Volume 239, number 2, 313–318 FEB 06486 November 1988

Sequence of the gene for ribosomal protein L23 from the archaebacterium *Methanococcus vannielii*

Andreas K.E. Köpke and Brigitte Wittmann-Liebold

Max-Planck-Institut für Molekulare Genetik, Abteilung Wittmann, Berlin-Dahlem, Germany

Received 29 July 1988

The N-terminal sequence of HPLC-purified protein L23 from the *Methanococcus vannielii* ribosome has been determined by automated liquid-phase Edman degradation. Using the N-terminal amino acid sequence, an oligonucleotide probe complementary to the 5'-end of the gene was synthesized. The 26-mer oligonucleotide, containing two inosines, was used for hybridization with digested *M. vannielii* chromosomal DNA. The hybridizing band from *Hpa*II-digested genomic DNA was ligated into pUC18 to yield plasmid pMvaZ1 containing the entire gene of protein L23. The nucleotide sequence complemented the partial amino acid sequence, and the gene codes for a protein of 9824 Da. The amino acid sequence of protein L23 form *M. vannielii* was compared to that of ribosomal proteins from other archaebacteria as well as from eubacteria and eukaryotes. The number of identical amino acids is highest when the *M. vannielii* protein is compared to the homologous protein from yeast and lowest vs that from tobacco chloroplasts. Interestingly, the secondary structures of the proteins as predicted by computer programs are more conserved than the primary structures.

Ribosomal protein; Protein L23; Nucleotide sequence; Evolution; Oligonucleotide hybridization; (*M. Vannielii*)

1. INTRODUCTION

The investigation and comparison of ribosomal components from eubacterial, eukaryotic and archaebacterial sources have given interesting hints as to the evolution of these organisms (review [1]). Furthermore, it is of interest to determine which parts of the ribosomal RNAs and proteins have been highly conserved during evolution, since it is likely that these regions are important for the structure and/or function of the ribosome.

It has been shown that eubacteria, eukaryotes and archaebacteria have high sequence similarities among the homologous ribosomal proteins within each kingdom [2]. However, it has been more difficult to correlate sequences derived from different kingdoms. In general, more sequence similarities are found between archaebacteria and eukaryotes than between eubacteria and eukaryotes [3]. In addition, archaebacterial ribosomal sequences have been shown to vary in their degree of similarity to eubacterial and eukaryotic proteins. Also, some proteins show no similarity with either eukaryotic or eubacterial ribosomal proteins, which may be due to the incomplete protein sequence data available for eukaryotes or the presence of additional proteins in archaebacterial ribosomes.

Schmid and Böck [4] have described the ribosomal protein pattern of *Methanococcus vannielii* in two-dimensional polyacrylamide gels. The L23 protein from this organism (originally named L7 according to the gel electrophoresis pattern) was isolated by reverse-phase HPLC, and the 19 N-terminal amino acids were determined. Using this protein sequence we have designed a 26-mer oligonucleotide containing two inosine nucleotides and allowing GT pairing. Here, the complete nucleotide sequence determination of the *M. vannielli* L23 protein gene, detected and isolated using

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

The nucleotide sequence presented here has been submitted to the EMBL/GenBank database under the accession no. Y00772

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/88/$3.50

oligonucleotide hybridization, is reported. In addition, this archaebacterial protein is compared with its homologous sequences from organisms of all three kingdoms.

2. EXPERIMENTAL

2.1. *Materials*

Polynucleotide kinase was purchased from Boehringer (Mannheim). The M13-^{35}S sequencing kit and all endonucleases were from Biolabs (Beverly, USA); [γ-^{32}P]ATP and [α-^{35}S]thio-dATP were supplied by Amersham (Bucks, England). Gene-Screen membrane was from NEN (Boston, USA). Agarose (ultrapure), DNA ligase and *E. coli* DH5α strain were from BRL (Gaithersburg, USA). Some components of the growth media were from Difco (Detroit, USA). All other chemicals were pro-analysis grade from Merck. All enzymes and the M13 sequencing kit were used under the conditions recommended by the supplier, unless otherwise specified.

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 [5]. The multiple-prediction program was made by M. Dzionara and A. Beck from this institute (unpublished).

