The EMBO Journal vol.3 no.8 pp.1799–1802, 1984

The pro- and mature forms of the *E. coli* K-12 outer membrane phospholipase A are identical

Pieter de Geus, Hubertus M.Verheij, Nico H.Riegman, Wiel P.M.Hoekstra[1] and Gerard H.de Haas

Laboratory of Biochemistry, and [1]Department of Molecular Cell Biology, State University of Utrecht, Transitorium III, University Centre 'de Uithof', Padualaan 8, NL-3584 CH Utrecht, The Netherlands

Communicated by L.L.M.van Deenen

The nucleotide sequence of the *pldA* gene, coding for the outer membrane (OM) phospholipase A of *Escherichia coli* K-12, and flanking sequences, was determined. Data were obtained from sequences of overlapping deletions which had been generated *in vitro* from both ends of the gene, using DNase I in the presence of Mn^{2+} and *Bal*31 nuclease. The deduced amino acid sequence of the *pldA* gene product is the first primary sequence of a membrane-bound phospholipase. The complete PldA protein contains 260 amino acids, which include a putative signal sequence, and has a calculated mol. wt. of 29 946 similar to that of the purified protein. Furthermore we found the N terminus of the purified protein to be blocked and the overall amino acid composition to be consistent with the one deduced from the complete *pldA* gene. Analysis of proteins synthesized in minicells with a *pldA* coding plasmid in the presence of 8% ethanol did not reveal any new bands on polyacrylamide gels, whereas the control β-lactamase clearly showed its unprocessed form under the same conditions. These data are consistent with the empirical prediction from the primary sequence, that the PldA protein lacks any signal peptidase 'target' site. We therefore conclude that the PldA protein is exported to the OM without proteolytic removal of the signal peptide.

Key words: outer membrane proteins/signal sequence/nucleotide sequence/phospholipase A/protein export

Introduction

Phospholipase A is one of the very few enzymes located in the outer membrane (OM) of *Escherichia coli* K-12. The phospholipase has an apparent mol. wt. of 29 000 daltons (29 K) and is absolutely dependent on Ca^{2+} ions for enzymatic activity. The protein is considered to be very tightly bound to the lipid constituents of the OM (Scandella and Kornberg, 1971), and is only partly susceptible to proteolytic attack (Nishijima *et al.*, 1977). Thus, one may expect that the enzyme is an integral membrane protein embedded in the OM lipid bilayer of phospholipids and lipopolysaccharides (LPS). Consequently, interesting questions arise about the enzyme's regulation of activity and possible function, particularly since membrane-bound phospholipases are so widespread in nature (for a review, see van den Bosch, 1980). So far, no definite answers have been obtained because studies on membrane-bound phospholipases are always limited by the very low amount of these proteins present.

We cloned the *pldA* gene of *E. coli* K-12 on multicopy plasmids and showed it to be the structural gene of the phospholipase (de Geus *et al.*, 1983). The cloned *pldA* gene in principle should allow us to enhance the low level of phospholipase in *E. coli* and in addition provide new approaches for research on this enzyme. We have now determined the primary structure of the enzyme from the nucleotide sequence of the *pldA* gene. This is, we believe, the first example of a primary structure of a membrane-bound phospholipase. The deduced amino acid sequence, together with some data obtained on the purified enzyme from overproducing strains (to be published), has some intriguing implications for the biogenesis and possibly also for enzymatic activity of PldA protein.

Results

Nucleotide sequence determination

We started the sequence determination with a semi-random approach, because *pldA* lacks sufficient restriction sites. Initial cloning of the 4.0-kb DNA stretch containing *pldA* into mp9 was performed as described in Materials and methods. This placed unique restriction sites at both ends of the gene, resulting in a derivative phage M17 (Figure 1). Before further subcloning, the stability of the inserted DNA was determined by cloning the 4.0-kb *pldA* fragment from M17 into *Eco*RI-*Hind*III treated pBR322. The resulting hybrid restored the *pldA*$^+$ phenotype of strain CE1303 in our test system, thereby showing that the *pldA* gene is stable in M13 vectors in JM103. A deletion library of *pldA* was then constructed *in vitro* using DNase I and *Bal*31 nuclease (Figure 1). As can be seen in Figure 1, the deletions moved progressively inwards, either from the N-terminal – (series I and III) or C-terminal – (series II) side of the PldA protein, thereby providing sequence data of both strands of the complete *pldA* gene and flanking regions.

