Volume 15 Number 5 1987 Nucleic Acids Research

Nucleotide sequence of the *melA* gene, coding for α-galactosidase in *Escherichia coli* K-12

Pirkko L.Liljeström[1]* and Peter Liljeström[2]

[1]Research Laboratories of the Finnish State Alcohol Company (Alko Ltd), PO Box 350, SF-00101 Helsinki and [2]Molecular Genetics Laboratory, Department of Genetics, University of Helsinki, Arkadiankatu 7, SF-00100 Helsinki, Finland

Received December 29, 1986; Accepted February 4, 1987

ABSTRACT

Melibiose uptake and hydrolysis in *E.coli* is performed by the MelB and MelA proteins, respectively. We report the cloning and sequencing of the *melA* gene. The nucleotide sequence data showed that *melA* codes for a 450 amino acid long protein with a molecular weight of 50.6 kd. The sequence data also supported the assumption that the *mel* locus forms an operon with *melA* in proximal position. A comparison of MelA with α-galactosidase proteins from yeast and human origin showed that these proteins have only limited homology, the yeast and human proteins being more related. However, regions common to all three proteins were found indicating sequences that might comprise the active site of α-galactosidase.

INTRODUCTION

Melibiose utilization in *E.coli* is dependent on the melibiose locus (*mel*), which is located at 93 min on the genetic map (1). The *mel* locus is induced by several α-D-galactosides (2,3) and its expression is controlled by the cAMP-CRP regulatory circuit (4). Melibiose is also an inducer of the *lac* operon and can be transported into the cell by the lactose permease protein LacY. In fact, as expression of *melB* is temperature sensitive, influx of melibiose at 38-40°C is dependent on LacY (5,6).

The exact structure of the *mel* is poorly defined. It is believed to form an operon consisting of at least 2 structural genes, *melA* and *melB*, coding for α-galactosidase and the melibiose carrier respectively (7). Moreover, the inducibility of the operon indicates that there might exist a repressor gene, similar to *lacI*. The promoter proximal gene *melA* codes for an α-galactosidase which hydrolyzes melibiose into glucose and galactose. The *melA* gene product has been identified (7) and purified (8) and the results from these studies have indicated a tetrameric structure with a molecular weight of 200kd for the active enzyme.

The *melAB* region has been cloned previously and the promoter region together with the *melB* gene has been sequenced (9,10). We are studying melibiose utilization and report here the cloning and sequencing of the complete *melA* gene together with 5' and 3'-flanking regions. In addition, the deduced amino acid se-

quence of *E.coli* α-galactosidase is compared with proteins of yeast and human origin. The results suggested common functional domains for all three proteins.

MATERIALS AND METHODS

Bacterial strains, plasmids and phages

Bacterial strains used in this study were: *Escherichia coli* K-12 strain B14, wild type and strain PC103, *pro*, *thi*, *endA*, *rpsL*, *supE44*, *hsdR17*, *melA6*, *zje*::Tn*10*. The *melA6* mutation was introduced into PC103 by P1-transduction from BJW43 using a selection for tetracycline resistance provided by the adjacent Tn10. The phage vectors used for sequencing were M13mpl8 and M13mp19 (11). Plasmid pALK68, used as source of DNA in the sequence determination, carries the *melAB* operon on a 4.7kb fragment cloned into the *Eco*RI site of pBR322. For sequence determination of the 5' region of the *melA* gene plasmid pALK70 was used. This plasmid carries the 5' region of the *melAB* operon as a 1.6 kb *Hin*dIII-*Bgl*II fragment in pUC18.

Enzymes and reagents

Restriction endonucleases, T4-DNA ligase, DNA-polymerase I Klenow fragment and M13-hybridization primers were purchased from Boehringer-Mannheim. Deoxynucleoside triphosphates and dideoxynucleoside triphosphates were from P.L.Biochemicals. α-^{35}S-dATP-α-S (600 Ci/mmol) from Amersham was used for labeling.

DNA-sequencing

Determination of DNA sequence was performed by the dideoxy-chain termination method using bacteriophage M13 (12,13) or by sequencing directly from plasmid DNA (14).

RESULTS AND DISCUSSION

Isolation of the *melA* gene

The *mel* region of *E.coli* K-12 was shotgun cloned from total DNA of strain B14 into pBR322 as a partial *Sau*3A fragment. The ligation mixture was transformed into strain PC103 (*melA6*) and transformants were selected on minimal plates containing melibiose as a sole carbon source and ampicillin. One of the plasmids obtained (pALK61) carried a 8.0 kb insert and was chosen for further study. Subsequent subcloning and complementation analysis showed that the *melA* gene was carried on a 4.7 kb *Eco*RI fragment (Fig.1). The restriction map obtained was in agreement with that reported earlier (7).

