Acta Biochimica Polonica

Vol. 45 No. 3/1998

669–676

QUARTERLY

Cloning and sequencing of cDNA encoding the rice methionyl-tRNA synthetase[○][□]

Marzanna Deniziak[1], Marc Mirande[2] and Jan Barciszewski[1✉]

[1]*Institute of Bioorganic Chemistry of the Polish Academy of Sciences, 60-704 Poznań, Poland,* [2]*Laboratoire d'Enzymologie et Biochimie Structurales, UPR 9063 du Centre National de la Recherche Scientifique, 91190 Gif-sur-Yvette, France*

Received: 21 September, 1997; revised: 17 July, 1998; accepted: 28 July, 1998

Key words: aminoacyl-tRNA synthetases, plant methionyl-tRNA synthetase, rice MetRS, molecular cloning, λ library screening

Three overlapping clones of cDNA, Mos43, Mos28 and Mos60, coding for methionyl-tRNA synthetase were obtained by screening the *Oryza sativa* λgt11 library. Their nucleotide sequence of 2850 bp was determined. The deduced amino-acid sequence of the isolated clones contains a HLGN and KFSKS motifs, which are conserved for this family of enzymes and have been proposed to be the signature sequences for class I aminoacyl-tRNA synthetases. A comparison of the rice MetRS primary structure with those deposited in EMBL/GenBank points to its high homology to yeast, human and *Caenorhabditis elegans* MetRSs. Interestingly, a great similarity of its C terminus to endothelial-monocyte-activating polypeptide II (EMAPII) and yeast protein G4p1 was observed.

[○]This work was partly supported by the Polish-French Biotechnology Centre (CNRS, KBN). M.D. was partly supported by the "Jumelage Franco-Polonais" program from CNRS.

[□]GenBank accession numbers: *Oryza sativa* MetRS cDNA – AF040700; MetRSs: *Caenorhabditis elegans* – Z73427, *Saccharomyces cerevisiae* – V01316, *Saccharomyces cerevisiae* mitochondrial – X14629, *Saccharomyces pombe* – AB004537, *Saccharomyces pombe* mitochondrial – Z98978, *Homo sapiens* – X94754, *Methanococcus jannaschii* – U67567, *Methanobacterium thermoautotrophicum* – AE000841, *Archaeoglobus fulgidus* – AE001003, *Haemophilus influenzae* – U32807, *Borrelia burgdorferi* – AE001160, *Escherichia coli* – U0007, *Bacillus stearothermophilus* – X57925, *Bacillus subtilis* – D26185, *Thermus thermophilus* – M64273, *Helicobacter pylori* – AE000557, *Thermotoga maritima* – U76417, *Synechocystis* sp. – D64002, *Mycoplasma pneumoniae* – AE000015, *Mycobacterium tuberculosis* – Z94752, *Mycoplasma genitalium* – U39680, *Arabidopsis thaliana* chloroplastic – Y13943; *Saccharomyces cerevisiae* G4p1 – U31348; *Homo sapiens* EMAPII – U10117.

[✉]To whom correspondence should be addressed: Prof. dr hab. Jan Barciszewski, Institute of Bioorganic Chemistry of the Polish Academy of Sciences, Z. Noskowskiego 12/14, 60-704 Poznań, Poland; tel. (48 61) 852 8503 ext. 132; fax (48 61) 852 0532; e-mail: jbarcisz@ibch.poznan.edu.pl

Abbreviations: aa, amino acid(s); bp, base pair(s); EMAPII, endothelial-monocyte-activating polypeptide II; MetRS, methionyl-tRNA synthetase; aaRSs, aminoacyl-tRNA synthetases.

Aminoacyl-tRNA synthetases (aaRSs) are a family of housekeeping enzymes, which catalyze the ATP-dependent esterification of specific amino acid at the 3′ end of its cognate tRNA. Although they carry out the same reaction, aaRSs differ significantly in their primary sequence as well as in size and quaternary structure. Amino-acid sequences of these enzymes contain conserved motifs, which are a basis for dividing them into two classes of ten members each [1, 2]. The class I enzymes show two consensus amino-acid sequence motifs: His-Ile-Gly-His (HIGH) and Lys-Met-Ser-Lys-Ser (KMSKS) [3, 4]. They bind ATP through an α/β domain called the Rossman fold, which is common to many nucleotide binding proteins [5]. The aaRSs of the class II share three other sequence motifs (I, II and III), which occupy key spatial and functional positions in a conserved active site, formed around an antiparallel β sheet [6]. Interestingly, this classification coincides with biochemical data and class I of aaRSs catalyzes amino acid charging to 2′-OH of the terminal adenosine residue of tRNA before its isomerisation to 3′-OH, whereas class II enzymes add an amino acid directly to 3′-OH [7, 8].

