Volume 247, number 2, 167-172 FEB 07010 April 1989

An archaebacterial gene from *Methanococcus vannielii* encoding a protein homologous to the ribosomal protein L10 family

Andreas K.E. Köpke, Gottfried Baier* and Brigitte Wittmann-Liebold

Max-Planck-Institut für Molekulare Genetik, Abteilung Wittmann, Ihnestrasse 73, D-1000 Berlin 33, Germany and **Institut für Medizinische Mikrobiologie, Fritz-Pregl-Str. 3, A-6020 Innsbruck, Austria*

Received 21 February 1989

An open reading frame upstream of the *Methanococcus vannielii* L12 gene has been detected. The beginning of this open reading frame agrees with the N-terminal region of a protein (MvaL10) which has been isolated from the 50 S ribosomal subunit of *M. vannielii* and sequenced. The length of this gene is 1008 nucleotides, coding for 336 amino acids. Excellent sequence similarities were found to the L10-like ribosomal proteins from *Halobacterium halobium* and man. The N-terminal part of the MvaL10 protein shows significant sequence similarities to the *E. coli* L10 protein. MvaL10 is more than twice as long as *E. coli* L10 but is of length similar to those of the homologous halobacterial and human proteins. Interestingly, the C-terminal region of MvaL10 shows exceptionally high similarity to the C-terminal sequence of the MvaL12 protein. This is not the case for the *E. coli* proteins but was also observed for the human, *Halobacterium* and *Sulfolobus* proteins.

Ribosomal protein; Protein L10; Nucleotide sequence; Evolution

1. INTRODUCTION

Investigations of ribosomal proteins from archaebacteria have already led to interesting information being obtained about the evolution of these organisms [1-4]. Similarities of the archaebacterial proteins were found with respect to the eukaryotic and/or eubacterial counterparts. On the other hand, for some archaebacterial proteins no detectable sequence homology with any other organism has been reported thus far. The degree of similarity depends on the protein examined. For example, the L12 proteins from archaebacteria and eukaryotes are homologous to each other but cannot be unambiguously aligned to the eubacterial L12 proteins [5]. In contrast, the L23 proteins are similar in all kingdoms [6].

The evolutionary relationships of archaebacteria could be further investigated if the sequences of more ribosomal proteins from this kingdom could be compared to those of eubacteria and eukaryotes. Here, we report the nucleotide sequence of the gene for ribosomal protein MvaL10 from the archaebacterium *Methanococcus vannielii* and compare its deduced amino acid sequence to those of other organisms.

Ribosomal protein L10 is known to constitute a complex with protein L7/L12 in *E. coli* [7], and with the corresponding protein in *Sulfolobus* (Henning, P., this institute, unpublished) and yeast [8]. It is localized in the 50 S particle of *E. coli* near the stalk protuberance and protein L11 [9].

2. EXPERIMENTAL

2.1 *Materials*

The M13, ^{35}S, sequencing kit, endonucleases and T_4 polymerase were from Biolabs, Beverly (USA). [γ-^{32}P]ATP and [α-^{35}S]thio-dATP were supplied by Amersham (Bucks, England). Agarose (ultrapure) and DNA ligase were from BRL, Gaithersburg (USA). Some components of growth media were

Correspondence address: A.K.E. Köpke, Max-Planck-Institut für Molekulare Genetik, Abteilung Wittmann, Ihnestrasse 73, D-1000 Berlin 33, Germany

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/89/$3.50 © 1989 Federation of European Biochemical Societies

from Difco (Detroit, USA). All other chemicals were analytical grade (p.A.) from Merck (Darmstadt). Enzymes and the M13 sequencing kit were used under the conditions recommended by the supplier, unless otherwise specified. The Gen-Pak™ column for ion-exchange chromatography of DNA fragments was supplied by Waters (Milford, USA). The HPLC system contained two pumps (model 64, Knauer, Berlin) with analytical pump heads, a Milton Roy gradient program running on an Apple IIe computer, a dynamic mixing chamber (Knauer), a variable-wavelength UV detector (Knauer), a Kipp & Zonen BD41 recorder and an LKB (Uppsala) 7000 Ultropac fraction collector.

2.2. *Computer programs*

The hydrophobicity plots, secondary structure predictions and homology searches were performed on a VAX 8600 computer with programs from the Genetics Computer Group [10]. The homology search in the RIBO database (Köpke, A.K.E. and Wittmann-Liebold, B., unpublished) was performed with the ALIGN program [11].