Nucleotide sequence of the *melA* gene

The complete nucleotide sequence of the *melA* region was determined from

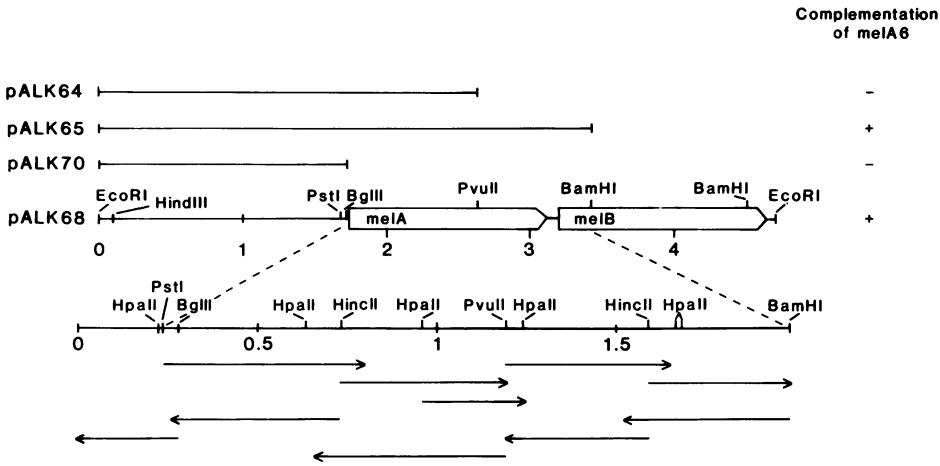

Fig.1. Complementation analysis of mel subclones and sequencing strategy of the *melA* gene. The restriction sites used for the sequencing are shown along with arrows indicating the length and direction of sequence determined by the chain termination method. The position of the *melA* and *melB* genes is shown.

both strands and the junction points were sequenced from overlapping fragments. The sequencing strategy is shown in Figure 1 and Figure 2 presents the nucleotide sequence obtained. Parts of the sequence (nucleotides -229 to +157 and 1411-1504) has been published earlier (7,10). We detected some differences from the published sequence: The G's at -219, -220 and +138 (Fig.2) were earlier reported as A, C and A respectively, and the C at +1416 as a T. Moreover, the G at -67 is missing in the previous sequence.

Coding region

It has been suggested (7) that the *melA* coding region would be located downstream of the *Pst*I site at -52 (Fig.2). According to the previous study the ATG codons at -85 and +1 (Fig.2) were in the same reading frame and hence it was uncertain, which one was used in vivo. However, the revised nucleotide sequence has an additional G at position -67 which brings the two ATG codons in different reading frames. Accordingly, the ATG at position +1 is most probably used as translation initiation of *melA*. To confirm this observation the following analysis was performed. The plasmid pALK69 was opened at the *Pst*1 site at -52 to -46 (Fig.2) and treated with Klenow enzyme. The religated plasmid pALK73 had lost 4 nucleotides (-51 to -45) between the ATG codons; This was confirmed by DNA-sequence analysis (data not shown). If the ATG at -85 would be used in vivo, then the *melA* gene on pALK73 would not direct the synthesis of wild type α-galactosidase since the reading frame is changed 11 amino acids from the start.