Nucleotide sequence of pldA

The nucleotide sequence of *pldA* is shown in Figure 2. Several

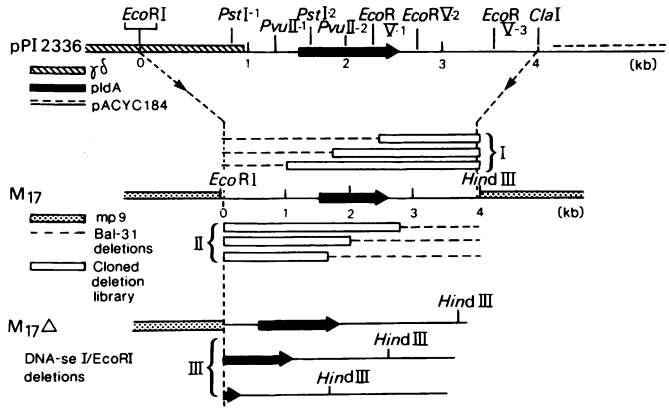

Fig. 1. Sequencing strategy. The upper part of the figure shows the construction of M17, the mp9 vector carrying the entire *pldA* gene. The middle and lower parts schematically show the construction of the *pldA* deletion library. Symbols are explained in the figure except for the Roman numerals I, II and III, which are explained in Materials and methods.

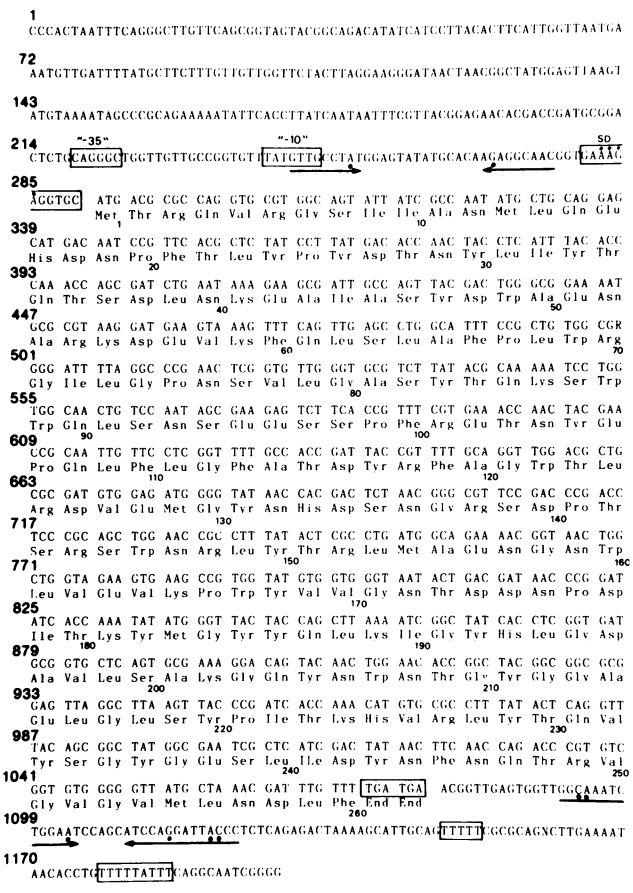

Fig. 2. Sequence of the 1.2-kb DNA fragment containing the *pldA* gene, and translation product of *pldA*. Shown is the 5′ to 3′ non-coding strand with sequences discussed in the text in boxes. These include a '−35' and '−10' promoter region, a ribosome binding site [or Shine and Dalgarno sequence (SD)], and possible terminator structures. Inverted repeats that can form stable hairpin loops are marked by arrows with dots representing mismatches.

regulatory factors can be assigned to some sequences indicated in Figure 2. They include a ribosome binding site (SD) between positions 280 and 290 where eight out of 11 nucleotides base-pair with the 3′-terminal nucleotides from 16S rRNA (Shine and Dalgarno, 1974). A Pribnow box (consensus -TATAATG-) and '−35' region (consensus -TTGACA-), which have been suggested to be part of the *E. coli* promoter (Siebenlist *et al.*, 1980; Rosenberg and Court, 1979), are located at position 245 (-TATGTTG-) and position 220 (-CAGGGC-), respectively, although the latter contains a poor homology with the consensus sequence.

Between the two stop codons (at positions 1071 and 1074) and the first poly(T) sequence (at 1148) a rather weak inverted repeat can be found (Figure 2). According to Tinocco *et al.*(1973) the calculated free energy of the possible stem and loop structure is −9.6-kcal/mol. These data indicate the presence of a rho-independent termination site, though far from optimal with respect to relative positions of the different termination structures (cf. Rosenberg and Court, 1979).