2.3. *HPLC isolation of DNA fragments*

To separate DNA fragments obtained after digestion of plasmids or inserts with restriction enzymes, a HPLC ion-exchange column was used. The running buffer (buffer A) comprised 25 mM Tris-Cl at pH 7.5, with the eluent (buffer B) being 1 M NaCl in buffer A. A gradient formed as a root function (exponent 0.3) was applied to change the mixture from 45 to 65% buffer B in 45 min. The pooled peak fractions were precipitated with ethanol.

2.4. *Subcloning for DNA sequencing*

The plasmids pMvaX1 and L10e were digested by different restriction enzymes. The resulting fragments were isolated by agarose gel electrophoresis or ion-exchange HPLC and then subcloned in M13mp18 and M13mp19. Nucleotide sequencing was carried out according to the dideoxy chain-termination method [12] using [α-^{35}S]thio-dATP.

2.5. *Nomenclature*

Proteins of the large subunit of *M. vannielii* had originally been named ML1, ML2, etc. according to their position on two-dimensional electropherograms [13]. In those cases where unambiguous homology to the ribosomal proteins of *E. coli* was later found by sequence analysis, the proteins were renamed in order to express this homology. For examaple, protein ML2 thus became MvaL10 (this paper), with ML8 being renamed MvaL12 [5].

3. RESULTS AND DISCUSSION

3.1. *Amino acid sequence determination*

The N-terminal sequence of the HPLC-purified protein MvaL10 was determined in a liquid-phase sequencer according to Köpke and Wittmann-Liebold [6]. The amino acids were identified up to position 11 with a gap at position 6 (fig.1).

3.2. *Nucleotide sequence of the MvaL10 gene*

The 350 b *Kpn*I(polylinker)/*Pst*I fragment obtained from the 5′-end of pMvaX1 was subcloned in M13mp18 and M13mp19 and then sequenced in both directions. The 365 b *Pst*I/*Pst*I fragment from the middle of pMvaX1 was sequenced as described [5] and the overlap between these two parts was found on a 129 b *Dde*I/*Rsa*I fragment. The 535 b *Pst*I/*Eco*RI subfragment of the L10e plasmid was sequenced in both directions and yielded an 18-base overlap at its 3′-end with the 5′-end of pMvaX1. The approx. 650 b *Xba*I(polylinker)/*Pst*I subfragment of L10 e was ligated into *Xba*I/*Pst*I-digested M13mp18. 260 bases were sequenced and the beginning of the L10 gene was found after 124 bases. An overlap was sequenced on the 254 b *Nsi*I/*Dra*I fragment cloned into the *Pst*I/*Sma*I site of M13mp18.

3.3. *Comparison of the nucleotide and the N-terminal protein sequence*

The protein sequence deduced from the nucleotide sequence at the 5′-end of the gene was in agreement with the N-terminal amino acid sequence of the purified MvaL10 protein. This finding and the excellent homology to other L10 proteins showed that the protein sequence was correctly identified. The difficulties in identifying the amino acid at position 6 (fig.1) resulted from the destruction of serine residues during the sequencing and from the small amounts of protein available.

Protein ML9 (named according to the 2D gel pattern) was also purified and gave the same N-terminal amino acid sequence as MvaL10. Protein ML9 was present in all 50 S protein preparations. It has a length of about 100 amino acids and is probably an N-terminal fragment of MvaL10, since N-terminal sequence analysis indicated identity with MvaL10. In addition, ML9 was observed in peaks together with MvaL12 which might be due to the known ability of protein L10 and L12 to form a complex [7].

3.4. *Primary structure of protein MvaL10*

The protein composition of MvaL10 is: Ala_{44}, Cys_1, Asp_{18}, Glu_{35}, Phe_9, Gly_{20}, His_2, Ile_{27}, Lys_{36}, Leu_{29}, Met_8, Asn_{10}, Pro_{16}, Gln_7, Arg_5, Ser_{18}, Thr_{16}, Val_{30}, Trp_1, Tyr_4. The relative molecular mass of MvaL10 is 35956 Da. With 43 basic