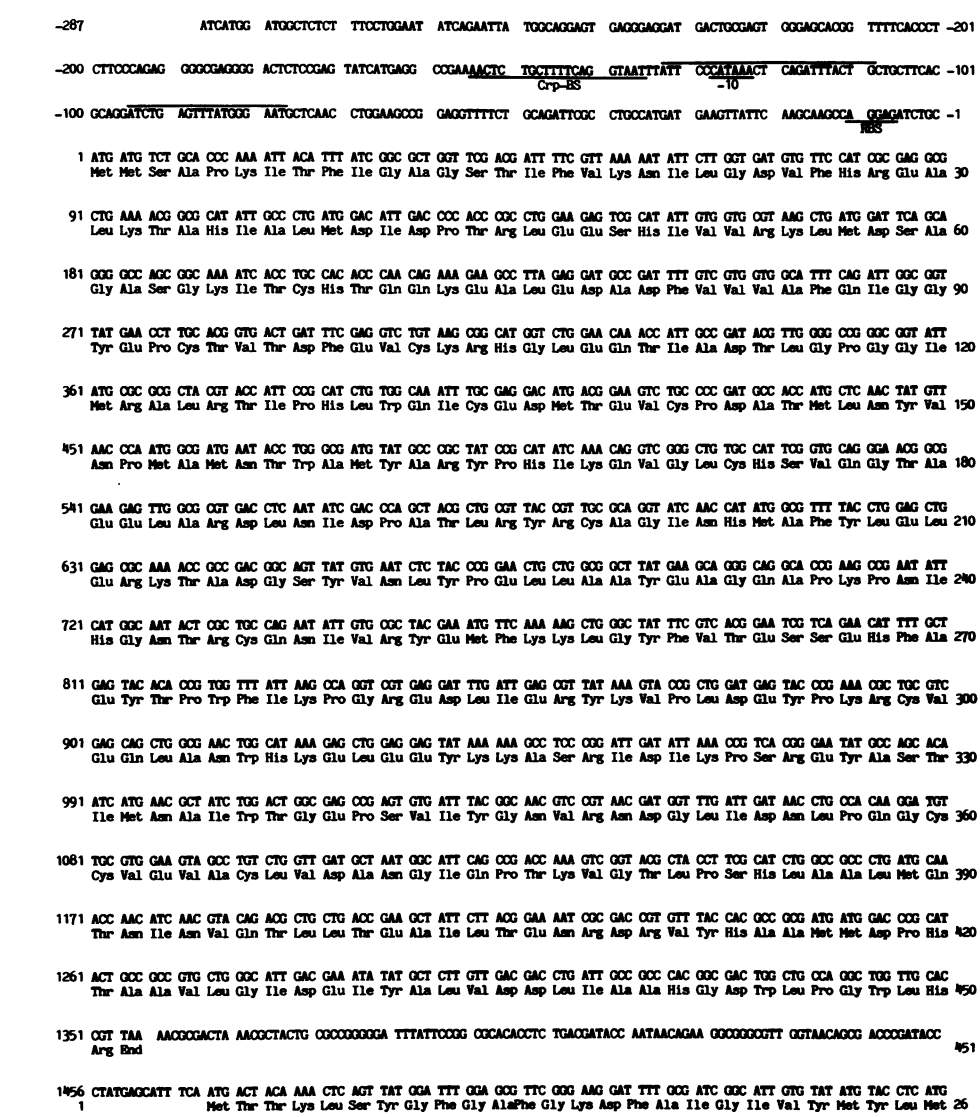

Fig.2. Nucleotide and amino acid sequence of the *melA* gene. The complete sequence of a 1986 nucleotide region containing the *melA* structural gene is shown. The numbering of nucleotides begins at the A in the translation initiating ATG. The putative -10 region is underlined. RBS denotes the ribosome binding site. Overlinings indicate an inverted repeat.

However, pALK73 did complement the *melA* mutation in PC103, confirming that the second ATG is used as start codon. This ATG is preceeded by a typical ribosome binding site (15,16).