Primary structure and biogenesis of pldA

Figure 2 also shows the primary structure of the PldA protein deduced from the nucleotide sequence. The complete absence of cysteine and the high content of tryptophan (9) is noteworthy (see Table I). Inspection of the codons used for the various amino acids shows a clear preference for codons leading to a low expression, as pointed out by Grosjean and Fiers (1982). This is consistent with the low level of phospholipase − 500 molecules/cell − in wild-type cells.

Table I. Amino acid composition of phospholipase A

	Protein[a] data	DNA[b] data		Protein data	DNA data		Protein data	DNA data
Asx	35.0	35	Cys	0	0	Lys	n.d.	9
Thr[c]	18.9	17	Val	15.4	14	His	4.3	4
Ser[c]	19.6	19	Met[d]	5.8	6	Arg	11.7	13
Glx	27.7	24	Ile	9.0	9	Trp[e]	8−9	9
Gly	24.2	22	Leu	23.0	25	Pro	11.0	10
Ala	14.6	13	Tyr[d]	23.3	22			
			Phe	10.0	9			

[a]Average value from five analyses. Estimated error 5 − 10%.
[b]Amino acid composition predicted for the whole unprocessed protein.
[c]Corrected for losses during hydrolysis.
[d]Determined by hydrolysis in the presence of 4% thioglycollic acid.
[e]Determined spectrophotometrically.

The molecular weight (29 946) and the amino acid composition (Table I) of the total unprocessed translation product agree with those determined for the purified active protein. Within experimental accuracy this agreement is not compatible with the processing of exported proteins normally observed (Blobel and Dobberstein, 1975). Searching for a possible signal sequence in Figure 2, we suggest a 'signal function' for the sequence between Met_1 and Thr_{22}. The distribution of charged and hydrophobic residues in this sequence, and the locations of predicted secondary structures (data not shown), all answer the criteria deduced by Perlman and Halvorson (1983) for signal peptides. The sequence -Pro_{20}-Phe_{21}-Thr_{22}- however, is a very unlikely − if not impossible − candidate for proteolytic cleavage (Perlman and Halvorson, 1983).

This predicted absence of signal processing is in agreement with the following observations. (i) Upon dansylation of the purified protein no modified N-terminal residue could be detected. (ii) Edman degradation yielded no free PTH-amino acid derivatives. Both observations indicate a blocked NH_2 terminus. (iii) The amino acid composition agrees well with the unprocessed form of PldA protein. This is particularly prominent for the amino acids His, Met and Ile (cf. Table I). (iv) Probably the strongest evidence against processing was obtained from experiments with minicells in which protein synthesis was allowed to occur in the presence of 8% ethanol. Minicells containing a *pldA* coding plasmid under these conditions did not produce any precursor form of *pldA* protein (Figure 3). In the control experiment with pBR322 and β-lactamase, a clear accumulation of precursor form could be evoked with 8% ethanol as described by Palva *et al.* (1981).

Discussion

The nucleotide sequence of *pldA* clearly defined a 780-bp coding region for the OM phospholipase. Transcription and translation regulatory signals are present in *pldA* (Figure 2). Experimental data on regulation of *pldA*, however, have so far been lacking, but should be interesting with regard to the yet unknown function of the 'dormant', i.e. inactive in natural state, phospholipase (de Geus *et al.*, 1983).

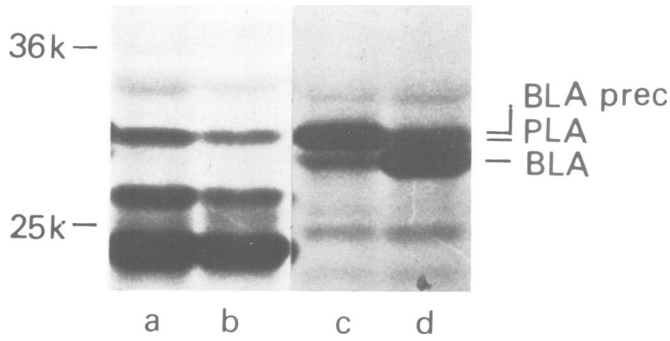

Fig. 3. Autoradiogram of SDS-polyacrylamide gel electrophoresis patterns of proteins synthesized in minicells in the presence of [^{35}S]methionine. **Lane a**, minicells with pPI 232 producing PldA protein; **lane b**, same as **lane a**, with 8% EtOH present during incubation; **lane c**, minicells with pBR322 and 8% EtOH present during incubation, producing large amounts of β-lactamase precursor (BLA-prec); **lane d**, same as **lane c** without EtOH.

