Cloning, sequence analysis and over-expression of the gene for the Class II fructose 1,6-bisphosphate aldolase of *Escherichia coli*

Peter R. ALEFOUNDER,* Stephen A. BALDWIN,† Richard N. PERHAM‡ and Nicholas J. SHORT§

Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge CB2 1QW, U.K.

Nucleotide sequence analysis of the *Escherichia coli* chromosomal DNA inserted in the plasmid pLC33-5 of the Clarke and Carbon library [Clarke & Carbon (1976) Cell 9, 91–99] revealed the existence of the gene, *fda*, encoding the Class II (metal-dependent) fructose 1,6-bisphosphate aldolase of *E. coli*. The primary structure of the polypeptide chain inferred from the DNA sequence of the *fda* gene comprises 359 amino acids, including the initiating methionine residue, from which an M_r of 39 146 could be calculated. This value is in good agreement with that of 40 000 estimated from sodium dodecyl sulphate–polyacrylamide gel electrophoresis of the purified dimeric enzyme. The amino acid sequence of the Class II aldolase from *E. coli* showed no homology with the known amino acid sequences of Class I (imine-forming) fructose 1,6-bisphosphate aldolases from a wide variety of sources. On the other hand, there was obvious homology with the *N*-terminal sequence of 40 residues already established for the Class II fructose 1,6-bisphosphate aldolase of *Saccharomyces cerevisiae*. These Class II aldolases, one from a prokaryote and one from a eukaryote, evidently are structurally and evolutionarily related. A 1029 bp-fragment of DNA incorporating the *fda* gene was excised from plasmid pLC33-5 by digestion with restriction endonuclease *Hae*III and subcloned into the expression plasmid pKK223-3, where the gene came under the control of the *tac* promoter. When grown in the presence of the inducer isopropyl- β -D-thiogalactopyranoside, *E. coli* JM101 cells transformed with this recombinant expression plasmid generated the Class II fructose 1,6-bisphosphate aldolase as approx. 70 % of their soluble protein. This unusually high expression of an *E. coli* gene should greatly facilitate purification of the enzyme for any future structural or mechanistic studies.

INTRODUCTION

Two forms of the enzyme fructose 1,6-bisphosphate aldolase (EC 4.1.2.13) are known, designated Class I and Class II. The Class I enzymes function by imine formation between the substrate and a catalytically essential lysine residue in the active site, which acts to stabilize the intermediate carbanion (reviewed by Morse & Horecker, 1968; Horecker *et al.*, 1972). The Class II enzymes utilize a divalent metal ion to act as the electron sink in what appears to be an otherwise similar catalytic mechanism (Mildvan *et al.*, 1971).

In other respects the two classes of aldolase are quite different (Rutter, 1964). The enzymes of eukaryotes generally fall into Class I and are tetramers (M_r approx. 160 000) of identical polypeptide chains. They occur as isoenzymes in different tissues and their amino acid sequences are homologous, especially in the region of the imine-forming lysine residues, indicating that they have arisen by divergent evolution from a common origin (Horecker *et al.*, 1972; Benfield *et al.*, 1979). The complete amino acid sequence of rabbit muscle aldolase was established some years ago (Lai *et al.*, 1974; Sajgo & Hajos, 1974; Benfield *et al.*, 1979). Nucleotide sequences of cDNAs for the rabbit muscle (Tolan *et al.*, 1984),

human liver (Besmond *et al.*, 1983; Paoletta *et al.*, 1984), rat liver (Tsutsumi *et al.*, 1984) and *Trypanosoma brucei* (Clayton, 1985) enzymes have also been determined, from which the primary structures of the proteins could be inferred. An X-ray crystallographic analysis of human muscle aldolase is in progress (Millar *et al.*, 1981).

In contrast with this wealth of information about the Class I aldolases, comparatively little is known about the Class II enzymes. The aldolase of yeast is metal-dependent, as are the aldolases of most prokaryotes studied (Rutter, 1964), but at least some prokaryotes, such as *Escherichia coli* (Stribling & Perham, 1973; Baldwin & Perham, 1978), *Lactobacillus casei* (London, 1974) and *Mycobacterium smegmatis* (Bai *et al.*, 1975), have been shown also to contain Class I enzymes. The Class II aldolases of *Saccharomyces cerevisiae* (Rutter, 1964; Jack & Harris, 1971) and *E. coli* (Baldwin *et al.*, 1978) are perhaps the best characterized. Both are dimers (M_r approx. 80 000) of identical polypeptide chains (Jack & Harris, 1971; Baldwin *et al.*, 1978). For the yeast enzyme, there are indications that thiol groups of cysteine residues may be involved in the catalytic activity (Lin *et al.*, 1971) and in the coordination of the zinc ion (Ingram, 1969). However, these proteins have proved relatively hard to purify in quantity, little is known of

Abbreviations used: IPTG, isopropyl- β -D-thiogalactopyranoside

This paper is dedicated to the memory of the late Dr. J. I. Harris (deceased 2 April 1978).

* Present address: University Chemical Laboratory, Lensfield Road, Cambridge, CB2 1EW, U.K.