The nucleotide sequence in Figure 2 containes only one long open reading

Table 1. Codon usage of the *melA* gene of *Escherichia coli*

TTT	Phe	6	(1.3)	TCT	Ser	1	(0.2)	TAT	Tyr	11	(2.4)	TGT	Cys	3	(0.7)
TTC	Phe	5	(1.1)	TCC	Ser	1	(0.2)	TAC	Tyr	8	(1.8)	TGC	Cys	9	(2.0)
TTA	Leu	1	(0.2)	TCA	Ser	3	(0.7)	TAA	End	1	(0.2)	TGA	End	0	(0.0)
TTG	Leu	5	(1.1)	TCG	Ser	5	(1.1)	TAG	End	0	(0.0)	TGG	Trp	7	(1.5)
CTT	Leu	3	(0.7)	CCT	Pro	2	(0.4)	CAT	His	13	(2.9)	CGT	Arg	10	(2.2)
CTC	Leu	3	(0.7)	CCC	Pro	3	(0.7)	CAC	His	4	(0.9)	CGC	Arg	9	(2.0)
CTA	Leu	2	(0.4)	CCA	Pro	5	(1.1)	CAA	Gln	5	(1.1)	CGA	Arg	0	(0.0)
CTG	Leu	25	(5.5)	CCG	Pro	13	(2.9)	CAG	Gln	9	(2.0)	CGG	Arg	3	(0.7)
ATT	Ile	23	(5.1)	ACT	Thr	4	(0.9)	AAT	Asn	9	(2.0)	AGT	Ser	2	(0.4)
ATC	Ile	8	(1.8)	ACC	Thr	11	(2.4)	AAC	Asn	10	(2.2)	AGC	Ser	2	(0.4)
ATA	Ile	1	(0.2)	ACA	Thr	3	(0.7)	AAA	Lys	15	(3.3)	AGA	Arg	0	(0.0)
ATG	Met	16	(3.5)	ACG	Thr	11	(2.4)	AAG	Lys	5	(1.1)	AGG	Arg	0	(0.0)
GTT	Val	5	(1.1)	GCT	Ala	8	(1.8)	GAT	Asp	13	(2.9)	GGT	Gly	9	(2.0)
GTC	Val	8	(1.8)	GCC	Ala	18	(4.0)	GAC	Asp	12	(2.7)	GGC	Gly	13	(2.9)
GTA	Val	3	(0.7)	GCA	Ala	6	(1.3)	GAA	Glu	16	(3.5)	GGA	Gly	2	(0.4)
GTG	Val	12	(2.7)	GCG	Ala	11	(2.4)	GAG	Glu	17	(3.8)	GGG	Gly	4	(0.9)

frame beginning with ATG at position +1 and ending with TAA at +1354. The molecular weight of α-galactosidase calculated from the predicted amino acid sequence is 50,662 d, which is in good agreement with the apparent molecular weight of the identified *melA* gene product of 50 kd estimated by SDS-gel electrophoresis (7).

Observed patterns of codon usage correlate with the availability of isoaccepting tRNAs, base composition and codon-anticodon interaction energies. These are means of controlling the expressivity of genes and the accuracy of their translation (17-21). However, genes expressed at low levels appear to have a less restrictive codon usage and therefore use of rare codons is unusually high. The codon usage of *melA* is of the type found in moderately expressed genes (Table 1).

5' noncoding region

The area upstream of the *melA* coding region does not show any obvious sites with strong homology with the consensus promoter (22-25). The site indicated in Figure 2 as -10 has been suggested earlier (9). Although this site is not preceeded by a consensus -35 sequence, it is close to a region with strong homology (85%) with the CRP binding site (26). Genes which are under positive transcriptional control frequently lack the -35 consensus region, as they use regulatory proteins such as CRP to enhance transcription initiation (27). Hence, as *mel* expression is known to be catabolite repressible (4), it is probable that the promoter of *melAB* is contained within this region. The promoter area also contains a stretch of dyad symmetry, which shows strong resemblance to the REP sequences, and which are believed to function either in mRNA stabilization or in organization of chromosome structure (28-30).

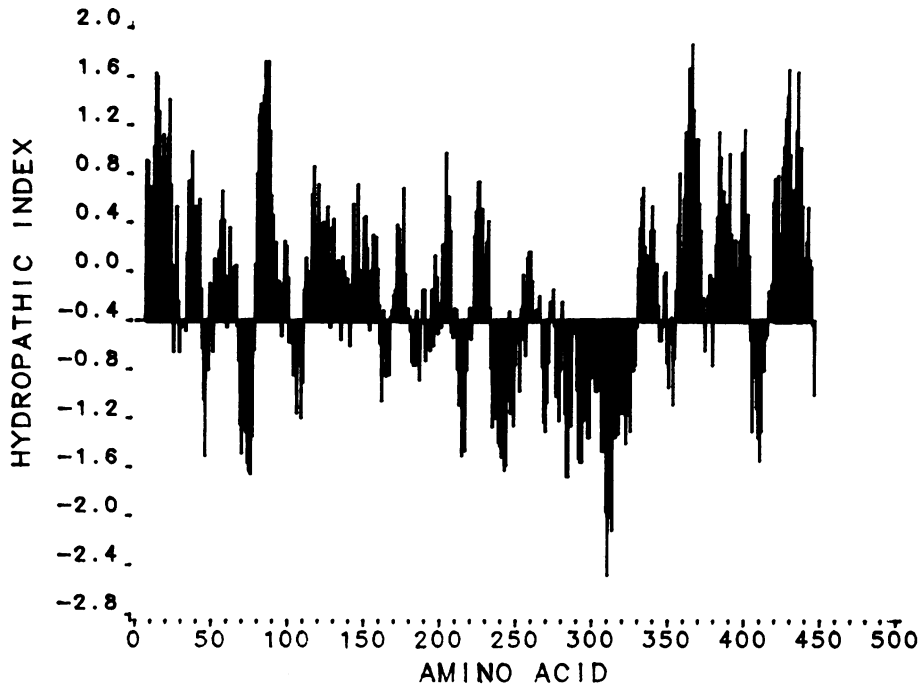

Fig.3. Hydropathic profile of α-galactosidase encoded by *melA*. Hydropathic index is calculated according to Kyte and Doolittle (31) with the base line set at a value of -0.4. Positive and negative values indicate hydrophobic and hydrophilic regions, respectively.