The deduced amino acid sequence led us to suggest that the N-terminal peptide Met_1-Thr_{22} functions as a signal peptide, in accordance with the 'signal hypothesis' for exported proteins introduced by Blobel and Dobberstein (1975). Upon comparison of the tentative signal sequence of PldA protein with those reviewed by Perlman and Halvorson (1983), we conclude that, although the peptide shares most features of a signal sequence, no permissive signal peptidase 'target' site is present in phospholipase. It is generally agreed that signal sequences being present either as distinct N-terminal peptides or as integral sequences (Lingappa *et al.*, 1979) are a prerequisite for protein translocation across membranes. Processing of the N-terminal signal peptides is described for all native exported proteins, irrespective of the proposed mechanism of export. Only for a mutant lipoprotein has evidence been provided of no processing during or after export (Lin *et al.*, 1978).

Despite the fact that the majority of signal peptides are processed, we propose phospholipase to be an exception. This conclusion is based on (i) the molecular weight, (ii) the amino acid composition, (iii) the amino terminus, and (iv) the nature of the proteins synthesized in minicells in the presence of ethanol. The validity of arguments (i) and (ii) was examined for possible artefacts caused by over-production of the purified protein using an inducible expression vector with *pldA*. No differences were detected between phospholipase from the uninduced, i.e., natural, state and the fully induced state (P. de Geus, in preparation).

Others have previously shown that phospholipase occurs naturally in the OM (Bell *et al.*, 1971) and in fact the enzyme is often used as a marker for the OM. Co-purification of lipopolysaccharides and resistance to extraction by Triton X-100 further confirm that the protein resides in the OM. We think, therefore, that this is the first example of an unprocessed, but exported, protein. The implications of this phenomenon are, however, not yet clear.

The bacterial phospholipase releases fatty acids mainly at the SN-1 position in phospholipids, but the 2-position is also attacked. Comparison of the sequence to known sequences of phospholipases A_2 from pancreatic and snake venom sources, revealed no homology, although certain lipid-binding domains and catalytic residues might be expected to be homologous on the basis of common substrate molecules and the cleavage of identical bonds. Only the His_{133}-Asp_{134} couple in PldA protein could be related to the active site His-Asp present in all phospholipases (Dijkstra *et al.*, 1978).

Present research on the OM phospholipase of *E. coli* K-12 is focussed on the properties of the purified enzyme from over-producing strains (P.de Geus, in preparation), with respect to different substrates and regulation of enzymatic activity. The role of the 'signal' sequence in biogenesis and in enzymatic activity will be studied further using protein fusions and deletions.

Materials and methods

Materials: strains, phages, plasmids

E. coli K-12 CE1303, F$^-$, *bio*, *end*A, *sup*$^-$, *rec* A56, *pldA* (de Geus *et al.*, 1983) was used to propagate *pldA* carrying plasmids and to test new recombinant plasmids for *pldA* presence. *E. coli* K-12 JM103 was used in all experiments with M13 strains mp8 and mp9 (Messing, 1981). Plasmid pPI2336 (de Geus *et al.*, 1983) was used as the source of the *pldA* gene (Figure 1). Phospholipase phenotype determination of strain CE1303 with or without plasmids carrying *pldA* and minicell experiments were done as previously described (de Geus *et al.*, 1983).

Media

L-Broth was used in all cases to culture bacteria, with two exceptions: (i) strain JM103 was maintained on minimal salt medium with glucose to select against loss of F episome and (ii) this strain was grown on 0.004% x-gal containing Sinsheimer agar plates (10 g/l tryptone, 2.5 g/l NaCl) when single M13 phage plaques were to be isolated.

Enzymes

T4 DNA ligase, *Bal*31 nuclease, pancreatic DNase I and the Klenow fragment of *E. coli* DNA polymerase I were obtained from Boehringer Mannheim. Restriction enzymes *Bam*HI, *Cla*I, *Eco*RI, *Hind*III, *Pst*I, *Sma*I, *Eco*RV and *Pvu*II, were from New England Biolabs., and were used according to the supplier's instructions.

Biochemicals

[α-^{32}P]ATP for DNA sequence determination was purchased from New England Nuclear (10 mCi/ml, ±3000 Ci/mmol). Dideoxynucleotide triphosphates (ddNTPs) as well as all four dNTPs were from P-L Biochemicals.