2.3. *Protein isolation and sequencing*

M. vannielii cells (DSM 1224) were prepared as described by Schmid and Böck [4] and kindly provided by Dr A. Böck. Ribosomes were prepared as described by Wittmann [6]. The total protein mixture of the 50 S subunits was chromatographed on a reversed-phase Vydac C18 column (10 μm particles, 30 nm pore size) [7,8]. The identity and purity of the protein were checked by two-dimensional polyacrylamide gel electrophoresis [9]. The N-terminal region was sequenced in a Berlin liquid-phase sequenator [10] employing on-line detection of the PTH amino acid derivatives by isocratic HPLC [11].

2.4. *DNA – preparation and blotting*

M. vannielii chromosomal DNA was prepared using the sarcosyl method of Hofman et al. [12]. Total genomic digests were performed overnight. Gel electrophoresis of the digested chromosomal DNA (10 μg/lane) was carried out in 150 × 150 × 7 mm 0.8% agarose gels at 1.0 V/cm in TAE buffer [13] for 16 h. The DNA was transferred onto Gene-Screen membranes, denatured and cross-linked to the nylon membrane by UV illumination for 10 min at 302 nm, as described by Church and Gilbert [14].

2.5. *Preparation of the oligonucleotide probe and hybridization*

The oligonucleotide mixture was synthesized automatically, purified and labeled as described [3]. Prehybridization was carried out for 1–4 h at 42°C in polypropylene bags with 10 ml of 10 × Denhardt's solution, 6 × SSC [13], 0.1% SDS and 0.1% sodium pyrophosphate, filtered through a 0.2 μm filter. Hybridization was performed overnight in 10 ml of 1 × Denhardt's solution, 6 × SSC, 0.1% SDS and 0.1% sodium pyrophosphate. The labeled oligonucleotide mixture (1 × 10^7 cpm) was used for hybridization of the nylon-bound DNA at 40°C overnight. Washing after hybridization was done in 6 × SSC and 0.1% SDS three times at room temperature for 1, 2 and 3 min, respectively, each time with 300 ml of the solution. The last stringent wash was carried out in 1 × SSC, 0.1% SDS at 45°C for 1 min. Autoradiographs of the hybridized membranes were obtained after exposure for 12–48 h at −80°C with an intensifying screen.

2.6. *Cloning of the gene*

For preparation of the DNA fragments, 100 μg *M. vannielii* DNA were digested overnight with 100 units of restriction endonuclease (*Hin*dIII, *Hpa*II) at 37°C. The digest was separated on a preparative 0.8% agarose gel. The DNA in the region of the hybridizing band was cut out, and the DNA eluted in a Biotrap (Schleicher and Schüll, Dassel, FRG) overnight at 4°C at 2 V/cm. The eluted fragments were phenol-extracted and ligated into the pUC18 vector. The ligation mixture was transformed into competent DH5α cells using the calcium chloride procedure [13].

The cells were plated along with 70 μl X-Gal solution (2% 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside in dimethyl formamide) on LB plates containing 100 μg ampicillin/ml. The plates were incubated at 37°C overnight. Recombinants (colorless) were grown in 2 ml LB broth with ampicillin (100 μg/ml) at 37°C overnight, and mini-plasmid preparations were done in pools of up to 10 clones. The plasmid DNA was cut with the appropriate restriction enzyme (*Hin*dIII, *Hpa*II) and separated on a 1% agarose minigel (85 × 70 × 5 mm, at 8 V/cm), blotted onto Gene-Screen and probed as described previously. Pooled plasmid preparations which gave a strong signal were then rescreened separately to identify the positive clones.

2.7. *Subcloning for DNA sequencing*

The inserts from such clones (pMvaY1, pMvaZ1) were prepared by gel elution and then subcloned in M13mp18 and M13mp19. Subfragments were made from the Z1 fragment by digestion with several restriction endonucleases and cloning in M13. Nucleotide sequencing was carried out according to the dideoxy chain-termination method [15] using [α-^{35}S]thio-dATP.

3. RESULTS

3.1. *Amino acid sequence determination*

The N-terminal sequence determination of the HPLC-purified intact protein was performed using a liquid-phase sequencer. The primary structure could be determined up to position 19 with a gap of two unidentified residues at positions 13 and 14 (see fig.2).