While the structure-function relationships of bacterial aaRSs have been a subject of numerous studies, a present knowledge concerning eukaryotic aminoacyl-tRNA synthetases is rather limited. This is particularly true to plant aaRSs. Generally, the quaternary structure of individual aaRS is well conserved among prokaryotes and eukaryotes. However, there are some exceptions. *Escherichia coli* MetRS is a homodimer, each subunit of which consists of two distinct functional domains, a catalytic domain and a C-terminal extension essential for dimerization [9]. Removal of 120 amino-acid residues at the C terminus of the polypeptide chain by mild proteolysis generates a monomeric fully active enzyme [10]. In contrast, *Saccharomyces cerevisiae* cytoplasmic MetRS is a monomer lacking the C terminal extension [11]. However, it has an N-terminal extension of about 200 amino acids, which is essential for activity and stability of the enzyme [12]. It has been found that yeast MetRS is associated with a protein called Arc1p or G4p1 [13, 14] which shows also a tRNA binding capacity. In fact, that protein strongly increases the apparent affinity of MetRS for $tRNA^{Met}$ and may be required for specific aminoacylation *in vivo* [13]. Interestingly, the C-terminus of Arc1p showed a high homology to that of bacterial MetRS [13]. On the other hand, nematode (*Caenorhabditis elegans*) MetRS polypeptide chain, deduced from its nucleotide sequence, covers an enzymatic part as well as the C-terminal Arc1p-like domain [15].

In addition, it has been established that MetRSs from higher eukaryotes are associated in supramolecular multi-enzyme complexes consisting of nine aminoacyl-tRNA synthetases and three other peptides [16–18]. However, the detailed function of these complexes is largely unknown.

Untill now limited data are available on plant aaRSs. Their structural organization in multi-aaRS complexes have not been demonstrated. It has been only suggested that the plant enzymes are organized in a way different from that found in *Drosophila melanogaster* or mammals (reviewed in [19]).

To fill in a gap in a knowledge on plant aminoacyl-tRNA synthetases, we have started a project of cloning and sequencing of plant aaRSs. The aim of the present work was to clone a coding sequence of plant methionyl-tRNA synthetase.

MATERIALS AND METHODS

A rice cDNA clone C2054 was a gift of Dr. Yoshiaki Nagamura (NIAR/STAFF, Japan). The 605 bp *Xho*I fragment obtained by digestion of pBluescriptII SK(+) containing this insert was labeled by random oligonucleotide priming using [α^{32}P]dATP (Amersham), and used as a probe for screening a λgt11 *Oryza sativa* cDNA library purchased from Clontech

(U.S.A.). Approximately 6×10^5 of recombinant phages were plated on *E. coli* Y1090r$^-$. Inserts from positive clones were amplified using primers complementary to the β-galactosidase portion of the λgt11 template. The nucleotide sequences of used λgt11 the forward and reverse primers were: 5'-GGT GGC GAC GAC TCC TGG AGC CCG-3' and 5'-TTG ACA CCA GAC CAA CTG GTA ATG-3', respectively. Restriction analysis of PCR products was performed with *Eco*RI and *Rsa*I endonucleases (Boehringer). Selected inserts were partially sequenced with AmpliCycle Sequencing Kit (Perkin Elmer) and λgt11 primers. The obtained results were compared with a content of EMBL/GenBank using BLASTX program for local alignment [20]. The purified λgt11 DNAs containing inserts selected for sequencing were digested with *Eco*RI endonuclease. The isolated inserts were further subcloned into M13 bacteriophage vector. For Mos60 only the 5' end of the insert of 312 bp (the product of *Eco*RI–B*gl*II digestion) was ligated. *E. coli* JM101Tr strain was used for transformation and phage propagation. To determine the nucleotide sequence of obtained cDNAs, nested sets of deletion clones were generated. The DNA sequencing of both strands was carried out by the standard Sanger method [21] with the T7 Sequencing Kit (Pharmacia). The homology of rice sequence to known MetRSs, as well as to G4p1 and EMAPII was found with BESTFIT and PILEUP GCG's programs [22].