Ultrathin (0.2 mm) polyacrylamide gels were poured with acrylamide from Serva, bis-acrylamide and urea from BRL. These reagents were additionally purified before use by shaking them with 5 g/100 ml of Amberlite mixed bed ion-exchange resin.

Methods: DNA techniques

Small- as well as large-scale preparation of double-stranded plasmid or phage RF DNA was done by the method of Birnboim and Doly (1979), for large-scale preparations followed by CsCl-ethidium bromide isopycnic centrifugation. Isolation of single-stranded (ss) M13 DNA was done essentially according to Heidecker *et al.* (1980).

Cloning and analysis of restriction fragments were done as previously described (de Geus *et al.*, 1983). Recombinant M13 ssDNA was analyzed on submerged horizontal 1% agarose gels in Tris-Borate buffer (Maniatis *et al.*, 1982).

Construction of overlapping deletion library

Because of lack of restriction sites in *pldA*, we subcloned the 4.0-kb *Eco*RI-*Cla*I fragment from pPI 2336 (see Figure 1), after filling in the 3′ OH recessed ends with Klenow's enzyme into the *Sma*I site of mp9, resulting in phage M17. Thereby we created a unique *Eco*RI site facing the N-terminal coding region of *pldA* and a unique *Hind*III site at the C-terminal part, as was shown with *Pst*I and *Eco*RI restriction analysis. These *Eco*RI and *Hind*III sites were used to generate a series of directional deletions with *Bal*31 nuclease, as was described essentially by Belfort *et al.* (1983). However, we omitted the use of linkers, but instead cloned the *Bal*31-generated blunt-ends directly into the *Sma*I site of mp8 (Figure 1, series II) and mp9 (Figure 1, series I).

In another line of deletion construction we used DNaseI in the presence of Mn^{2+} (Melgar and Goldthwait, 1968) to generate a set of randomly single cut M17 DNA. For this purpose 20 μg of M17 DNA was incubated with 1.5 ng freshly prepared DNase I in 200 μl of 25 mM Tris/HCl buffer of pH 7.3 with 5 mM $MnCl_2$ at 0°C. After 15 min the reaction was stopped by the addition of cold EDTA solution to a final concentration of 20 mM. The appearance of linear M17 DNA molecules was monitored on 0.6% agarose gels, after which the DNA was phenol-extracted once, EtOH-precipitated and resuspended in

500 μl *Bal*31 buffer. A subsequent 5 min incubation with 1 U of *Bal*31 nuclease was found advantageous because it transformed most of the single-strand nicked DNA (~50% of the product after DNase action) to the double strand cut form, which is the one of interest. The DNA was again phenol-extracted, cut with *Eco*RI, treated with Klenow's polymerase and ligated. With these deletions it was necessary to run preliminary single sequence tracks to discard those clones that had deleted at the wrong side of *Eco*RI, resulting in the loss of the polymerase priming site.

Significant time was gained in general by selecting recombinant ss phage DNA by insert size before actual sequencing. The agarose gel system that we used (Maniatis *et al.*, 1982) permitted convenient size determination.

DNA sequencing

The sequence of both strands of *pldA* and adjacent sequences was determined by the dideoxy chain termination method of Sanger *et al.* (1977, 1980) with ssDNA templates of appropriate clones from the deletion library. Clones were propagated in M13 vectors mp8 or mp9 (Messing *et al.*, 1980; Messing, 1981). Electrophoresis was performed with thermostated (55°C) acrylamide gels (0.2 mm) of 40 and 60 cm for 45 and 240 min, respectively, at 2400 V after which the gels were fixed in 10% acetic acid and dried at 80°C before autoradiography. Usually 350–450 nucleotides were read from the combined gels. Sequence data were processed on an Apple II computer with a version of the Larson and Messing software (1983), modified by Dr. Hans Bergmans in our institute.

Protein conformational predictions

The primary structure of PldA protein was run through a computer program to predict secondary structure, based on the method of Chou and Fasman (1978).

Amino acid analysis of proteins

Amino acid analysis of proteins was done by the method of Spackman *et al.* (1958) on a Kontron liquid chromatograph Liquimat III equipped with an automatic integrator. Protein samples were hydrolysed in 6 N HCl for 24 h at 110°C in vacuum sealed glass tubes. For tyrosine and methionine determination, 4% thioglycolic acid was included as suggested by Matsubara and Sasaki (1969). Analysis of N-terminal residues was done with the dansyl chloride procedure as described by Gray (1972). Automated Edman degradation on a Beckman sequencer, model 890 C was performed as described by Klapper *et al.* (1978).

