JMB

Selenoprotein Synthesis in Archaea: Identification of an mRNA Element of *Methanococcus jannaschii* Probably Directing Selenocysteine Insertion

R. Wilting, S. Schorling, B. C. Persson and A Böck*

Lehrstuhl für Mikrobiologie der Universität München, Maria-Ward-Straße 1a D-80638, München, Germany

Selenocysteine is encoded by a UGA codon in all organisms that synthesise selenoproteins. This codon is specified as a selenocysteine codon by an mRNA secondary structure, which is located immediately 3′ of the UGA in the reading frame of selenoprotein genes in Gram-negative bacteria, whereas it is located in the 3′ untranslated region of eukaryal selenoprotein genes. The location and the structure of a similar mRNA signal in archaea has so far not been determined. Seven selenoproteins were identified for the archaeon *Methanococcus jannaschii* by labelling with ^{75}Se and by SDS/polyacrylamide electrophoresis. Their size could be correlated with open reading frames possessing internal UGA codons from the total genomic sequence. One of the open reading frames, that of the VhuD subunit of a hydrogenase, possesses two UGA codons and appears to code for a selenoprotein with two selenocysteine residues. A strongly conserved mRNA element was identified that is exclusively linked to selenoprotein genes. It is located in the 3′ untranslated region in six of the mRNAs and in the 5′ untranslated region of the *fdhA* mRNA. This element, which is present in the 3′ non-translated region of two selenoprotein mRNAs from *Methanococcus voltae*, is proposed to act in decoding of the UGA with selenocysteine.

© 1997 Academic Press Limited

Keywords: archaea; selenoproteins; SECIS element; untranslated region; UGA codon

*Corresponding author

Selenocysteine insertion into selenoproteins is directed by a UGA codon plus a structural motif of the mRNA (for a review, see Baron & Böck, 1995). It has been shown that in several bacteria this mRNA element is located within the reading frame of the mRNA and that it serves as a target for binding of a specific translation factor, the SelB protein (Baron *et al.*, 1993; Kromayer *et al.*, 1996). SelB also binds selenocysteyl-tRNA[Sec] and GTP, and fulfils a function homologous to that elongation factor Tu has in inserting the 20 canonical amino acids (Forchhammer *et al.*, 1989).

Eukaryal selenoprotein mRNAs, on the other hand, carry the selenocysteine insertion element (SECIS element) in the 3′ untranslated region (for a review, see Low & Berry, 1996). It acts at considerable distance in directing the decoding of an in-frame UGA codon with selenocysteine. A minimal distance of 51 to 111 bases between the SECIS element and the 3′ end of the open reading frame is required (Martin *et al.*, 1996); no protein has been identified yet in eukarya that plays a function simi-

lar to that of SelB in the bacterial system. It is thought that the placement of the SECIS element in the untranslated region of the mRNA renders it independent of the constraint of the reading frame and thereby enables a single element to insert selenocysteine into proteins with diverse primary structures. Also, incorporation of more than one selenocysteine residue into a protein is possible by these means (Hill *et al.*, 1991).