RESULTS AND DISCUSSION

Molecular cloning and protein expression studies, as well as recent genome sequencing projects have led to identification of over twenty MetRSs primary structures. Most of them are of eubacterial origin (*Escherichia coli, Bacillus stearothermophilus, Bacillus subtilis, Thermus thermophilus, Haemophilus influenzae, Helicobacter pylori, Thermotoga maritima, Synechocystis* sp., *Mycoplasma pneumoniae, Mycobacterium tuberculosis, Mycoplasma genitalium, Borrelia burgdorferi*). Three originate from archaea (*Methanococcus jannashii, Methanobacterium thermoautotrophicum, Archaeoglobus fulgidus*) [9, 23–36]. Four eukaryotic MetRSs – from *Saccharomyces cerevisiae, Saccharomyces pombe, Caenorhabditis elegans* and *Homo sapiens* have been sequenced [11, 15, 37, 38]. Three organellar MetRSs were also identified: mitochondrial of *S. cerevisiae* and *S. pombe*, and chloroplastic from *Arabidopsis thaliana* [39]. Since there are no structural data on the plant enzyme, we decided to determine the coding sequence of rice MetRS. Our study were initiated by identification of cDNA EST-type sequence (DDBJ accession number D23020) as being homologous to C terminus of *E. coli* MetRS. A 5'-proximal region of this clone was used as a probe for screening of cDNA library from rice. Sixteen positive clones were obtained, their inserts were PCR amplified and analysed by restriction digestions. The partial sequences of seven selected inserts were compared with data bases. This study resulted in identification of 1228 bp DNA of clone Mos11. PCR amplified Mos11 was digested by *Eco*RI and *Sac*I endonucleases and 417 bp DNA fragment obtained in this way was a probe in second library screening to clone a missing part of the sequence. Eleven positive transformants were isolated and analyzed as above. Two clones: Mos28 (1693 bp) and Mos43 (1409 bp) were selected. The digestion of Mos28 insert with *Bam*HI endonuclease enabled us to isolate a fragment of 767 bp, which was used furthermore to screen the library in order to identify the 5' end of the rice sequence. Out of ten positive clones, a cDNA of about 1700 bp (Mos60) was isolated.

Three overlapping inserts, Mos43, Mos28 and Mos60, were selected for nucleotide sequence determination (Fig. 1). The rice cDNA is 2850 bp long. It comprises an open reading frame for 804 amino-acid protein, which con-

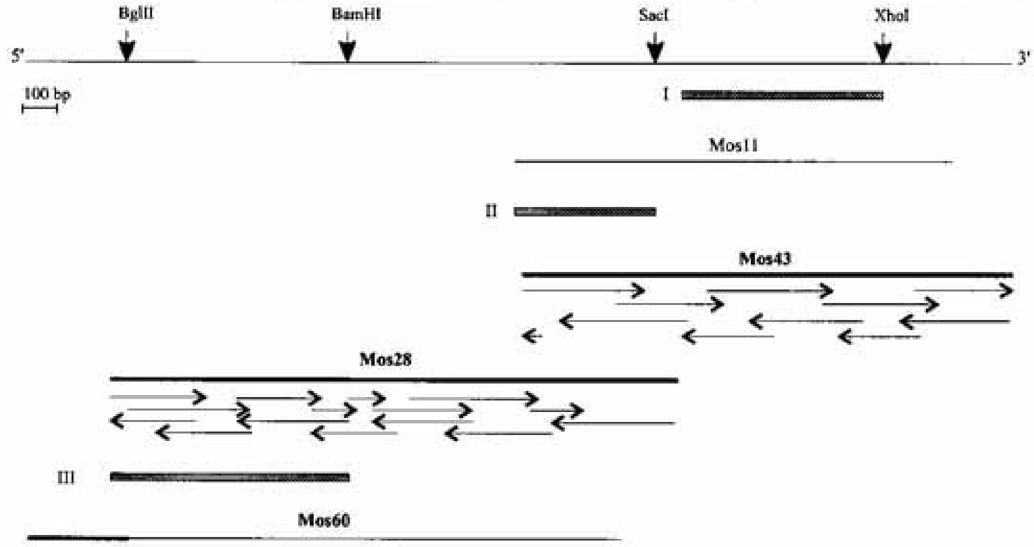

Figure 1. The cDNA for rice MetRS was cloned during three cycles of library screening.

