Volume 284, number 1, 51–56 FEBS 09800 June 1991

ADONIS 001457939100510Y

Extended N-terminal sequencing of proteins of the large ribosomal subunit from yeast mitochondria*

Lutz Grohmann**, Hanns-Rüdiger Graack***, Volker Kruft, Theodora Choli†, Sabine Goldschmidt-Reisin and Madoka Kitakawa††

Max-Planck-Institut für Molekulare Genetik, Abteilung Wittmann, Ihnestr. 73, D-1000 Berlin 33, Germany

Received 11 March 1991; revised version received 15 April 1991

We have determined the N-termini of 26 proteins of the large ribosomal subunit from yeast mitochondria by direct amino acid micro-sequencing. The N-terminal sequences of proteins YmL33 and YmL38 showed a significant similarity to eubacterial ribosomal (r-) proteins L30 and L14, respectively. In addition, several proteins could be assigned to their corresponding yeast nuclear genes. Based on a comparison of the protein sequences deduced from the corresponding DNA regions with the N-termini of the mature proteins, the putative leader peptides responsible for mitochondrial matrix-targeting were compiled. In most leader sequences a relative abundance of aromatic amino acids, preferentially phenylalanine, was found.

Mitoribosomal protein; Amino acid sequencing; Mitochondria; Yeast; *Saccharomyces cerevisiae*

1. INTRODUCTION

Ribosomes in mitochondria of *Saccharomyces cerevisiae* contain two rRNA molecules and about 80 proteins (for review see [1]). The two ribosomal RNAs and a single protein of the small ribosomal subunit are encoded in the mitochondrial genome. Like the majority of mitochondrial proteins, all remaining r-proteins are encoded in the nucleus, synthesised on cytoplasmic ribosomes and then transported into the mitochondria. In order to trace the evolutionary origin of the mitochondrial translation apparatus we have recently started a programme for the isolation and sequencing of mitochondrial r-proteins [2]. N-Terminal sequence data were used for cloning of the nuclear genes by oligonucleotide hybridisation [3,4]. In this way several nuclear genes coding for mitochondrial r-proteins have been cloned and analysed [3–7], and it was found that 3 of the sequences are significantly related to eubacterial r-proteins.

More generally, yeast nuclear genes coding for mitochondrial proteins were cloned by genetic complementation of pet mutants (for review see [8]). Pet mutant genes often affect gene products that are either directly involved in the oxidative metabolism of mitochondria or are necessary for expression of its activity [9]. For example, the PET-genes MRP1 and MRP2 were found to code for mitochondrial r-proteins [10]. It is to be expected that several more MRPs will be analysed, since most of the mitochondrial r-proteins are essential for the translational activity of mito-ribosomes [4–7,10,11]. Recently, McMullin et al. [11] identified PET123 as a component of the small ribosomal subunit in yeast. Nevertheless, in most cases it is difficult to define the exact function of the PET-genes, and additional biochemical analysis is necessary to assign them a function. Even with more data this function might remain unknown [12]. Therefore, N-terminal sequences of mitochondrial r-proteins could be a fruitful basis not only for cloning of the corresponding genes, but also for the identification and/or assignment of other PET-gene products in the future.

In the present study we show the results of extended N-terminal amino acid sequence analysis of 26 mitochondrial r-proteins. Proteins MRP7, YMR-26 and

*In commemoration of Prof. H.G. Wittman.

**Present address:* Institut für Genbiologische Forschung Berlin, Ihnestr. 63, D-1000 Berlin 33, Germany.

***Present address:* Ruhr-Universität Bochum, Fakultät für Chemie, Lehrstuhl für Biochemie, Universitätsstr. 150, D-4630 Bochum 1, Germany.

†*Present address:* Aristotelian University of Thessaloniki, School of Chemistry, Laboratory of Biochemistry, Thessaloniki 54006, Greece.

††*Present address:* Kobe University, Department of Biology, Faculty of Science, Rokkodai, Nada-ku, Kobe 657, Japan.

Abbreviations: PVDF, polyvinylidene difluoride; TP50, total proteins of the 50 S mitochondrial ribosomal subunit; PAGE, polyacrylamide gel electrophoresis; MRP, mitochondrial ribosomal protein gene; SDS, sodium dodecylsulphate

Correspondence address: L. Grohmann, Institut für Genbiologische Forschung, Ihnestr. 63, D-1000 Berlin 33, Germany.
Fax: (49) (30) 8300 0736.