Acknowledgements

The authors wish to thank Dr. P.Weisbeek for the use of his sequence facilities and Renske Teertstra and Carlo van den Berg for their help in performing the sequence experiments. Dr. Hans Bergmans is gratefully thanked for his assistance with the computer programs and for correcting the manuscript.

References

Belfort,M., Maley,G., Pedersen-Lane,J. and Maley,F. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 4914-4918.

Bell,R.M., Mavis,R.D., Osborn,M.J. and Vagelos,P.R. (1971) *Biochim. Biophys. Acta*, **249**, 628-635.

Birnboim,H.C. and Doly,J. (1979) *Nucleic Acids Res.*, **7**, 1513-1524.

Blobel,G. and Dobberstein,R. (1975) *J. Cell. Biol.*, **67**, 835-851.

Chou,P. and Fasman,G. (1978) *Adv. Enzymol. Relat. Areas Mol. Biol.*, **47**, 45-148.

de Geus,P., van Die,I., Bergmans,H., Tommassen,J. and de Haas,G.H. (1983) *Mol. Gen. Genet.*, **190**, 150-155.

Dijkstra,B.W., Drenth,J., Kalk,K.H. and Vandermaelen,P.J. (1978) *J. Mol. Biol.*, **124**, 53-60.

Gray,W.R. (1972) *Methods Enzymol.*, **25B**, 333-334.

Grosjean,H. and Fiers,W. (1982) *Gene*, **18**, 199-209.

Heidecker,G., Messing,J. and Gronenborn,B. (1980) *Gene*, **10**, 69-73.

Klapper,D.G., Wilde,C.E. and Capra,J.D. (1978) *Anal. Biochem.*, **85**, 126-131.

Larson,R. and Messing,J. (1983) *DNA*, **2**, 31-35.

Lin,J.J.C., Kanazawa,H., Ozols,J. and Wu,H.C. (1978) *Proc. Natl. Acad. Sci. USA*, **75**, 4891-4895.

Lingappa,V.R., Lingappa,J.R. and Blobel,G. (1979) *Nature*, **281**, 117-121.

Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning. A Laboratory Manual*, published by Cold Spring Harbor Laboratory Press, NY.

Matsubara,H. and Sasaki,R.M. (1969) *Biochem. Biophys. Res. Commun.*, **35**, 175-181.

Melgar,E. and Goldthwait,D.A. (1968) *J. Biol. Chem.*, **243**, 4409-4416.

Messing,J., Crea,R. and Seeburg,P.H. (1980) *Nucleic Acis Res.*, **9**, 309-321.

Messing,J. (1981) in Walton,E.A. (ed.), *Third Cleveland Symposium on Macromolecules: Recombinant DNA*, Elsevier, Amsterdam, pp. 143-153.

Nishijima,M., Nakaike,S., Tamori,Y. and Nojima,S. (1977) *Eur. J. Biochem.*, **73**, 115-124.

Palva,E.T., Hvist,T.R., Hardy,S.J.S., Holmgren,J. and Randall,L. (1981) *J. Bacteriol.*, **146**, 325-330.

Perlman,D. and Halvorson,H.O. (1983) *J. Mol. Biol.*, **167**, 391-409.

Rosenberg,M. and Court,D. (1979) *Annu. Rev. Genet.*, **13**, 319-353.

Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463-5467.

Sanger,F., Coulson,A.R., Barrell,B.G., Smith,A.J.H. and Roe,B.A. (1980) *J. Mol. Biol.*, **143**, 161-178.

Scandella,C.J. and Kornberg,A. (1971) *Biochemistry (Wash.)*, **10**, 4447-4456.

Shine,J. and Dalgarno,L. (1974) *Proc. Natl. Acad. Sci. USA*, **71**, 1342-1346.

Siebenlist,U., Simpson,R.B. and Gilbert,W. (1980) *Cell*, **20**, 269-281.

Spackman,D.H., Stein,W.H. and Moore,S. (1958) *Anal. Chem.*, **30**, 1190-1206.

Tinoco,I., Borer,P.N., Dangler,B. and Levine,M.D. (1973) *Nature*, **246**, 40-41.

van den Bosch,H. (1980) *Biochim. Biophys. Acta*, **604**, 191-246.

Received on 24 April 1984; revised on 28 May 1984