The molecular probes present on the scheme are marked as I, II and III; probes II and III were prepared from 5' regions of previously isolated clones by digestions with restriction enzymes; Mos11, Mos28, Mos43, Mos60 – isolated clones; the cDNA fragments selected for sequence analysis are in bold; the length and direction of sequencing for cDNA fragments (nested deletions) of Mos28 and Mos43 is marked by arrows.

```
                                                 *            *
mrsmge    LSHGWIVDGEGRKMSKSLNNVISPEQLIDQF..GVDGT.RYCLLKEMRLD.KDNR
mrsmpn    LSHGWIVDGNGHKMSKSLGNVISPEELLAQF..GVDGT.RYCLLKEMRLD.KDNR
mrsatc    FGHGFLT.KDGMKMGKSLGNTLEPFELVQKF..GPDAV.RYFFLREV.EFGNDGD
mrssyn    FGHGFLT.KDGQKMGKSLGNTVDPLDLINRY..GEDAF.RYYFLKEI.EFGKDGD
mrsbst    FGHGWLLMKDG.KMSKSKGNVVDPVMIIDRY..GLDAL.RYYLLREV.PFGSDGV
mrsbsu    FAHGWLLMKDG.KMSKSKGNVVDPVTLIERY..GLDEL.RYYLLREV.PFGSDGV
mrstma    FAHGWLTV.NGQKISKSLGNAIDPRFFVKRY..GNDVV.RYYLIRDI.MFGKDGD
mrsmtu    FAHGFLHNR.GEKMSKSVGNIVDPVALAEAL..GVDQV.RYFLLREV.PFGQDGS
mrstth    NVGGFLLGPDGRKMSKTLGNVVDPFALLEKY..GRDAL.RYYLLREI.PYGQDTP
mrshpy    CVHGWWTI.EGVKMSKSLGNVLDAQKIAMEY..GIEEL.RYFLLREV.PFGQDGD
mrsscm    VVHGHWLC.NGMKMSKSLGNVVDPIDMARYY..GADIV.RWFLLENS.KLEEDGD
mrsspm    LVHSHWTM.NKVKMSKSLGNVVDPFWLIEKY..GVDTI.RYYLLKRG.RLTSDSN
mrsscc    NTTEYLQYENG.KFSKSRGVGVFG.NNAQDSGISPS.VWRYYLA.SVRPESSDSH
mrsspc    NTTDYLNYETG.KFSKSRGVGVFG.NTAQDIGLSPS.VWRYYLL.SSRPETSDTM
mrsosa    SVTEYLNYEAG.KFSKSHGIGVFG.NDAKDTNIPPE.VWRYYLL.TNRPEVSDTL
mrshsa    IATEYLNYEDG.KFSKSRGVGVFG.DMAQDTGIPAD.IWRFYLL.YIRPEGQDSA
mrscel    CATEYLNYEDT.KFSKSRGTGIFG.DAAQGTEIPAD.IWRFYLL.YMRPESQDTA
mrsbbu    SSSEYLNYENL.KFSKSEGTGIFG.NDAITTGIPSD.IWRFYIY.YNRPEKSDFQ
mrseco    FVHGYVTV.NGAKMSKSRGTFIKA..STWLNHFDADSL.RYYYTAKLSSRIDDID
mrshin    FAHGYVTV.DGAKMSKSRGTFIQA..STYLNHIDPECL.RYYYAAKLNDRIEDLD
mrsmja    VSGGYLTL.EGRKMSTSKRWVVWV..KDFVKNFDADYL.RYYLIMS.APLFKDCD
mrsmth    IAGEYLSL.EGQKMSTSKNWVVWT..SDFLERFDRDLL.RYYLTVN.APLTRDTD
mrsafu    VASGMV.KVEGKTFSKSRGYVVWVEEDYLKSGLSPDYL.RYYIVNY.TSHQKDLN
```

Figure 2. Alignment of the fragment of *Oryza sativa* predicted amino-acid sequence (osa) with other known methionyl-tRNA synthetases (mrs).