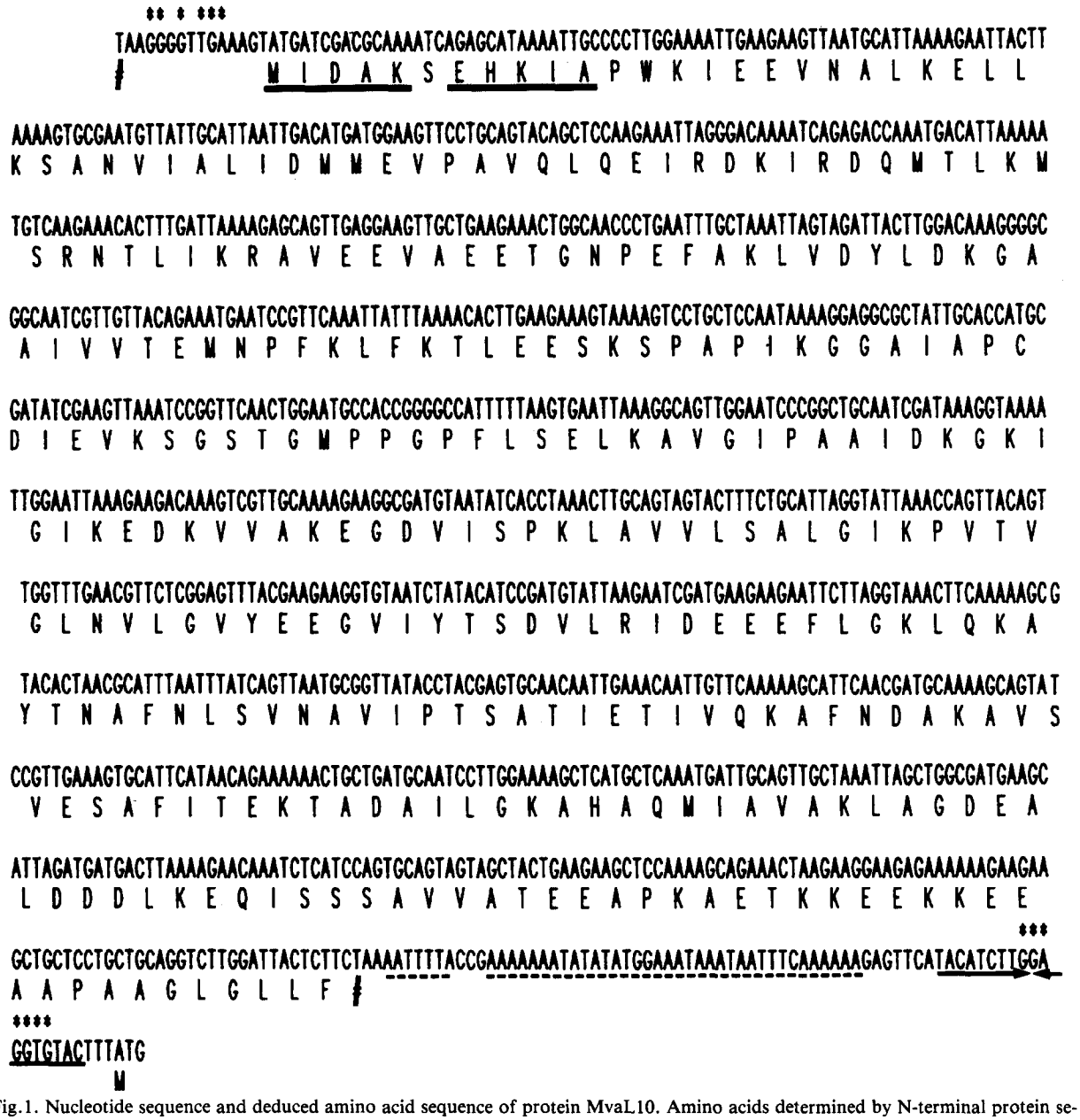

Fig.1. Nucleotide sequence and deduced amino acid sequence of protein MvaL10. Amino acids determined by N-terminal protein sequencing of the purified protein are underlined. Stop codons are denoted by (#). Shine-Dalgarno sequences complementary to the 3′-end of the 16 S rRNA are indicated by asterisks (*) above the DNA sequence. The AT-rich DNA sequence in the spacer region is underlined, and the putative stem-loop structure is shown with arrows below the DNA sequence.

residues (His, Lys, Arg) and 53 acidic residues (Glu, Asp) this protein is acidic, and has an unusually high content of alanine.