Protein characteristics

To characterize the α-galactosidase protein as deduced from the *melA* sequence, its amino acid sequence was analyzed for hydropathic profile (31) as shown in Figure 3. MelA (α-galactosidase) was shown to be composed of alternating hydrophobic and hydrophilic regions with an average hydropathy of -0.18. This is consistent with α-galactosidase being a soluble protein, as the average hydropathy for membrane embedded proteins should be greater than +0.5 (31). There are also no hydrophobic sequences in α-galactosidase that would be long and hydrophobic enough to span a membrane (31,32).

The hydropathic profile of MelA (Fig.3) was also compared with that of yeast (33) and human α-galactosidase (34). However, no obvious similarities to either one were found. The amino acid sequences of the three α-galactosidases were also compared. The yeast and the human enzymes showed several regions with good homology (Fig.4). In comparison, only a few short homologous regions were

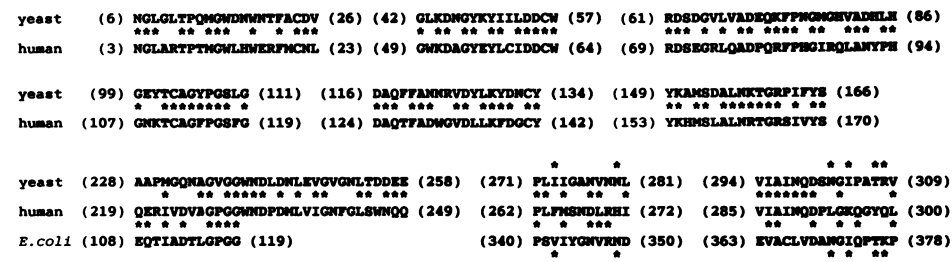

Fig.4. Amino acid sequence homologies between yeast, human and *E.coli* α-galactosidases. The amino acids have been grouped according to similarity: 1, L I V M; 2, F Y W; 3, K R H; 4, D N; 5, E Q; 6, S T. Asterisks indicate similarites between sequences. Asterisks outside the yeast and bacterial sequences indicate homologies between these sequences that are not shared by the human protein.

found between MelA of *E.coli* and the -galactosidase proteins of yeast and human origin.

α-galactosidases (α-D-galactoside galactohydrolase, EC 3.2.1.22) catalyse the hydrolysis of saccharides containing α-1,6,-galactoside linkages. The three α -galactosidases compared in Fig.4 all catalyse the same reaction, but are localized in different cellular compartments: The *E.coli* enzyme is cytoplasmic, the human enzyme is lysosomal and the yeast enzyme extracellular. Therefore,although the active enzyme from all three species have nearly the same molecular weight, structural similarities, as well as differences, are expected. It is not surprising that the yeast and the human enzymes are more similar with each other than with the *E.coli* enzyme, because they are both secretory proteins while MelA is cytoplasmic. The fact that only limited regions of human α -galactosidase and yeast α-galactosidase are highly conserved strongly suggests that these regions specify important common structural and functional domains of the enzyme. Because only three of these regions share homology to the *E.coli* α-galactosidase,it may be that these regions are involved in substrate recognition or are parts of the active catalytic site. The third region homologous in all three enzymes (Fig.4) can be predicted (35) to have similar secondary structures. An interesting possibility for an active site is position 369 (*E.coli*), which has an aspartate residue initiating a β-turn between two β-sheet structures. However, more work using site specific mutagenesis is needed to pinpoint the exact location of the active site.

ACKNOWLEDGEMENTS

We wish to thank Outi Nikkilä for skillful technical assistance, Richard Fagerström, Matti Korhola and Roy Tubb for useful discussions and John Denslow for

computerizing hydroplot drawing. The strain BJW43, origin of the *melA6* mutation was a kind gift from Brian Smith-White.