3.2. *Synthesis of oligonucleotide probes and hybridization*

The partial protein sequence of *M. vannielii* L23 was used for synthesis of an oligonucleotide probe

N-terminal protein sequence of MvaL23:

Met	Asp	Ala	Phe	Asp	Val	Ile	Lys	Thr	Pro
ATG	GAT	GCA	TTT	GAT	GTA	ATT	AAA	ACA	...
	C	T	C	C	T	C	G	T	
		G			G	A		G	
		C			C			C	

Oligonucleotide synthesized:

ATG	GAT	GCI	TTT	GAT	GTI	ATT	AAG	AC
						C		
						A		

Fig.1. Design of an oligonucleotide probe for the *Methanococcus* L23 ribosomal protein. Two inosines were used at wobble positions of amino acids coded by four different codons. To reduce the degeneration, GT pairing was allowed. At wobble positions with either T or C, always T was used. At positions where either G or A could be the correct match, G was used.

for hybridization with *M. vannielii* chromosomal DNA. The protein sequences were converted into the corresponding nucleotide sequence, and oligonucleotides were synthesized according to the most unequivocable N-terminal part (positions 1–9) as given in fig.1. The mixture of three different oligonucleotides had a length of 26 bases. The oligonucleotide was hybridized and washed as described in section 2. At 45°C the oligonucleotide mixture gave a clear hybridization signal. *Eco*RI, *Hin*dIII, *Cla*I, *Pst*I, *Hpa*II, *Sau*3A and *Dra*I chromosomal digests each yielded one hybridizing band by Southern blot analysis. Signals were detected at 8.5 kb for the *Eco*RI digest, at 500 kb for the *Hin*dIII digest, at 3.3 kb for the *Cla*I digest, at 1.2 kb for *Hpa*II, at 2.0 kb for the *Sau*3A digest and 1.3 kb for *Dra*I.

3.3. *Cloning of Mva L23*

The 500 bp *Hin*dIII digested DNA fragment was chosen for cloning. Direct excision of the band and ligation of the hybridizing fragment into pUC18 yielded the desired clone within a small number of recombinants. The resultant recombinant plasmid pMvaY1 was digested with *Hin*dIII, and the 500 bp fragment was subcloned into M13mp18 and sequenced. Since no stop codon was found in the L23 sequence, the 1.2 kb *Hpa*II fragment which also hybridized with the oligonucleotide was cloned in pUC18. The recombinant plasmid named pMvaZ1 was subcloned for sequencing using *Hin*dIII, *Rsa*I, *Acc*I restriction sites. The L23 gene sequence was determined from bidirectional overlapping fragments; the entire sequence was determined as given in fig.2.

3.4. *Homology of the primary protein structure*

The similarity of the protein sequence of L23 derived from *M. vannielii* with all other archaebacterial and eukaryotic L23 ribosomal proteins of known sequence is shown in fig.3. *M. vannielii* L23 is most similar to the *Halobacterium marismortui* L23 sequence. The yeast L25 protein possesses an N-terminal extended sequence and shows significant similarities to Mva L23. There is also substantial similarity of Mva L23 to eubacterial L23 proteins but less similarity to chloroplast L23 proteins. The alignment scores are shown in fig.4, in which the highest scores are marked.

3.5. *Hydrophobicity of L23 proteins*

Hydrophobicity plots were made, calculating the relative hydrophobicity by normalizing a region of nine residues vs the entire protein [16]. Similar patterns of high hydrophilicity were obtained for all L23 proteins (not shown).

3.6. *Secondary structure prediction*

Secondary structure predictions [17] with 'Predict multi' for the L23 proteins were per-

```
        *******
AGATTAATCTGAGGTGAATGGCATGGATGCCTTTGATGTTATAAAAACACCAATCGTTAGTGAAAAAACAATGAAACTCATTGAAGAAGAAAATCGATTGGTATTT
                  MetAspAlaPheAspValIleLysThrProIleValSerGluLysThrMetLysLeuIleGluGluGluAsnArgLeuValPhe

TACGTTGAAAGAAAAGCTACAAAAGAAGATATTAAAGAAGCAATTAAACAGTTATTCAATGCTGAAGTTGCTGAAGTAAACACGAACATTACTCCAAAAGGACAG
TyrValGluArgLysAlaThrLysGluAspIleLysGluAlaIleLysGlnLeuPheAsnAlaGluValAlaGluValAsnThrAsnIleThrProLysGlyGln
                                                                                                  ******
AAAAAAGCATACATAAAATTAAAAGACGAATACAACGCTGGAGAAGTAGCTGCAAGCTTAGGAATCTATTAATTAATTTAAGAAAACACTATTGAGTAATATGGTGAT
LysLysAlaTyrIleLysLeuLysAspGluTyrAsnAlaGlyGluValAlaAlaSerLeuGlyIleTyrEnd
```