3.5 *Primary structure of the DNA*

A Shine-Dalgarno sequence (GGGGTTGA) was identified, starting 12 bases upstream of the structural gene. The L10 gene begins with ATG for methionine and ends with a TAA stop codon as found for all methanogenic ribosomal genes that we have sequenced. A spacer region of 60 nucleotides was identified between the stop codon

of the L10 gene and the Shine-Dalgarno sequence of the following L12 gene. In this spacer region, we located an AT-rich structure (AAAAAAATATAT ATAT) and a stem-loop structure (TACATCTTG GAGGTGTA).

3.6. *Homology between MvaL10 and ribosomal proteins from other organisms*

When protein MvaL10 was compared to all ribosomal protein sequences presently known (RIBO database) it was found to be most similar to the corresponding protein from *Halobacterium halobium*. A comparison score of 54 SD units for these two proteins was obtained by the program ALIGN (100 random runs, mutation data matrix, break penalty of 20). 122 identical residues and 57 conservative exchanges were aligned with only three gaps, namely from position 73 to 79, at position 286, and from position 324 to 342. Four stretches of more than three amino acids in a row were identical in the two proteins.

When the sequence of MvaL10 was compared to the corresponding protein from man, an ALIGN score of 25 SD units was obtained. 83 identical amino acids, 43 conservative exchanges and four gaps were found by this alignment. However, the alignment from position 261 to the C-terminus is not significant.

Comparison of MvaL10 with protein EcoL10 from *E. coli* gave significant sequence similarity for positions 64–73 of the MvaL10 protein with the corresponding region of EcoL10 (fig.2). For comparison between the first 75 residues of MvaL10 with the first 70 residues of EcoL10 an alignment by the program GAP yielded a score of more than 5 SD units (standard parametrs, 100 random runs).

A multiple sequence alignment was carried out for all proteins of the L10 family reported so far

```
          1                                                                                                  100
  MvaL10  MI..DAKSEH KIAPWKIEEV NALKELLKSA NVIALIDMME VPAVQLQEIR DKIRDQ.MTL KMSRNTLIKR AVEEVAEETG NPEFAKLVDY LDKGAAIVVT
  HhaL10  MSAEEQRTTE EVPEWKRQEV AELVDLLETY DSVGVVVNVTG IPSKQLQDMR RGLHGQ.AAL RMSRNTLLVR ALEE.....A GAGLDTLTEY VEGEVGLVAT
Con.Arch. M...*.**.. .*..WK..EV ..L.*LL.*. ..****.... *P..QLQ*.R ..*.Q...L *MSRNTL*.R A*EE.....* .....L.*Y **...**V.T

  ManP0   MPRED..... .RATWKSNYF LKIIQLLDDY PKCFIVGADN VGSKQMQQIR MSLRGKAVVL .MGKNTMMRK AIRGHLEN.. NPALEKLLPH IRGNVGFVFT
  EcoL10  ALNLQDKQ.. .......AIV AEVSEVAKGA LSAVVADSRG VTVDKMTELR KAGREAGVYM RVVRNTLLRR AVEG------ ---------- ----------
Con.  L10 m..ed.k... ..a.wk..ev aei.ellk.. .s..vvd..g vpskq.qeiR ..lrgq.v.l rmsrNTllrr Ave...e... np.l.kL..y ..g.vg.V.T
          ----------------------------------------------------------------------------------------------------

          101                                                                                                200
  MvaL10  EMNPFKLFKT LEESKSPAPI KGGAIAPCDI EVKSGSTGMP PGPFLSELKA VGIPAAIDKG KIGIKEDKVV AKEGDVISPK LAVVLSALGI KPVTVGLNVL
  HhaL10  NDNPFGLYQQ LENSKTPAPI NAGEVAPNDI VVPEGDTGIA PGPFVGELQT IGANARIQEG SIQVLDDSVV TEEGETVSDD VSNVLSELGI EPKEVGLDLR
Con.Arch. ..NPF.L*.. LE.SK*PAPI .*G.*AP.DI .V..G.TG.. PGPF*.EL.. *G..A.I..G .I.*.**.VV ..EG*.*S..

  ManP0   KEDLTEIRDM LLANKVPAAA RAGAIAPCEV TVPAQNTGLG PEK.TSFFQA LGITTKISRG TIEILSDVQL IKTGDKVGAS EATLLNMLNI SPFSFGLVIQ
Con.  L10 ..npf.I... Le.sK.PApi .aGaiAPcdi .Vp.g.TG.. Pgpf.selqa .Gi.a.I..G .I.il.D.vv .keGd.vs.. .a.vLs.Lgl .P..vGL...
          ----------------------------------------------------------------------------------------------------

          201                                                                                                300
  MvaL10  GVYEEGVIYT SDVLRIDEEE FLGKLQKAYT NAFNLSVNAV IPTSATIETI VQKAFNDAKA VSVESAFITE KTADAILGKA HAQMIAV.AK LAGDEALDDD
  HhaL10  GVFSEGVLFT TEELEIDVDE YRADIQSAAA SARNLSVNAA YPTERTAPDL IAKGRGEAKS LGLQASVESP DLADDLVSKA DAQVRALAAQ IDDEDALPEE
Con.Arch. GV*.EGV**T **.L.ID.*E *.*.*Q.A.. .A.NLSVNA. .PT..T...* *.K*..*AK. *.*.....*. ..AD.**.KA .AQ..A*.A. *..**AL.**

  ManP0   QVFDNGSIYN PEVLDITEET LHSRFLEGVR NVASVCLQIG YPTVASVPHS IINGYKRVLA LSVETDYTFP LAEKVKAFLA DPSAFVAAAP VAA.....AT
Con.  L10 gVf.eGviyt .evL.Ideee .....q.a.. na.nlsvna. yPT.at.p.. i.kg...aka Isve.....p ..ad....kA daq..a.aA. .a...al...
          ----------------------------------------------------------------------------------------------------

          301                                                  358
  MvaL10  L......... .........KE QISSSAVVAT EEAPKAETKK EEKKEEAAPA AGLGLLF.
  HhaL10  LQDVDAPAAP AGGEADTTAD EQSDETQASE ADDADDSDDD DDDDDGNAGA EGLGEMFG
Con.Arch. L......... .........* ..S..... .*........ **..**.A.A .GLG..*.

  ManP0   TAAPAAAAAP AKVEA..... ..........  .........K EES.EES.DE DMGFGLFD
Con.  L10 I....a.aap a..ea..... ..s.....  .........k ee..ee.a.a .gIg.IF.
```