M. genitalium (mge), *M. pneumoniae* (mpn), *A. thaliana* chloroplastic (atc), *Synechocystis* sp. (syn), *B. stearothermophilus* (bst), *B. subtilis* (bsu), *T. maritima* (tma), *M. tuberculosis* (mtu), *T. thermophilus* (tth), *H. pylori* (hpy), *S. cerevisiae* mitochondrial (scm), *S. pombe* mitochondrial (spm), *S. cerevisiae* cytoplasmic (scc), *S. pombe* cytoplasmic (spc), *H. sapiens* (hsa), *C. elegans* (cel), *B. burgdorferi* (bbu), *E. coli* (eco), *H. influenzae* (hin), *M. jannaschii* (mja), *M. thermoautotrophicum* (mth) and *A. fulgidus* (afu). The conserved „KMSKS"-like motif is marked in bold; the residues homologous to rice sequence are shaded; amino acids which are identical in all compared sequences are marked with asterisks.

tains a HLGN (not shown) and KFSKS (Fig. 2) motifs. They are counterparts of HIGH and KMSKS proposed earlier as a "signature sequences" for class I aaRSs, where methionyl-tRNA synthetase is a member.

The similarity searching of the EMBL/GenBank data bases [20] clearly identified methionyl-tRNA synthetases of yeast, human and *C. elegans* as significantly homologous to rice clone. The data revealed 61.4%, 59.9% and 59.2% of sequence similarity for yeast, human and *C. elegans* MetRS, respectively.

C. elegans and rice MetRSs sequences show the carboxy-terminal extensions. These two domains were compared and they show 50.6% identity in the 170 amino acid overlap. The same analysis was performed for two previously reported homologues of the C termini of bacterial and *C. elegans* MetRSs: yeast protein G4p1 (described also as Arc1p) and human endothelial-monocyte-activating polypeptide II (EMAPII) [13, 14, 40]. The identity with yeast protein is 52.1% for 167 amino acid overlap, and 52.2% with EMAPII – for the overlap of 186 amino acids.

The results of sequence analysis confirm that the obtained rice cDNA comprises the coding region for plant methionyl-tRNA synthetase.

We thank Dr. Yoshiaki Nagamura for providing the rice cDNA clone C2054 and Dr. Maciej Szymański for help with computational analysis.

REFERENCES

1. Eriani, G., Delarue, M., Poch, O., Gangloff, J. & Moras, D. (1990) Partition of aminoacyl-tRNA synthetases into two classes on the basis of two mutually exclusive sets of sequence motifs. *Nature* **347**, 203–206.

2. Cusack, S., Berthet-Colominos, C., Härtlein, M., Nassar, N. & Leberman, R. (1990) A second class of synthetase structure revealed by X-ray analysis of *E. coli* seryl-tRNA synthetase at 2.5 Å. *Nature* **347**, 249–255.

3. Webster, T.A., Tsai, H., Kula, M., Mackie, G. & Schimmel, P. (1984) Specific homology and three-dimensional structure of an aminoacyl-tRNA synthetase. *Science* **226**, 1315–1317.

4. Hountoudji, C., Dessen, P. & Blanquet, S. (1986) Sequence similarities among the family of aminoacyl-tRNA synthetases. *Biochimie* **68**, 1071–1078.

5. Rossmann, M.G., Moras, D. & Olsen, K.W. (1974) Chemical and biological evolution of a nucleotide binding domain. *Nature* **250**, 194–199.

6. Moras, D. (1992) Structural and functional relationships between tRNA synthetases. *Trends Biochem. Sci.* **17**, 159–164.

7. Hecht, S.M. (1979) 2'OH *vs* 3'OH specificity in tRNA aminoacylation; in *tRNA Structure, Properties and Recognition* (Schimmel, P., Söll, D. & Abelson, J.N., eds.) pp. 345–360, Cold Spring Harbor Laboratory, NY.

8. Schimmel, P. & Söll, D. (1979) Aminoacyl-tRNA synthetases: General features and tRNA recognition. *Annu. Rev. Biochem.* **48**, 601–648.

9. Koch, G.L.E. & Bruton, C.J. (1974) The subunit structure of methionyl-tRNA synthetase from *Escherichia coli*. *FEBS Lett.* **40**, 180–182.