Fig.2. Nucleotide sequence of the *M. vannielii* L23 protein gene. Shine-Dalgarno sequences are marked with stars above the sequence. Underlined residues are identical to those derived from N-terminal sequencing of the HPLC-purified protein.

```
          -55                                                    0                   24
Yea MAPSAKATAA KKAVVKGTNG KKALKVRTSA TFRLPKTLKL ARAPKYASKA VPHYNRLDSY KVIEQPITSE TAMKKVEDGN
Mva ..... ..... ..... ..... ..... ..... .....MDAF DVIKTPIVSE KTMKLIEEEN
Hma ..... ..... ..... ..... ..... ..... .....SW DVIKHPHVTE KAMNDMDFQN
Eco ..... ..... ..... ..... ..... ..... ..MIREERLL KVLRAPHVSE KASTAMEKSN
Bst ..... ..... ..... ..... ..... ..... .....MKDPR DIIKRPIITE NTMN.LIGQK
Mpo ..... ..... ..... ..... ..... ..... .....M NQVKYPVLTE KTIR.LLEKN
Nta ..... ..... ..... ..... ..... ..... .....M DGIKYAVFTD KSIR.LLGKN
Con ..... ..... ..... ..... ..... ..... ..... DVIK.PivTE K.M..le..N

          25                                                                    95
Yea ILVFQVSMKA NKYQIKKAVK ELY.....EV NI..... ....LVRPNG TKKAYVRLTA DYDALDIANR IGYI
Mva RLVFYVERKA TKEDIKEAIK QLFNAEVAEV NT..... ....NITPKG QKKAYIKLKD EYNAGEVAAS LPIY
Hma KLQFAVDDRA SKGEVADAVE EQYDVTVEQV NT..... ....QNTMDG EKKAVVRLSE DDDAQEVASR IGVF
Eco TIVLKVAKDA TKAEIKAAVQ KLFEVEVEVV NTLVVKGKVK RHGQRIGRRS DKKAYVTLKE GQNLDFVGGA E...
Bst KYTFEVDVKA NKTEVKDAVE KIFGVKVEKV NIMNYKGKFK RVGRYSGYTN RKKAIVTLTP DSKEIELFEV ....
Mpo QYSFDVNIDS NKTQIKKWIE LFFNVKVISV NSHRLPKKKK KIGTTTGYTV RYKRMIIKLQ SGYSIPLFSN K...
Nta QYTSNVESGS TRTEIKHWVE LFFGVKVIAM NSHRLPGKSR RMGPIMGHTM HYRRMIITLQ PGYSIPPLRK KRT.
Con ..vF.V..kA .KteIK.AVE .lF.VkVe.V NT.....k.. ..g...g.tg rKKAyv.L.. d..aievo.. ....
```

Fig.3. Comparison of ribosomal L23 protein sequences from yeast [18], *Methanococcus vannielii* (this paper), *Halobacterium marismortui* [19], *E. coli* [20], *Bacillus stearothermophilus* [21], liverwort [22] and tobacco [23]. Symbols in the consensus sequence (Con): Lower case characters were used at positions of three identical residues and capital letters at positions of at least four identical residues. Underlined capital letters indicate that at least six amino acids out of the seven sequences are identical at this position. For abbreviations see fig.5.

formed in order to search for homologous regions in the multiple alignment (fig.5). These predictions gave similar results for various L23 sequences.

	YeaL25	MvaL23	HmaL23	BstL23	EcoL23	MpoL23	NtaL23
YeaL25	----						
MvaL23	[18.6]	----					
HmaL23	15.2	[21.0]	----				
BstL23	5.4	9.7	11.7	----			
EcoL23	7.8	[17.2]	10.8	9.5	----		
MpoL23	3.7	9.6	4.3	[18.4]	5.9	----	
NtaL23	0.0	6.7	1.9	14.5	4.8	[37.3]	----

Fig.4. Alignment scores for comparison of L23 proteins. Parameters: mutation data matrix (MD) comparison table, 100 random runs, break penalty was set to 20. For each organism the highest score was framed. For abbreviations see fig.5.

4. DISCUSSION

4.1. *Comparison of the nucleotide and amino acid sequence*

When the nucleotide sequence at the 5′-end of the gene was compared to the amino acid sequence at the N-terminal region of the protein, full agreement was found. These results and the excellent homology to all other L23 proteins show that the protein and the reading frame are correctly identified.

4.2. *Primary structure of the DNA*

A Shine-Dalgarno sequence (GAGGTGA) was found, starting 12 bases upstream of the structural gene. The L23 gene starts with an ATG codon for methionine and ends with a TAA stop codon. Another Shine-Dalgarno sequence was found in front of a new open reading frame located 40 bases downstream of the stop codon.