Fig.2. Primary structure similarities between *Methanococcus vannielii* MvaL10 (this paper), *Halobacterium halobium* HhaL10 [1], human ManP0 [15] and the 65 N-terminal amino acids of *E. coli* EcoL10 [18] shown by multiple sequence alignment. The sequence similarities at the C-terminal part are rather poor; therefore the alignment shown is not the only one possible. The C-terminal region of EcoL10 is not given, since the sequence similarity is too low for unambiguous alignment. The consensus of the two archaebacterial sequences (Con.Arch.) denotes identical residues by capitals and conservative exchanges by an asterisk (*). Con. L10 is the consensus of all compared L10 sequences: amino acids identical in all sequences are indicated using a capital and those identical in at least two sequences are shown by lower-case letters.

```
MvaL12  72 AAPAEVKKEEKKEDTTAAAAAGLGALFM  99
           |  || |||||||| || ||||| ||
MvaL10 312 APKAETKKEEKKEE..AAPAAGLGLLF. 336
```

Fig.3. Sequence similarity of the C-terminal regions of proteins MvaL10 and MvaL12. Identical residues and conservative exchanges are connected by vertical bars. Numbers indicate the amino acid positions in the proteins.

(fig.2). This alignment is only unambiguous for part of the protein chain whereas other regions with more sequence variability were difficult to align. Therefore, the alignment shown in fig.2 is not the only possible one.

Interestingly, the C-terminal region of MvaL10 shows considerable homology to that of MvaL12 (>6 SD units for comparison of MvaL10 residues 312–336 with MvaL12 residues 72–99; see fig.3). This homology is also reflected at the nucleotide level. Similar results have also been obtained for other archaebacteria [1,14] and eukaryotes [15]. This type of sequence homology between the C-termini of proteins L10 and L12 of the same organism was found for neither L10 and L12 nor for any other ribosomal protein pairs in *E. coli* [16].