YMR-44 [5,13,14] were assigned to the large subunit of yeast mitoribosomes. Furthermore, a compilation of leader peptide sequences of mitochondrial r-proteins was made, in order to characterise common features of leader sequences within the group of mitochondrial matrix-targeted proteins.

2. MATERIALS AND METHODS

The 50 S subunits of mito-ribosomes were prepared from *Saccharomyces cerevisiae* strain 07173 (α/a) as described [2]. Improved reverse-phase HPLC separation of TP50, two-dimensional PAGE and SDS-PAGE of HPLC fractions containing several proteins followed by blotting onto PVDF membranes were done according to Walsh et al. [15].

Amino acid sequencing of individual protein peaks and of blotted protein spots or bands which were excised from the membranes was performed essentially as described [16].

To digest proteins blotted onto PVDF membrane, the membrane was soaked in 0.2% polyvinyl pyrrolidone 40/methanol for 30 min and washed extensively with water and digesting buffer. Enzymatic digestion was performed with 0.02 U of endoproteinase LysC (Boehringer Mannheim, anal. grade)/0.1 M Tris buffer (pH 8.0) overnight at 37°C. The buffer containing the digested and released proteins was removed, the membrane was washed with 80% formic acid and water successively. All solutions were combined and subjected to reverse-phase HPLC analysis.

3. RESULTS AND DISCUSSION

The proteins isolated from the large subunit of mito-ribosomes from *S. cerevisiae* wild-type strain 07173 were subjected to reverse-phase HPLC. By improving our separation conditions we could purify additional proteins to homogeneity in a form suitable for direct micro-sequencing. However, in cases of low resolution where several proteins of different molecular masses were present in the HPLC fractions they were separated by SDS-PAGE. After blotting on PVDF-membranes, single protein bands were excised from the membranes and subjected to N-terminal amino acid sequencing. During the extensive SDS-PAGE analysis of different HPLC-separated proteins combined with direct sequencing it became clear that the actual number of r-proteins in the large subunit of yeast mito-ribosomes is higher than has been reported [2]. We completed our previous two-dimensional map of the 50 S subunit and found that it contains at least 46 proteins (Fig. 1). Individual proteins which eluted from the HPLC column at different positions, and which were not identical in their N-terminal sequences, were found to be localised in one single spot (see Fig. 1B and Table I). Conversely, for spots 15/16 and 28 [2] an endopeptidase digestion of the individual protein moiety followed by reverse-phase HPLC separation of the peptides revealed almost identical peptide compositions (data not shown). These proteins possibly undergo modifications such as methylation or acetylation as is the case with several bacterial r-proteins [17]. Alternatively, this result could reflect minimal amino acid sequence differences between the