*To whom correspondence should be addressed

REFERENCES

1. Bachmann, B.J. (1983). Microbiol. Rev. 47, 180-230.
2. Prestidge, L.S. and Pardee, A.B. (1965). Biochim. Biophys. Acta 100, 591-593.
3. Schmitt, R. (1968). J. Bacteriol. 96, 462-471.
4. Okada, T., Ueyama, K., Niiya, S., Kanazawa, H., Futai, M. and Tsuchiya, T. (1981). J. Bacteriol. 146, 1030-1037.
5. Beckwith, J.R. (1963). Biochim. Biophys. Acta 76, 162-164.
6. Prestidge, L.S. and Pardee, A.B. (1965). Biochim. Biophys. Acta 100, 591-593.
7. Hanatani, M., Yazyu, H., Shiota-Niiya, S., Moriyama, Y., Kanazawa, H., Futai, M. and Tsuchiya, T. (1984). J. Biol. Chem. 259, 1807-1812.
8. Burstein, C. and Kepes, A. (1971). Biochim. Biophys. Acta 230, 52-63.
9. Shimamoto, T., Yazyu, H., Futai, M. and Tsuchiya, T. (19894). Biochem. Biophys.Res.Commun. 121, 41-46.
10. Yazyu, H., Shiot-Niiya, S., Shimamoto, T., Kanazawa, H., Futai, M. and Tsuchiya, T. (1984). J. Biol. Chem. 259, 4320-4326.
11. Norrander, J., Kempe, T. and Messing, J. (1983). Gene 26, 101-106.
12. Sanger, F., Nicklen, S. and Coulsen, A.R. (1977). Proc. Natl. Acad. Sci USA 74, 5463-5467.
13. Messing, J. (1983). Methods in Enzymol. 101, 20-78.
14. Zagursky, R.J., Baumeister, K., Lomax, N. and Berman, M.L. (1985). Gene Anal. Techn. 2, 89-94.
15. Shine, J. and Dalgarno, L. (1979). Proc.Natl.Acad.Sci.USA 71, 1342-1346.
16. Kozak, M. (1983). Micobiol.Rev. 47, 1-45.
17. Ikemura, T. (1981). J. Mol. Biol. 146,1-21.
18. Ikemura, T. (1981). J. Mol. Biol. 151, 389-409.
19. Gouy, M. and Gautier, C. (1982). Nucl. Acids Res. 10, 7055-5054.
20. Grosjean, H. and Fiers, W. (1982). Gene 18, 199-209.
21. Bibb, M.J., Findlay, P.R. and Johson, M.W. (1984). Gene 30, 157-166.
22. Hawley, D.K. and McClure, W.R. (1983). Nucl.Acids Res. 11, 2237-2255.
23. von Hippel, P.H., Bear, D.G., Morgan, W.D. and McSwiggen, J.A. (1984). Ann.Rev.Biochem. 53, 389-446.
24. McClure, W.R. (1985). Ann.Rev.Biochem. 54, 171-204.
25. Reznikoff, W.S., Siegele, D.A., Cowing, D.W. and Gross, C.A. (1985). Ann.Rev.Genet. 19, 355-387.
26. Cossart, P., Groisman, E.A., Serre, M.-C., Casadaban,M.J. and Gicquel-Sanzey, B. (1986). J.Bacteriol. 167, 639-646.
27. Raibaud, O. and Schwartz. M. (1984). Ann.Rev.Genet. 18, 173-206.
28. Gilson, E., Clement, J.-M., Brutlag, D. and Hofnung, M. (1984). EMBO J. 3, 1417-1421.
29. Stern, M.J., Ames, G.F.-L., Smith, N.H., Robinson, E.C. and Higgins, C.F. (1984). Cell 37, 1015-1026.
30. Becerril, B., Valle, F., Merino, E., Riba, L. and Bolivar, F. (1985). Gene 37, 53-62.
31. Kyte, J. and Doolittle, R.F. (1982). J. Mol. Biol. 157, 105-132.
32. Eisenberg, D., Schwarz, E., Komaromy, M. and Wall, R.J. (1984). J. Mol. Biol. 179, 125-142.
33. Liljeström, P.L. (1985). Nucl. Acids Res. 13, 7257-7268.
34. Bishop, D.F., Calhoun, D.H., Bernstein, H.S., Hantzopoulos, P., Quinn, M. and Desnick, R.J. (1986). Proc. Natl. Acad. Sci. USA 83, 4859-4863.
35. Chou, P.Y. and Fasman, G.D. (1978). Adv.Enzymol. 47,45-148.