3.7. *Secondary structure prediction*

Secondary structure predictions according to Chou and Fasman [17] were performed for the L10 proteins with the computer program PEPTIDE-STRUCTURE (GCG software). In fig.4 the predicted secondary structure together with the hydrophobicity pattern and the surface probability are shown for protein MvaL10. When the secondary structures predictions of the L10 proteins of the various organisms were compared, two general features were observed: (i) The methanococcal, halobacterial and human L10 proteins contain a large number of β-turns, and high surface probabilities were obtained for the C-terminal region and that near the N-terminus as shown for MvaL10

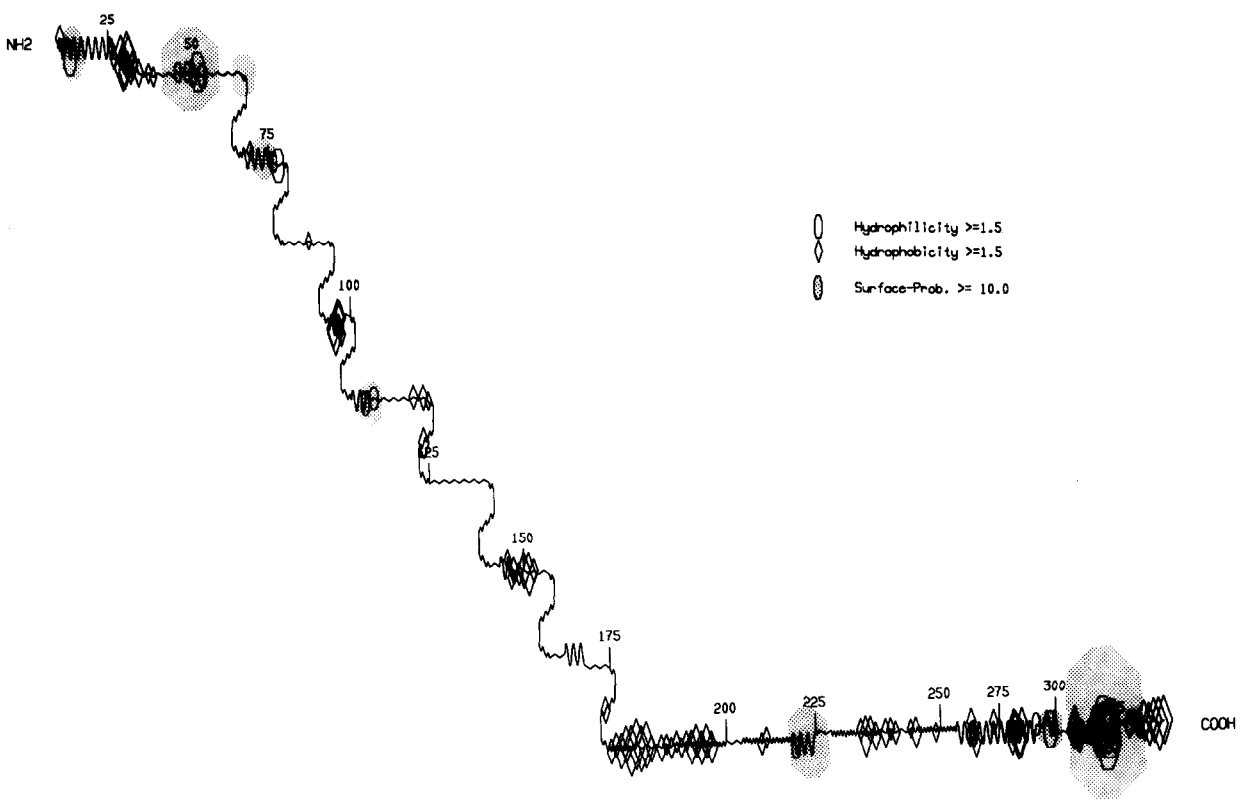

Fig.4. Secondary structure predictions of MvaL10 according to Chou and Fasman [17]; hydrophobic and hydrophilic amino acids are marked and regions of high surface probability [19] are indicated.

(fig.4). (ii) The predicted secondary structure of the *E. coli* protein L10 was different from that of the other proteins, since only two β-turns were observed in this case. However, also for this protein a high surface probability was predicted for regions near the N-terminal part of the protein chain.

3.8. *Conclusions about the evolution of the L10 proteins*

The ribosomal L10 proteins of various organisms can be divided in two major groups. One group consists of the archaebacterial and eukaryotic L10 proteins, and the other of the eubacterial L10 proteins. The two groups can be distinguished by means of three parameters: the length of the protein chains, their sequence similarity and the predicted secondary structure. A similar division into the two groups has already been observed for the L12 proteins [5].