10. Cassio, D. & Waller, J.P. (1971) Modification of *E. coli* methionyl-tRNA synthetase by proteolytic cleavage and properties of the trypsin modified enzyme. *Eur. J. Biochem.* **20**, 283–300.

11. Fasiolo, F., Gibson, B.W., Walter, P., Chatton, B., Biemann, K. & Boulanger, Y. (1985) Cytoplasmic methionyl-tRNA synthetase from bakers' yeast. A monomer with a post-translationally modified N terminus. *J. Biol. Chem.* **260**, 15571–15576.

12. Walter, P., Weygand-Durasevic, I., Sanni, A., Ebel, J.-P. & Fasiolo, F. (1989) Deletion analysis in the amino-terminal extension of methionyl-tRNA synthetase from *Saccharomyces cerevisiae* shows that a small region is important for the activity and stability of the enzyme. *J. Biol. Chem.* **264**, 17126–17130.

13. Simos, G., Segref, A., Fasiolo, F., Hellmuth, K., Shevchenko, A., Mann, M. & Hurt, E.C. (1996) The yeast protein Arc1p binds to tRNA and functions as a cofactor for the methionyl- and glutaminyl-tRNA synthetases. *EMBO J.* **15**, 5437–5448.

14. Frantz, D.J. & Gilbert, W. (1995) A novel yeast gene product, G4p1, with a specific affinity for quadruplex nucleic acids. *J. Biol. Chem.* **270**, 20692–20697.

15. Wilson, R., Ainscough, R., Anderson, K., Baynes, C., Berks, M., Bonfield, J., Burton, J., Connell, M., Copsey, T., Cooper, J., Coulson, A., Craxton, M., Dear, S., Du, Z., Durbin, R., Favello, A., Fulton, L., Gardner, A., Green, P., Hawkins, T., Hillier, L., Jier, M., Johnston, L., Jones, M., Kershaw, J., Kirsten, J., Laister, N., Latreille, P., Lightning, J., Lloyd, C., McMurray, A., Mortimore, B., O'Callaghan, M., Parsons, J., Percy, C., Rifken, L., Roopra, A., Saunders, D., Shownkeen, R., Smaldon, N., Smith, A., Sonnhammer, E., Staden, R., Sulston, J., Thierry-Mieg, J., Thomas, K., Vaudin, M., Vaughan, K., Waterston, R., Watson, A., Weinstock, L., Wilkinson-Sproat, J. & Wohldman, P. (1998) 2.2 Mb of contiguous nucleotide sequence from chromosome III of *C. elegans*. *Nature* (in press).

16. Kellerman, O., Brevet, A., Tonetti, H. & Waller, J.P. (1979) Macromolecular complexes of aminoacyl-tRNA synthetases from eukaryotes. 1. Extensive purification and characterization of the high-molecular-weight complex(es) of seven aminoacyl-tRNA synthetases from sheep liver. *Eur. J. Biochem.* **99**, 541–550.

17. Mirande, M., LeCorre, D. & Waller, J.P. (1985) A complex from cultured chinese hamster ovary cells containing nine aminoacyl-tRNA synthetases. *Eur. J. Biochem.* **147**, 281–289.

18. Godar, D.E., Garcia, V., Jacobo, A., Aebi, U. & Yang, D.C.H. (1988) Structural organization of the multienzyme complex of mammalian aminoacyl-tRNA synthetases. *Biochemistry* **27**, 6921–6928.

19. Berbec, H. (1990) High-molecular-weight complexes of aminoacyl-tRNA synthetases. *Postępy Biochemii* **3–4**, 41–52 (in Polish).

20. Altshul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410.

21. Sanger, F., Nicklen, S. & Coulson, A.R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463–5467.

22. Devereux, J., Haeberli, P. & Smithies, O. (1984) A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.* **12**, 387–395.

23. Dardel, F., Panvert, M. & Fayat, G. (1990) Transcription and regulation of expression of the *Escherichia coli* methionyl-tRNA synthetase gene. *Mol. Gen. Genet.* **223**, 121–133.