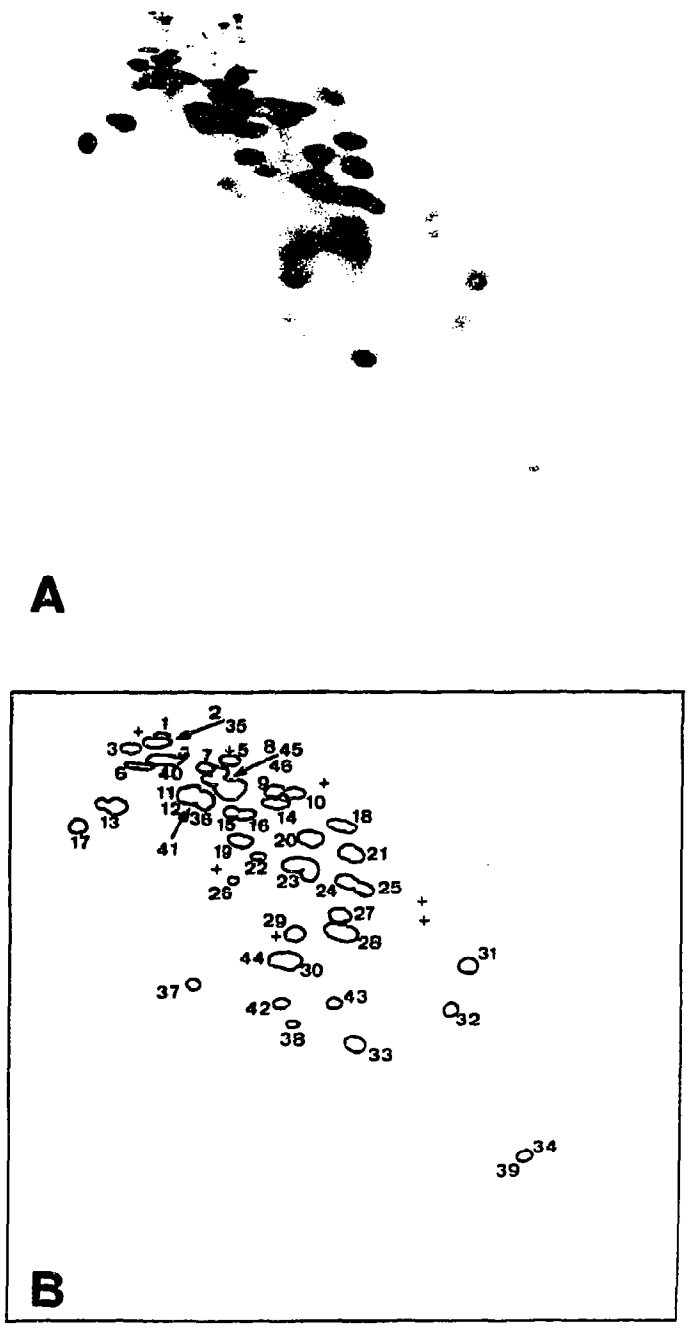

Fig. 1. (A) Two-dimensional electropherogram of the proteins of the large subunit of mito-ribosomes from *Saccharomyces cerevisiae* stained with Coomassie brillant blue. (B) Schematic drawing of the protein pattern of Fig. 1A. The numbering extends the previously reported two-dimensional map [2]. Additional proteins are numbered from spot 35 on. Spots containing more than one protein have been designated with all of their respective protein numbers.

parental haploid strains, since the proteins analysed were prepared from a diploid strain. Furthermore, the actual number of mitochondrial r-proteins increased, due to additional sequenced proteins which were clearly visible in silver-stained SDS-gels. These proteins were only faintly detectable in two-dimensional gels stained

Table I

Protein spots from the two-dimensional electropherogram shown in Fig. 1 which contain more than one protein

Number of spot in Fig. 1B	Proteins found in that position
2	YmL2, YmL35
4	YmL4, YmL40
8	YmL8, YmL45, YmL46
11/12	YmL11, YmL12, YmL36, YmL41
15/16	YmL15*
23	YmL23, ?
24	YmL24, YmL25
28	YmL28*
30	YmL30, YmL44
34	YmL34, YmL39

Asterisks indicate spots which appear in a form suggesting the existence of more than one protein, but were found to consist of only one protein.

with Coomassie Brillant Blue, and were therefore not counted previously (see spots 37 and 38 in Fig. 1B).

Table II shows the results of the extended sequencing of yeast mitochondrial r-proteins. In addition to the sequences which have been published [2,5] the N-termini of a further 26 proteins were determined. Two proteins (YmL19 and YmL23) were amino-terminally blocked. In general, smaller proteins with molecular masses between 10 and 20 kDa (proteins YmL20 to YmL39) gave longer N-terminal sequences. With the exception of proteins YmL18, YmL38 and YmL44 all other sequences do not start with a methionine, indicating that this residue is post-translationally cleaved off. This is also found for most of the eukaryotic cytoplasmic proteins, but for mitochondrial proteins it is probably due to their cleavable leader sequences (see below). For one of the mitochondrial r-proteins (YmL11) two different and partially overlapping N-terminal sequences were obtained (see Table II), suggesting two alternative cleavage sites or partial cleavage. Because the proteins were isolated from a diploid strain, the two alleles coding for YmL11 may differ in this specific protein region.

For several proteins (YmL9, YmL18, YmL20, YmL24, YmL27, YmL40, YmL44) we have determined internal amino acid sequences. This was done either to improve the deduced primary sequences of cloned genes [7] or to design oligonucleotides as primers for partial cloning of the corresponding nuclear gene by the polymerase chain reaction (Kang et al., to be published).

In those cases, where deduced protein sequences from the corresponding DNA sequences were available, these were compared with the N-terminal sequence data and some minor deviations were observed. As long as they did not affect the rest of the reading frame or the characteristics of the deduced protein sequence, we decided to show here only the data obtained from direct protein sequencing.