In archaebacteria and eukaryotes the amino acid sequences of the C-terminal regions of proteins L10 and L12 are very similar to each other, although drastic differences in these regions are observed when proteins L10 or L12 from different organisms are compared. This points to the coevolution of both proteins: when the sequence of one protein changes, an alteration in the other protein is necessary to retain full activity.

Since proteins L10 and L12 form a complex in ribosomes from eubacteria, archaebacteria and eukaryotes, a correlation between coevolution and complex formation can be assumed. However, this can only be true for archaebacteria and eukaryotes, since no homology at the C-terminal region was found for proteins L10 and L12 in eubacteria.

Two reasons for the coevolution of the two proteins observed in archaebacteria and eukaryotes can be postulated: (i) a homologous primary structure is important for a given ribosomal function; (ii) for the formation of a complex between proteins L10 and L12 homology of the C-terminal regions is necessary in archaebacteria and eukaryotes but not in eubacteria.

With regard to the evolution of proteins L10 and L12, it appears that archaebacteria and eukaryotes are more closely related to each other than is the case between these two kingdoms on one hand and eubacteria on the other hand.

Acknowledgement: We thank Dr August Böck for a gift of *Methanococcus vannielii* cells, and Dr Roza Maria Kamp and Helga Gaenge for purifying ribosomal proteins by HPLC and performing 2D gel electrophoresis. The help of Wolfgang Krömer is gratefully acknowledged.

REFERENCES

[1] Itoh, T. (1988) Eur. J. Biochem. 176, 297–303.
[2] Auer, J., Lechner, K. and Böck, A. (1989) Can. J. Microbiol., in press.
[3] Kimura, M., Kimura, J. and Arndt, E. (1989) Can. J. Microbiol., in press.
[4] Köpke, A.K.E. and Wittmann-Liebold, B. (1989) Can. J. Microbiol., in press.
[5] Strobel, O., Köpke, A.K.E., Kamp, R.M., Böck, A. and Wittmann-Liebold, B. (1988) J. Biol. Chem. 263, 6538–6546.
[6] Köpke, A.K.E. and Wittmann-Liebold, B. (1988) FEBS Lett. 239, 313–318.
[7] Liljas, A., Kirseboom, L.A. and Leijonmarck, M. (1986) in: Structure, Function and Genetics of Ribosomes (Hardesty, B. and Kramer; G. eds) pp. 379–390, Springer, Heidelberg.
[8] Sáenz-Robles, M.T., Vilella, M.D., Pucciarelli, G., Polo, F., Remacha, M., Ortíz, B.L., Vidales, F.J. and Ballesta, P.G. (1988) Eur. J. Biochem. 177, 531–537.
[9] Walleczek, J., Schüler, D., Stöffler-Meilicke, M., Brimacombe, R. and Stöffler, G. (1988) EMBO J. 7, 3571–3576.
[10] Devereux, J., Haeberli, P. and Smithies, O. (1984) Nucleic Acids Res. 12, 387–395.
[11] Orcutt, B.C., Dayhoff, M.O. and Barker, W.C. (1982) NBR Report 820501-08710, Program Version 1.0, National Biomedical Research Foundation, Georgetown University Medical Center, Washington, DC.
[12] Sanger, F. and Coulson, A.R. (1978) FEBS Lett. 87, 107–110.
[13] Schmid, G. and Böck, A. (1981) Bacteriol. 147, 282–288.
[14] Ramirez, C., Shimin, R., Louie, K.A. and Matheson, A.T. (1989) Can. J. Microbiol., in press.
[15] Rich, B.E. and Steitz, J.A. (1987) Cell. Biol. 7, 4065–4074.
[16] Wittmann-Liebold, B. (1986) in: Structure, Function and Genetics of Ribosomes (Hardesty, B. and Kramer, G. eds) pp. 326–361, Springer, Heidelberg.
[17] Chou, P.Y. and Fasman, G.D. (1978) Adv. Enzymol. 47, 45–148.
[18] Heiland, I., Brauer, D. and Wittmann-Liebold, B. (1976) Biol. Chem. Hoppe-Seyler 357, 1751–1770.
[19] Emini, E.A., Hughes, J.V., Perlow, D.S. and Boger, J. (1985) J. Virol. 55, 836–839.