The recent elucidation of the total genomic sequence of *Methanococcus jannaschii* provides the basis for identification of selenoproteins and selenocysteine insertion elements in an archaeon. Our approach was to label *M. jannaschii* cells with [75]Se (Cox *et al.*, 1981), in order to correlate the selenoproteins synthesised with open reading frames coding for selenoproteins. In those parts of the *M. jannaschii* genome that were suspected to contain selenoprotein genes, open reading frames (ORFs) were re-analysed with the GENEWORKS 2.5 program (IntelliGenetics, Inc.). Parameters were chosen, that were suitable to find very short ORFs.
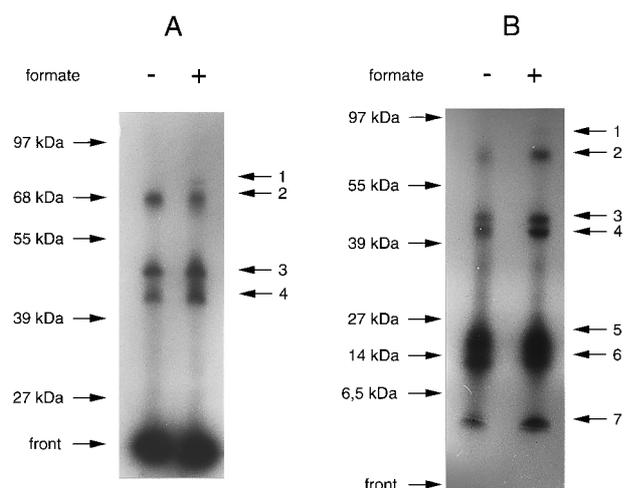
Figure 1. Autoradiographs of SDS/polyacrylamide gels according to (A) Laemmli (1970) and (B) Schägger and Jagow (1987) in which total lysates of cells from *M. jannaschii* (DSM 2661) were separated. Cells were grown in the medium described by Jones *et al.* (1983), in the absence and presence of 0.1 mM sodium formate; labelling was with $^{75}SeO_3^{2-}$ (2.2 and 4.4 μCi/ml; spec. act. 1170 and 2340 mCi/mmol). Lysates were prepared and subjected to electrophoresis as described (Cox *et al.*, 1981).

Since this program cannot predict ORFs with in-frame UGA codons, interrupted ORFs or reading frames coding proteins that are only partly homologous to known selenoproteins were manually searched for the presence of internal UGA codons. The coding region of putative selenoproteins and their 5′ and 3′ flanking segments were manually searched for consensus mRNA structural motifs. As a further test of the biological significance of these structures, it was determined whether they occur linked to selenoprotein genes in other methanogenic organisms.

Figure 1 displays the labelling pattern of *M. jannaschii* proteins separated by SDS/polyacrylamide gel electrophoresis; seven labelled bands were identified. To distinguish selenoproteins from selenium-modified RNAs, *M. jannaschii* lysates were incubated with either proteinase K or RNase. Band 5 is the only RNase-sensitive and proteinase K-resistant band, suggesting that this is a selenium-modified RNA, presumably a tRNA, because tRNAs migrate between 20 and 30 kDa in SDS/PAGE. The other bands are proteinase K-sensitive and RNase-resistant (data not shown). In their analysis of the genomic sequence (Internet: http://www.tigr.org/tdb/mdb/mjdb/mjdb.html), Bult *et al.* (1996) had recognised four putative selenoprotein genes on the basis that these ORFs have an internal UGA codon; namely, formate dehydrogenase (*fdhA*, MJ1353), heterodisulfide reductase (*hdrA*, MJ1190), formyl-methanofuran dehydrogenase (*fwdB*, MJ1194), and F_{420}-reducing hydrogenase (*fruA*, MJ0029; see Table 1). Band 1 of the autoradiograph in Figure 1A appears only when cells were grown in the presence of formate and corresponds in size to the *fdhA* gene product. Bands 2 and 3 correspond to those of *hdrA* and *fwdB*, respectively, and band 4 to that of *fruA* (Figure 1A) Since the autoradiograph indicates the existence of several small molecular mass selenoproteins (see Figure 1B), we have searched for further open reading frames with in-frame UGA codons. Three were found. One of these encodes the selenium donor protein SelD, selenophosphate synthetase, which, like the *Haemophilus influenzae* SelD (Fleischmann *et al.*, 1996), possesses a selenocysteine residue at position 17. The *Escherichia coli* SelD protein has a cysteine residue in this position (Figure 2C; and see Leinfelder *et al.*, 1990). A band corresponding to the size of SelD (39 kDa) could not be detected in the ^{75}Se-labelling experiments (Figure 1), presumably because of its low level of expression. Furthermore, two additional ORFs with internal UGA codons are present in the *M. jannaschii* genome; they are homologous to the *Methanococcus voltae* genes coding for the VhuD (Figure 2A) and VhuU (Figure 2B). These proteins are subunits of the F_{420} non-reducing hydrogenase (Halboth & Klein, 1992). Figure 2 shows the strong