24. Mechulam, Y., Schmitt, E., Panvert, M., Schmitter, J.M., Lapadat-Topolsky, M., Meinnel, T., Dessen, P., Blanquet, S. & Fayat, G. (1991) Methionyl-tRNA synthetase from *Bacillus stearothermophilus*: Structural and functional identities with the *Escherichia coli* enzyme. *Nucleic Acids Res.* **19**, 3673–3681.

25. Ogasawara, N., Nakai, S. & Yoshikawa, H. (1994) Systematic sequencing of the 180 kilobase region of the *Bacillus subtilis* chromosome containing the replication origin. *DNA Res.* **1**, 1–14.

26. Nureki, O., Muramatsu, T., Suzuki, K., Kohda, D., Matsuzawa, H., Ohta, T., Miyazawa, T. & Yokoyama, S. (1991) Methionyl-tRNA synthetase gene from an extreme termophile, *Thermus thermopilus* HB8. *J. Biol. Chem.* **266**, 3268–3277.

27. 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., Fitz-Hugh, 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.

28. Tomb, J.-F., White, O., Kerlavage, A.R., Clayton, R.A., Sutton, G.G., Fleischmann, R.D., Ketchum, K.A., Klenk, H.P., Gill, S., Dougherty, B.A., Nelson, K., Quackenbush, J., Zhou, L., Kirkness, E.F., Peterson, S., Loftus, B., Richardson, D., Dodson, R., Khalak, H.G., Glodek, A., McKenney, K., Fitzegerald, L.M., Lee, N., Adams, M.D., Hickey, E.K., Berg, D.E., Gocayne, J.D., Utterback, T.R., Peterson, J.D., Kelley, J.M., Cotton, M.D., Weidman, J.M., Fujii, C., Bowman, C., Watthey, L., Wallin, E., Hayes, W.S., Borodovsky, M., Karp, P.D., Smith, H.O., Fraser, C.M. & Venter, J.C. (1997) The complete genome sequence of the gastric pathogen *Helicobacter pylori. Nature* **388**, 539–547.

29. Kaneko, T., Sato, S., Kotani, H., Tanaka, A., Asamizu, E., Nakamura, Y., Miyajima, N., Hirosawa, M., Sugiura, M., Sasamoto, S., Kimura, T., Hosouchi, T., Matsuno, A., Muraki, A., Nakazaki, N., Naruo, K., Okumura, S., Shimpo, S., Takeuchi, C., Wada, T., Watanabe, A., Yamada, M., Yasuda, M. & Tabata, S. (1996) Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. *DNA Res.* **3**, 109–136.

30. Himmelreich, R., Hilbert, H., Plagens, H., Pirkl, E., Li, B.C. & Herrmann, R. (1996) Complete sequence analysis of the genome of the bacterium *Mycoplasma pneumoniae. Nucleic Acids Res.* **24**, 4420–4449.

31. Philipp, W.J., Poulet, S., Eiglmeier, K., Pascopella, L., Balasubramanian, V., Heym, B., Bergh, S., Bloom, B.R., Jacobs, W.R., Jr. & Cole, S.T. (1996) An integrated map of the genome of the tubercle bacillus, *Mycobacterium tuberculosis* H37Rv, and comparison with *Mycobacterium leprae. Proc. Natl. Acad. Sci. U.S.A.* **93**, 3132–3137.

32. 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., Fritchman, J.L., Weidman, J.F., Small, K.V., Sandusky, M., Fuhrmann, J.L., Nguyen, D.T., Utterback, T.R., Saudek, D.M., Phillips, C.A., Merrick, J.M., Tomb, J.-F., Dougherty, B.A., Bott, K.F., Hu, P.-C., Lucier, T.S., Peterson, S.N., Smith, H.O., Hutchison, C.A., III & Venter, J.C. (1995) The minimal gene complement of *Mycoplasma genitalium. Science* **270**, 397–403.

33. Fraser, C.M., Casjens, S., Huang, W.M., Sutton, G.G., Clayton, R.A., Lathigra, R., White, O., Ketchum, K.A., Dodson, R., Hickey, E.K., Gwinn, M., Dougherty, B., Tomb, J.-F., Fleischmann, R.D., Richardson, D., Peterson, J., Kerlavage, A.R., Quackenbush, J., Salzberg, S., Hanson, M., van-Vugt, R., Palmer, N., Adams, M.D., Gocayne, J.D., Weidman, J., Utterback, T., Watthey, L., McDonald, L., Artiach, P., Bowman, C., Garland, S., Fujii, C., Cotton, M.D., Horst, K., Roberts, K., Hatch, B., Smith, H.O. & Venter, J.C. (1997) Genomic sequence of a Lyme disease spirochete, *Borrelia burgdorferi. Nature* **390**, 580–586.