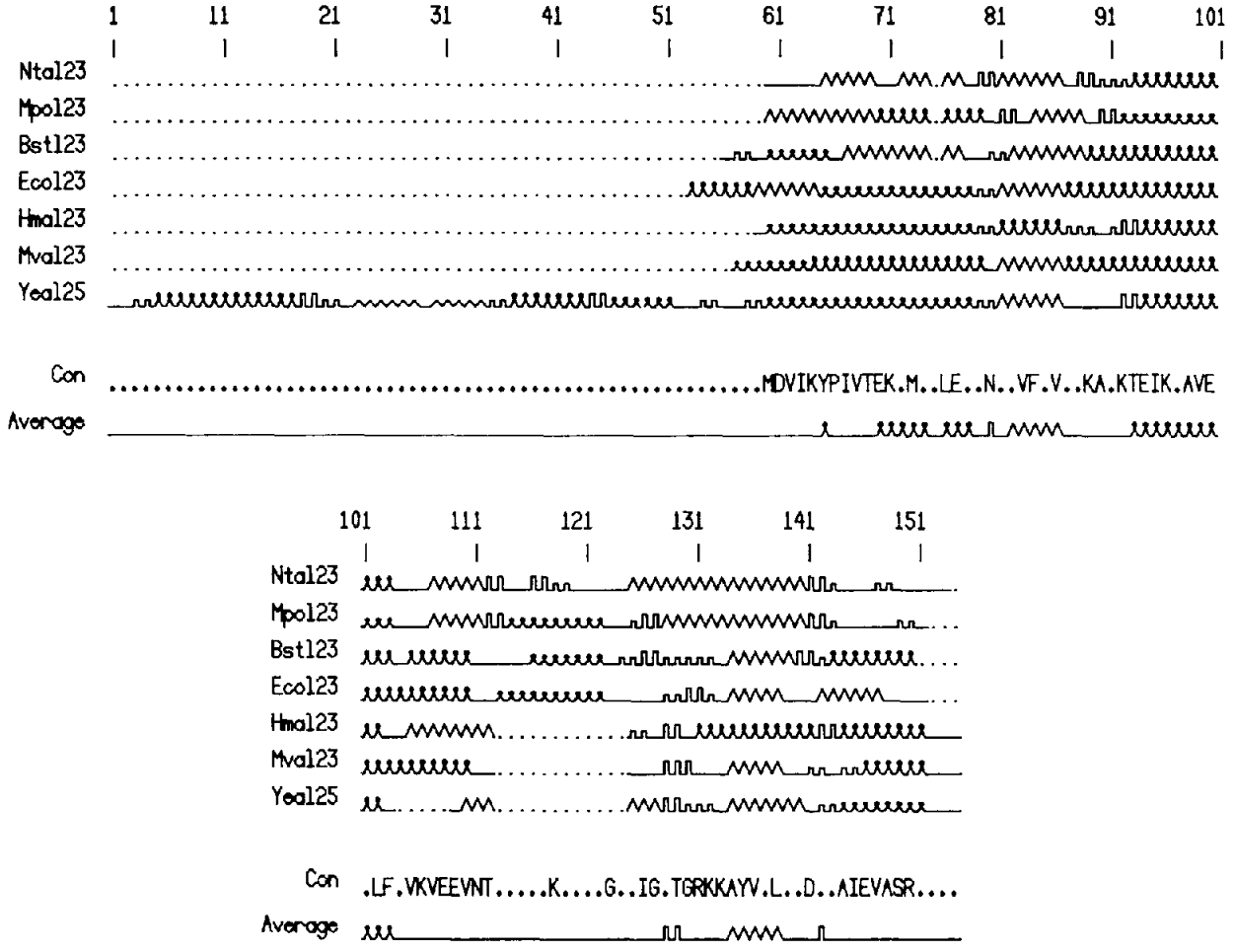

Fig.5. Multiple secondary structure prediction of the prealigned L23 protein sequences with the program PREDICT-MULTI. Yea, yeast; Mva, *Methanococcus vannielii*; Hma, *Halobacterium marismortui*; Eco, *E. coli*; Bst, *Bacillus stearothermophilus*; Mpo, liverwort; Nta, tobacco; Con, conserved residues. The 'average' has to be at least 60% similar for all sequences. Symbols: loops, α-helix; zig-zags, β-sheets; meanders, β-turns; straight lines, random coils.