Table 1. Selenoproteins and selenium-modified RNAs identified in *M. jannaschii*

Band	Putative selenoprotein:	Homologous gene	ORF	Calculated size (Da)	Apparent size (Da)
1	Formate dehydrogenase	*fdhA, M.f.*	MJ 1353	77,321	75,000 (±2%)
2	Heterodisulfide reductase	*hdrA, M.th.*	MJ 1190	71,956	72,000 (±1%)
3	Formyl-methanofuran-dehydrogenase	*fwdB, M.th*	MJ 1194	48,729	48,500 (±2%)
4	F420 red. hydrogenase	*fruA, M.v.*	MJ 0029	(46,635) 44,483[a]	44,500 (±2%)
–	Selenophosphate synthetase	*selD, E. c.*	MJ 1591	38,113	–
5	(RNA)	–	–	–	19,000 (±5%)
6	F_{420} non-reducing hydrogenase	*vhuD, M.v.*	–	14,778	14,600 (±3%)
7	F_{420} non-reducing hydrogenase	*vhuU, M.v.*	–	(5836) 3889[a]	3800 (±10%)

M.f., Methanobacterium formicicum; *M.th.,* Methanobacterium thermoautotrophicum; *M.c.,* Methanococcus voltae; *E.c.,* Escherichia coli.
[a] Calculated size after proteolytic processing.

Figure 2. Sequences of the VhuD and the VhuU proteins and of the N terminus of SelD aligned with the *M. voltae* (METVO) VhuD and VhuU and the *E. coli* (ECOLI) and *H. influenzae* (HINF) SelD sequences. Numbers indicate the nucleotides in the *M. jannaschii* (METJA) genome map (Bult *et al.*, 1996). Alignments were made by using the GENEWORKS 2.5 program (IntelliGenetics, Inc.).
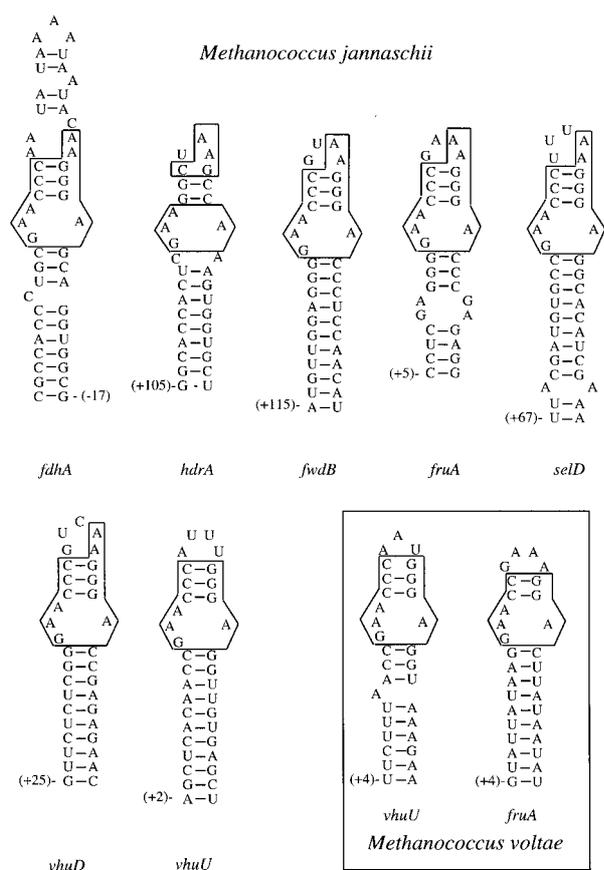
Figure 3. Secondary structure elements with putative selenocysteine insertion function. The numbers at the base of the stem−loop structure indicate the distance to the reading frame, where − indicates an upstream and + a downstream location. The motifs present in the 3′ untranslated region of two genes in *M. voltae* are shown in a separate box.

sequence conservation. As in *M. voltae* VhuU, the *M. jannaschii* sequence contains a single selenocysteine residue. However, surprisingly, VhuD from *M. jannaschii* possesses two selenocysteine residues corresponding to cysteine residues in the *M. voltae* sequence.