Although we have sequenced only parts of the entire proteins and therefore could not expect a high degree of similarity, we compared these sequences with those of all known ribosomal protein sequences. In two cases we found a significant sequence similarity (Fig. 2). The N-terminal sequence of protein YmL33 (11 kDa) has about 35% identity to proteins EcoL30 and BstL30 from *E. coli* and *B. stearothermophilus*, respectively. After cloning of the corresponding nucelar gene for protein YmL33 it was found that indeed two-thirds of the deduced protein sequence are highly conserved (W. Kang, to be published). YmL33 is one-third larger than the *E. coli* EcoL30, and its C-terminal extension showed sequence similarity to yeast cytoplasmic r-protein L16. For protein YmL38, for which only 26 residues were determined, a significant similarity was found to *B. stearothermophilus* BstL14 and to the chloroplast encoded protein ZmaL14 of maize. In contrast to YmL33, in this case the homologous eubacterial protein BstL14 is larger than the mitochondrial counterpart. With respect to the intracellular localisation of the genes it is interesting to note that the yeast L14-homologous protein YmL38 is nuclear-encoded, whereas in *Tetrahymena* an open reading frame significantly related to *E. coli* L14 was found within the mitochondrial genome [18].

Besides the similarities to eubacterial r-proteins, the comparison of the N-terminal sequence data gave complete correspondence to three known mitochondrial r-protein sequences. As has already been reported [3,4], protein YmL2 corresponds to the product of the MRP7 gene, which has been identified and cloned by screening of an expression library with antibodies raised against yeast mitochondrial r-proteins [13]. Recently the MRP genes YMR-44 [4] and YMR-26 [14] have been cloned, but the protein products were not assigned to one of the ribosomal subunits. We found that the N-termini of proteins YMR-44 and YMR-26 are identical to the sequences of YmL44 and YmL25, respectively. It thus became clear that both proteins (YMR-26 and YMR-44) are components of the large subunit of yeast mitoribosomes. Finally, since we sequenced almost all proteins of the large subunit in the size range of MRP2 (14 kDa, EcoS14 homologue [10]), and its N-terminus could not be found among the proteins sequenced so far, we assumed that MRP2 is a small subunit protein. This is underlined by the fact that in all cases where a similarity to *E. coli* large subunit proteins was observed, the corresponding mitochondrial r-proteins (YmL2, YmL8, YmL9, YmL33; [19]) are also components of the large ribosomal subunit.

In Table III a compilation of putative leader sequences of yeast mitochondrial r-proteins is given. These sequences are thought to function as signals for mitochondrial targeting and are cleaved off the protein after or during transport. Most of them revealed the known characteristic properties, namely a positive net

Table II

N-Terminal amino acid sequences of proteins of the large subunit of mito-ribosomes from *Saccharomyces cerevisae*