34. 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., Adams, M.D., Reich, C.I., Overbeek, R., Kirkness, E.F., Weinstock, K.G., Merrick, J.M., Glodek, A., Scott, J.D., Geoghagen, N.S., Weidman, J.F., Fuhrmann, J.L., Nguyen, D.T., Utterback, T., Kelley, J.M., Peterson, J.D., Sadow, P.W., Hanna, M.C., Cotton, M.D., Hurst, M.A., Roberts, K.M., Kaine, B.B., Borodovsky, M., Klenk, H.P., Fraser, C.M., Smith, H.O., Woese, C.R. & Venter, J.C. (1996) Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. *Science* **273**, 1058–1073.

35. Smith, D.R., Doucette-Stamm, L.A., Deloughery, C., Lee, H.-M., Dubois, J., Aldredge, T., Bashirzadeh, R., Blakely, D., Cook, R., Gilbert, K., Harrison, D., Hoang, L., Keagle, P., Lumm, W., Pothier, B., Qiu, D., Spadafora, R., Vicare, R., Wang, Y., Wierzbowski, J., Gibson, R., Jiwani, N., Caruso, A., Bush, D., Safer, H., Patwell, D., Prabhakar, S., McDougall, S., Shimer, G., Goyal, A., Pietrovski, S., Church, G.M., Daniels, C.J., Mao, J-i., Rice, P., Nolling, J. & Reeve, J.N. (1997) Complete genome sequence of *Methanobacterium thermoautotrophicum* delta H: Functional analysis and comparative genomics. *J. Bacteriol.* **179**, 7135–7155.

36. Klenk, H.P., Clayton, R.A., Tomb, J., White, O., Nelson, K.E., Ketchum, K.A., Dodson, R.J., Gwinn, M., Hickey, E.K., Peterson, J.D., Richardson, D.L., Kerlavage, A.R., Graham, D.E., Kyrpides, N.C., Fleischmann, R.D., Quackenbush, J., Lee, N.H., Sutton, G.G., Gill, S., Kirkness, E.F., Dougherty, B.A., McKenney, K., Adams, M.D., Loftus, B., Peterson, S., Reich, C.I., McNeil, L.K., Badger, J.H., Glodek, A., Zhou, L., Overbeek, R., Gocayne, J.D., Weidman, J.F., McDonald, L., Utterback, T., Cotton, M.D., Spriggs, T., Artiach, P., Kaine, B.P., Sykes, S.M., Sadow, P.W., D'Andrea, K.P., Bowman, C., Fujii, C., Garland, S.A., Mason, T.M., Olsen, G.J., Fraser, C.M., Smith, H.O., Woese, C.R. & Venter, J.C. (1997) The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon *Archaeoglobus fulgidus*. *Nature* **390**, 364–370.

37. Walter, P., Gangloff, J., Bonnet, J., Boulanger, Y., Ebel, J.-P. & Fasiolo, F. (1983) Primary structure of the *Saccharomyces cerevisiae* gene for methionyl-tRNA synthetase. *Proc. Natl. Acad. Sci. U.S.A.* **80**, 2437–2441.

38. Lage, H. & Dietel, M. (1996) Cloning of human cDNA encoding a protein with high homology to yeast methionyl-tRNA synthetase. *Gene* **178**, 187–189.

39. Tzagoloff, A., Vambutas, A. & Akai, A. (1989) Characterization of MSM1, the structural gene for yeast mitochondrial methionyl-tRNA synthetase. *Eur. J. Biochem.* **179**, 365–371.

40. Kao, J., Houck, K., Fan, Y., Haehnel, I., Libutti, S.K., Kayton, M.L., Grikscheit, T., Chabot, J., Nowygrod, R., Greenberg, S., Kuang, W.-J., Leung, D., Hayward, J.R., Kisiel, W., Heath, M., Brett, J. & Stern, D.M. (1994) Characterization of a novel tumor-derived cytokine. Endothelial-monocyte activating polypeptide II. *J. Biol. Chem.* **269**, 25106–25119.