A search for RNA sequences and/or secondary structures within the reading frames and in the 5′ and 3′ untranslated regions was performed. A strongly conserved element was detected in the non-coding regions of all seven putative selenoprotein mRNAs (Figure 3); in six of them it was located in the putative 3′ untranslated region, whereas in the case of the *fdhA* mRNA it was located upstream of the initiation codon covering the putative Shine-Dalgarno sequence. Unlike in the *E. coli* system, there was no conserved structure within the reading frame. Also, none of a considerable number of non-selenoprotein genes from *M. jannaschii* analysed, also from the same metabolic group, carried the respective RNA structure in their non-translated region. Apart from their strikingly strong conservation and from their exclusive

linkage with selenoprotein genes there are two further facts that suggest that these structures serve a biological role. First, they are rich in G + C; *M. jannaschii* has a G + C content of only 31.4% and the intergenic regions are almost exclusively composed of A + U residues (Bult *et al.*, 1996). Second, a mesophilic methanogenic organism, *M. voltae*, carried the same structural motifs in the putative 3′ region of its two selenoprotein genes identified thus far (Figure 3).

Although the evidence presented is only that by correlation, it is compelling to assume that the structural element identified directs the in-frame UGA to be decoded as selenocysteine. It is found in the untranslated region 3′ to the reading frame in the majority of the genes, as in eukaryal selenoprotein mRNAs (Low & Berry, 1996). In the case of the *fdhA* mRNA the element is upstream, which is reminiscent of the observation that the SECIS element in eukaryal mRNAs also works *in trans* and when placed in the 5′-region, although with weak activity (Berry *et al.*, 1993). Finally, the presence of two UGA codons in a reading frame (*vhuD*) represents the only prokaryotic example of a selenoprotein with more than one selenocysteine residue. It matches the finding that the selenocysteine insertion motif is located in non-translated regions.

The identification of an archaeal selenocysteine insertion motif will facilitate the search for an archaeal homologue of SelB and will enable experiments to analyse its mode of function.

References

Baron, C. & Böck, A. (1995). The selenocysteine-inserting tRNA species: structure and function. In *tRNA: Structure, Biosynthesis and Function* (Söll, D. & RhajBhandary, U. L., eds), pp. 529–544, American Society for Microbiology, Washington DC.

Baron, C., Heider, J. & Böck, A. (1993). Interaction of translation factor SELB with the formate dehydrogenase selenopolypeptide mRNA. *Proc. Natl Acad. Sci. USA,* **90**, 4181–4185.

Berry, M. J., Banu, L., Harney, J. H. & Larsen, P. R. (1993). Functional characterization of the eukaryotic SECIS elements which direct selenocysteine insertion at UGA codons. *EMBO J.* **12**, 3315–22.

Bult, C. J., White, O., Olsen, G. J., Zhou, L., Fleischmann, R. D., Sutton, G. G., Blake, J. A., FitzGerald, L. M., Clayton, R. A., Gocayne, J. D., Kerlavage, A. R., Dougherty, B. A., Tomb, J.-F., Adams, M. D., Reich, C. I., Overbeek, R., Kirkness, E. F., Weinstock, K. G., Merrick, J. M., Glodek, A., Scott, J. L., Geoghagen, N. S., Weidman, M., Fuhrmann, J. F., Nguyen, J. L., D., Utterback, T. R., Kelley, J. M., Peterson, J. D., Sadow, P. W., Hanna,

M. C., Cotton, M. D., Roberts, K. M., Hurst, M. A., Kaine, Borodovsky, B. P., M., Klenk, H.-P., C, Fraser, M., Smith, H. O., Woese, C. R. & Venter, J. C. (1996). Complete genome of the methanogenic archaeon *Methanococcus jannaschii*. *Science,* **273**, 1058–1073.