Protein (gene)	kDa	Sequence	Ref.
YmL1	35	SVTP	this work
YmL2 (MRP7)	40	SGK?AAGSR?SMKD?AGR?LG?K	this work [13]
YmL3	36	E?KRFL?ESELAKYKEYYQGLKSTVNEI	[2]
YmL4 (MRP-L4)	35	ARTKFTKPKPKQPVLPKDKIRP?TQLT	[2,7]
YmL6	32	?TLSPLPNIAIP?KYALVTVRSF??LP	this work
YmL7	28	KSANSLVKPVHHLWKIDK??L?PRF	this work
YmL8 (MRP-L8)	28.5	GVGIARKLSRDKAHRDALLKHLA?QLF	[4]
YmL9 (MRP-L9)	27.5	SVTRPFLVAPSIANSITTEAPAINHSPELANARK GTFGVMK (int. seq) NSFVK (int. seq)	[2,7]
YmL10	28.5	VSILGQLKPSKG?G	this work
YmL11	25	ALAHEQP?RKPVKPLD?RKVFLIDL EQPSRKPVKPLDKRKVFLID??K?NM[a]	this work
YmL12	32	W?DGYFVEVIVRFNP	this work
YmL13 (MRP-L13)	21	SSSNNKNRLELIPENFIGEGSRR?K?QKELKLAV	this work
YmL14	31	?Q??LI?T?KIAKQPNYQVG?AKPL	this work
YmL15	27	VIYLHKGPRINGL?RDPESYLRNP?GVLFS?VNAK??Q	this work
YmL18	23.5	MDIL?QL?P NFYDENAK (int. seq) ?FDH?M (int. seq)	this work
YmL19	16	blocked	
YmL20 (MRP-L20)	18	KQFGFPKTQVTTIYNKTKSASNYK TYM?GPQE?DEIR (int. seq)	[4]
YmL23	15	blocked	
YmL24 (MRP-L24)	17	RQ?RLIETRKIAKQPNYQVG?AKP ?P?Y?VGDAKPLHMP (int. seq) GLYGGSFVQFGNNISE (int. seq)	this work*
YmL25 (YMR-26)	18	SYIL?FLQL?D	this work [14]
YmL27 (MRP-L27)	15	LTRPWKKYRDGELFYGLSKVGNKRVPLTTKQ?NKTMYKG?N VRDGELFYGLS (int. seq) YRTYVLPDM?NFELKP (int. seq)	this work [7]
YmL28	14.5	?IFKSKVILVILAQRVV?QL?VK?ASP	this work
YmL30	9.5	SAPPKIKVGVLL??IPI??SE	this work
YmL31 (MRP-L31)	14.5	GGLLWKIPWRMSTHQKTRQRERLRNVDQVIKQLTLG	[2,3]
YmL32	11	AVPKKKVSHQKKRQ?LYGPGKKQLQMI?HLN?GPD	this work
YmL33 (MRP-L33)	11	VFYKVTLSRSLIGVPHTTKSIVK?LGLGKRGSIVYKKVNRAP??P	this work
YmL34	12.5	ANANRV??GDIDIAIVVR?KQRNM???D	this work
YmL35	42	GNATSERLPTDPVYP?V	this work
YmL36 (MRP-L36)	8.5	HYPGSTRITLPRRPAKKIQLGKSRPAIY?QFNVKM?L	this work*
YmL37	13.5	SVKSSQPAGTLLNLNI?KSGKDAVALEDKRYPN	this work
YmL38 (MRP-L38)	?	MIFLKSVIKVIDNSGAQLAF?IKVIR?G	this work*
YmL39	11	VKVKSKNSVIKLLSTAASGYSRYISIKRGAPLVTQVKY?PVVK??RLFK	this work
YmL40	36	TRTVAVRDVSFNG (int. seq) QF?LEGQTSHMITIPV?ILGK (int. seq)	this work
YmL41	32	?KSGSDQEV?INV	this work
YmL44 (YMR44)	13	MITKYF VARLVL (int. seq)	this work [5]

[a] Protein YmL11 gave two different but partially overlapping N-terminal sequences (see text). For proteins YmL4, YmL9, YmL13, YmL27 and YmL33 the corresponding genes have been cloned. Asteriks indicate proteins where genomic DNA fragments were successfully amplified by PCR and cloning is in progress. int. seq = N-terminal sequences determined from Lys-C endoproteinase peptides representing internal sequences of the corresponding protein. It follows that the preceding amino acid must be lysine.

charge, a high content of hydroxylated amino acids and hydrophobic residues, which are important for the amphipathic nature of leader sequences [20]. As can be seen in Table III, the length of the leader sequences varies from only 8 amino acids for YMR 31 to 33 residues for MRPS28p. The relative abundance of aromatic amino acids within these sequences should be noted, preferentially of phenylalanine. A basic amino acid can be observed adjacent to this phenylalanine or separated by one hydroxylated residue (serine). This is clearly neighboured by another aromatic amino acid, or alternatively such an amino acid is found 3–4 (one putative α-helical turn) or 7–8 residues (two putative α-helical turns) apart. Moreover in three of the proteins lacking a cleavable leader sequence (YmL25, YmL33 and YmL44) this motif is found at the N-terminus of

