding S-adenosylmethionine decarboxylase that is circadian-clock-regulated. Plant Mol. Biol. **30**, 1021–1033
33 Yoshida, I., Yamagata, H. and Hirasawa, E. (1999) Blue- and red-light regulation and circadian control of gene expression of S-adenosylmethionine decarboxylase in *Pharbitis nil*. J. Exp. Bot. **50**, 319–326

Received 10 July 2000/18 September 2000; accepted 7 November 2000