Cox, J. C., Edwards, E. S. & DeMoss, J. A. (1981). Resolution of distinct selenium-containing formate dehydrogenases from *Escherichia coli*. *J. Bacteriol.* **154**, 1317–1324.

Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., Bult, C. J., Tomb, J.-F., Dougherty, B. A., Merrick, J. M., Mckenney, K., Sutton, G., Fitzhugh, W., Fields, C. A., Gocayne, J. D., Scott, J. D., Shirley, R., Liu, L.-I., Glodek, A., Kelley, J. M., Weidman, J. F., Phillips, C. A., Spriggs, T., Hedblom, E., Cotton, M. D., Utterback, T. R., Hanna, M. C., Nguyen, D. T., Saudek, D. M., Brandon, R. C., Fine, L. D., Fritchman, J. L., Fuhrmann, J. L., Geoghagen, N. S. M., Gnehm, C. L., Mcdonald, L. A., Small, K. V., Fraser, C. M., Smith, H. O. & Venter, J. C. (1995). Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science,* **269**, 496–512.

Forchhammer, K., Leinfelder, W. & Böck, A. (1989). Identification of a novel translation factor necessary for incorporation of selenocysteine into protein. *Nature,* **342**, 453–456.

Halboth, S. & Klein, A. (1992). *Methanococcus voltae* harbors four gene clusters potentially encoding two [NiFe] and two [NiFeSe] hydrogenases, each of the cofactor F_{420}-reducing or F_{420}-non-reducing types. *Mol. Gen. Genet.* **233**, 217–224.

Hill, K. E., Lloyd, R. S., Yang, J.-G., Read, R. & Burk, R. F. (1991). The cDNA for rat Selenoprotein P contains 10 TGA codons in the open reading frame. *J. Biol. Chem.* **266**, 10050–10053.

Jones, W. J., Leigh, J. A., Mayer, F., Woese, C. R. & Wolfe, R. S. (1983). *Methanococcus jannaschii* sp. nov., an extremely thermophilic methanogen from a submarine hydrothermal vent. *Arch. Microbiol.* **136**, 254–261.

Kromayer, M., Wilting, R., Tormay, P. & Böck, A. (1996). Domain structure of the prokaryotic selenocysteine-specific elongation factor SELB. *J. Mol. Biol.* **263**, 413–420.

Laemmli, U. K. (1970). Cleavage of structurall proteins during the assembly of the the head of the bacteriophage T4. *Nature,* **227**, 680–685.

Leinfelder, W., Forchhammer, K., Veprek, B., Zehelein, E. & Böck, A. (1990). *In vitro* synthesis of selenocysteinyl-tRNA$_{UCA}$ from seryl-tRNA$_{UCA}$: involvement and characterisation of the *selD* gene product. *Proc. Natl Acad. Sci. USA,* **87**, 543–547.

Low, S. C. & Berry, M. J. (1996). Knowing when to stop: selenocysteine incorporation in eukaryotes. *Trends Biochem. Sci.* **21**, 203–208.

Martin, G. W., Harney, J. W. & Berry, M. J. (1996). Selenocysteine incorporation in eukaryotes: insights into mechanism and efficiency from sequence, structure, and spacing proximity studies of the type 1 deiodinase SECIS element. *RNA,* **2**, 171–182.

Schägger, H. & von Jagow, G. (1987). Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the seperation of proteins in the range from 1 to 100 kDa. *Anal Biochem.* **166**, 368–379.

Edited by D. E. Draper