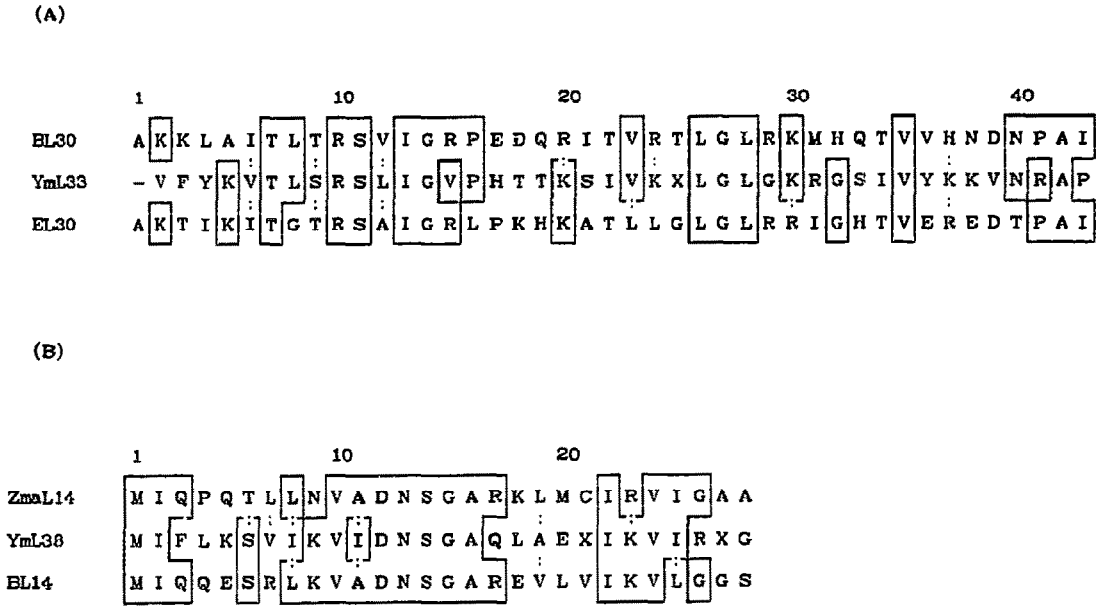

Fig. 2. Sequence similarities between N-terminal amino acid sequences. Identical amino acid residues are boxed and conservative exchanges are marked with colons. (A) Alignment of the YmL33 sequence with the N-terminus of ribosomal protein EcoL30 [24] and BstL30 [25] from *E. coli* and *B. stearothermophilus*, respectively. (B) Alignment of the YmL38 sequence with the N-terminus of ribosomal proteins EL14 [26] and ZmaL14 [27] from *B. stearothermophilus* and from *Zea mays* chloroplasts, respectively.

the sequence. Our observations are consistent with the results of Lemire et al. [21], where artificial leader sequences active in matrix-targeting also mostly contained a phenylalanine adjacent to a basic residue with another aromatic amino acid in a helical proximity. Furthermore, phenylalanine is part of a motif described by Hendrick et al. [22] which is possibly important for matrix processing protease activity.

Nevertheless, experiments to investigate the role of the aromatic amino acids within leader sequences are required. In this connection, another open question has been raised by recent reports concerning the receptors

Table III

Leader sequences of mitochondrial r-proteins as deduced by comparison of the gene sequence with the corresponding N-terminal amino acid sequence of the mature protein

Protein	Leader sequence	N-terminus of mature protein	Ref.
MRP2	MGNFRFPIKTKLPPG	FINARILRDN	[10]
MRPS28	MSIVGRNAILNLRISLCPLFMGKRSFVSSPVSN	SAKAVKFLKA	[6]
YmL2 (MRP7)	MWNPILLDTSSFSFQKHVSGVFLQVRN	ATKRAAGSRTSM	[13]
YmL4	MWKRSFHSQGGPLR	ARTKFTKPKP	[7]
YmL8	M	TVGIARKLSR	[4]
YmL9	MSKFLQGSIFSISKLHVRY	SSTRPFLVAP	[7]
YmL20	MIGRGVCCRSFHTAGSAW	KQFGFPKTQV	[4]
YmL25 (YMR-26)	M	SYKQYFDSLP	[14]
YmL27	MKGSPISQFSKTSINA	LTRPWKKYRD	[7]
YmL31	MFGPFKLTSPVA	GGLLWKIPWR	[3]
YmL33	M	VFYKVTLSRS	(W. Kang et al., unpublished)
YmL44 (YMR-44)		MITKYFSKVI	[5]
YMR-31	MIATPIRL	AKSAYEPMIK	[5]

Vertical bars indicate the putative cleavage site. Phenylalanine and other aromatic residues are in bold letters, the surrounding conserved amino acid residues are shaded (see also text for explanation). Above the sequence, the basic, hydrophobic and hydroxylated amino acids are indicated (+, *, and ·, respectively). The leader sequence of MRP2 was deduced by the comparison with its homologous *E. coli* and chloroplast r-protein.

for mitochondrial proteins [23]. Because mitochondrial r-proteins are a relatively large group consisting of about 80–90 members, it will also be interesting to investigate their possible common receptor-pathway directing them into mitochondria.