4.3. *Evolution of the L23 protein*

The number of identical residues of the *M. vannielii* L23 protein with other proteins varies from 36 with the yeast sequence, 30 with *H. marismortui*, 32 with *E. coli* and 24 with *B. stearothermophilus* to 18 for liverwort chloroplast and 13 for tobacco chloroplast L23 protein. In contrast, the align scores also consider the length and the conservative changes in the primary sequence and not only the identical residues, so that somewhat different results are obtained: The L23 protein from yeast has an N-terminal extended sequence, and therefore the align score is not as high as the number of identical residues would predict. In the *Halobacterium* sequence the amino acid exchanges are more conservative, so that the align score is higher than expected from the relatively low number of identical residues.

From the alignment scores it is obvious that good sequence similarities exist for the archaebacterial (HmaL23), eukaryotic (YeaL25) and Gram-negative eubacterial (EcoL23) representatives, while the Gram-positive eubacteria (BstL23) and the chloroplast proteins (MpoL23, NtaL23) are less related. Although the scores suggest a homology gradient in the order ar-

chaebacteria, eukaryotes and eubacteria, the differences in their values are not high enough to be significant.

Acknowledgements: We thank Dr August Böck for frozen *Methanococcus vannielii* cells, Roza Maria Kamp and Helga Eckardt for purifying the L23 protein by HPLC, Anita Marchfelder for experimental assistance and Christoph Weigel for valuable discussions. Dr J. McDougall is thanked for carefully reading the English version of the manuscript.

REFERENCES

[1] Wittmann, H.G. (1986) in: Structure, Function and Genetics of Ribosomes (Hardesty, B. and Kramer, G. eds) pp.1–27, Springer, New York.
[2] Wittmann-Liebold, B. (1986) in: Structure, Function and Genetics of Ribosomes (Hardesty, B. and Kramer, G. eds) pp.326–361, Springer, New York.
[3] Strobel, O., Köpke, A.K.E., Kamp, R.M., Böck, A. and Wittmann-Liebold, B. (1988) J. Biol. Chem. 263, 6538–6546.
[4] Schmid, G. and Böck, A. (1981) J. Bacteriol. 147, 282–288.
[5] Devereux, J., Haeberli, P. and Smithies, O. (1984) Nucleic Acids Res. 12, 387–395.
[6] Wittmann, H.G. (1974) in: Ribosomes (Nomura, M. et al. eds) pp.93–114, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
[7] Kamp, R.M., Bosserhoff, A., Kamp, D. and Wittmann-Liebold, B. (1984) J. Chromatogr. 317, 181–192.
[8] Kamp, R.M. and Wittmann-Liebold, B. (1988) Methods Enzymol., in press.
[9] Kaltschmidt, E. and Wittmann, H.G. (1970) Anal. Biochem. 36, 401–412.
[10] Wittmann-Liebold, B. (1983) in: Modern Methods in Protein Chemistry (Tschesche, H. ed.) pp.229–266, De Gruyter, Berlin.
[11] Ashman, K. and Wittmann-Liebold, B. (1985) FEBS Lett. 190, 129–132.
[12] Hofman, J.D., Lau, R.H. and Doolittle, W.F. (1979) Nucleic Acids Res. 7, 1321–1333.
[13] Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
[14] Church, M. and Gilbert, W. (1984) Proc. Natl. Acad. Sci. USA 81, 1991–1995.
[15] Sanger, F. and Coulson, A.R. (1978) FEBS Lett. 87, 107–110.
[16] Kyte, J. and Doolittle, R.F. (1982) J. Mol. Biol. 157, 105–132.
[17] Chou, P.Y. and Fasman, G.D. (1978) Adv. Enzymol. 47, 45–148.
[18] Leer, J.J., Van Raamsdonk-Duin, M.M.C., Hagendoorn, M.J.M., Mager, W.H. and Planta, R.J. (1984) Nucleic Acids Res. 12, 6685–6700.
[19] Hatakeyama, T. and Kimura, M. (1988) Eur. J. Biochem. 172, 703–711.
[20] Wittmann-Liebold, B. and Greuer, B. (1979) FEBS Lett. 108, 69–74.
[21] Kimura, M., Kimura, J. and Ashman, K. (1985) Eur. J. Biochem. 150, 491–497.
[22] Ohyama, K. et al. (1986) Nature 322, 572–574.
[23] Shinozaki, K. et al. (1986) EMBO J. 5, 2043–2049.