Acknowledgements: We are grateful to the late Prof. H.-G. Wittmann and to Prof. B. Wittmann-Liebold for their interest and support. We thank W. Kang for providing unpublished data and Dr R. Brimacombe for critical reading of the manuscript.

REFERENCES

[1] Curgy, J.J. (1985) Biol. Cell. 54, 1–38.
[2] Graack, H.-R., Grohmann, L. and Choli, T. (1988) FEBS Lett. 242, 4–8.
[3] Grohmann, L., Graack, H.-R. and Kitakawa, M. (1989) Eur. J. Biochem. 183, 155–160.
[4] Kitakawa, M., Grohmann, L., Graack, H.-R. and Isono, K. (1990) Nucleic Acids Res. 18, 1521–1529.
[5] Matsushita, Y., Kitakawa, M. and Isono, K. (1989) Mol. Gen. Genet. 219, 119–124.
[6] Dang, H. and Ellis, S.R. (1990) Nucleic Acids Res. 18, 6895–6901.
[7] Graack, H.-R. (1990) Ph.D. Thesis, Freie Universität Berlin, Germany.
[8] Tzagoloff, A. and Dieckmann, C.L. (1990) Microbiol. Rev. 54, 211–225.
[9] Tzagoloff, A. and Myers, A.M. (1986) Annu. Rev. Biochem. 55, 249–285.
[10] Myers, A.M., Crivellone, M.D. and Tzagoloff, A. (1987) J. Biol. Chem. 262, 3388–3397.
[11] McMullin, T.W., Haffter, P. and Fox, T.D. (1990) Mol. Cell. Biol. 10, 4590–4595.
[12] Michaelis, U. and Rödel, G. (1990) Mol. Gen. Genet. 223, 394–400.
[13] Fearon, K. and Mason, T.L. (1988) Mol. Cell. Biol. 8, 3636–3646.
[14] Kang, W., Matsushita, Y. and Isono, K. (1991) Mol. Gen. Genet. (in press).
[15] Walsh, M.J., McDougall, J. and Wittmann-Liebold, B. (1988) Biochemistry 27, 6867–6876.
[16] Choli, T., Kapp, U. and Wittmann-Liebold, B. (1989) J. Chromatogr. 476, 59–72.
[17] Wittmann, H.G., Littlechild, J.A. and Wittmann-Liebold, B. (1980) in: Ribosomes: Structure, Function and Genetics (Chambliss, G., Craven, G.R., Davis, J., Davis, K., Kahan, L. and Nomura, M. eds) pp. 51–88, University Park Press, Baltimore.
[18] Suyama, Y. and Jenny, F. (1989) Nucleic Acids Res. 17, 803.
[19] Wittmann-Liebold, B., Köpke, A.K.E., Krömer, E., Hatekeyama, T. and Wittmann, H.G. (1990) in: The Ribosome: Structure, Function and Evolution (Hill, W.E., Dahlberg, A., Garrett, R.A., Moore, P.B., Schlessinger, D. and Warner, J.R. eds) pp. 598–616, Am. Soc. Microbiol., Washington, DC.
[20] Nicholson, D.W. and Neupert, W. (1988) in: Protein Transfer and Organelle Biogenesis (Das, R.C. and Robbins, P.W. eds) pp. 677–746, Academic Press, New York.
[21] Lemire, B.D., Frankhauser, C., Baker, A. and Schatz, G. (1989) J. Biol. Chem. 264, 20206–20215.
[22] Hendrick, J.P., Hodges, P.E. and Rosenberg, L.E. (1989) Proc. Natl. Acad. Sci. USA 86, 4056–4060.
[23] Kiebler, M., Pfaller, R., Söllner, T., Griffiths, G., Horstmann, H., Pfanner, N. and Neupert, W. (1990) Nature 348, 610–616.
[24] Ritter, E. and Wittmann-Liebold, B. (1975) FEBS Lett. 60, 153–155.
[25] Kimura, M. (1984) J. Biol. Chem. 259, 1051–1055.
[26] Kimura, M., Kimura, J. and Ashman, K. (1985) Eur. J. Biochem. 150, 491–497.
[27] Markmann-Mulisch, U. and Subramanian, A.R. (1988) Eur. J. Biochem. 170, 507–514.