† Present address: Departments of Biochemistry and Chemistry, and of Protein and Molecular Biology, Royal Free Hospital School of Medicine (University of London), Rowland Hill Street, London, NW3 2PF, U.K.

‡ To whom correspondence should be addressed.

§ Present address: Department of Biophysics, King's College London (KQC), Strand, London, WC2R 2LS, U.K.

their primary structures and no crystals have been obtained.

As part of a genetic analysis of the glycolytic enzymes of *E. coli*, it was reported (Thomson *et al.*, 1979) that a gene *pgk*, encoding the enzyme phosphoglycerate kinase, was carried on plasmid pLC33-5 of the collection of Clarke and Carbon (1976). It was also noted that *E. coli* cells transformed with this plasmid exhibited raised levels of Class II fructose 1,6-bisphosphate aldolase activity, suggesting that the same plasmid might also carry the structural gene (*fda*) for this enzyme (Thomson *et al.*, 1979). In the present paper we describe the determination of the nucleotide sequence of part of the *E. coli* DNA insert of plasmid pLC33-5 and the direct identification of the *fda* gene encoding the Class II aldolase. The complete amino acid sequence of *E. coli* Class II aldolase can be inferred from the DNA sequence, making it possible for the structural and evolutionary relationships of Class I and Class II enzymes to be explored. We go on to describe the subcloning of the *fda* gene into an expression plasmid where it comes under control of the *tac* promoter, enabling Class II aldolase activity to be raised more than 70-fold above the level found in wild-type *E. coli* cells. This provides conclusive proof of the identity of the *fda* gene and should also facilitate purification of the enzyme for any future structural or mechanistic studies.

MATERIALS AND METHODS

Enzymes and reagents

The Class II aldolase was prepared from a glycerol-growth culture of *E. coli* (Crookes strain) and assayed for catalytic activity, as described by Baldwin *et al.* (1978). Restriction endonucleases and cloning enzymes were from New England Biolabs and Boehringer-Mannheim. Isopropyl- β -D-thiogalactopyranoside (IPTG) was obtained from Sigma and components of media were from Difco Laboratories. Deoxy- and dideoxyribonucleoside 5'-triphosphates and [³⁵S] α -thio-dATP were purchased from P-L Biochemicals and Amersham International, respectively. All other reagents were of analytical grade.

Bacterial strains, bacteriophages and plasmids

E. coli JM101 (F', *sup2*, *lacI*^q) (Messing *et al.*, 1981) and bacteriophage M13mp8 were gifts from Dr. J. Messing (University of Minnesota, St. Paul, MN, U.S.A.). *E. coli* JM101 was used for all cloning and sequencing procedures. The plasmid pLC33-5 was obtained from Dr. B. J. Bachmann (Yale University, New Haven, CT, U.S.A.) and plasmid pKK223-3 was purchased from Pharmacia Fine Chemicals.

DNA manipulation and sequencing procedures

Digestion with restriction endonucleases and incubations with bacteriophage-T₄ DNA ligase and calf intestinal alkaline phosphatase were carried out according to the manufacturers' instructions. DNA was electrophoresed on agarose gels and bands were visualized by staining with ethidium bromide (Maniatis *et al.*, 1982). For DNA sequencing, the intact plasmid pLC33-5 was sonically sheared into random fragments, the fragments were submitted to agarose gel electrophoresis, and those with a mean length of about 500 bp were selectively eluted and cloned into *Sma*I-cut and phosphatase-treated bacteriophage M13mp8, all as described by Deininger

(1983). DNA sequencing was carried out by the dideoxy chain termination method (Sanger *et al.*, 1980; Biggin *et al.*, 1983) and the sequence was assembled by using the DB system of computer programs (Staden, 1986).

Expression of the *fda* gene

Following the identification and sequence analysis of the *fda* gene, the plasmid pLC33-5 was digested with the restriction enzyme *Hae*III, chosen to release the complete gene. The resulting DNA fragments were separated by means of electrophoresis on low gelling temperature agarose gels (Maniatis *et al.*, 1982) and the appropriate band (1129 bp) was excised. The agarose was melted at 70 °C and extracted three or four times with phenol followed by two extractions with water-saturated diethyl ether. The isolated DNA fragment was ligated into *Sma*I-cut and alkaline phosphatase-treated vector pKK223-3. The products of ligation were transformed into *E. coli* JM101 cells grown in 2 $\times$ TY medium (16 g of Bactotrypon + 10 g of yeast extract + 5 g of NaCl per litre) supplemented with ampicillin (200 μ g/ml).

SDS/polyacrylamide gel electrophoresis

Whole *E. coli* cells were sonicated and the soluble proteins released were analysed by means of SDS/polyacrylamide gel electrophoresis, as described elsewhere (Packman *et al.*, 1984; Scrutton *et al.*, 1987).

RESULTS

Sequence analysis of the *fda* gene

The Clarke and Carbon library of *E. coli* DNA was constructed by using poly(dA,dT) tailing to insert fragments of the *E. coli* chromosome into the *Eco*RI site of the plasmid ColE1 (Clarke & Carbon, 1976). The insert therefore cannot readily be released from the plasmid, making restriction analysis and sub-cloning the normal route to isolation of a required gene [see, for example, Greer & Perham (1986)]. A preliminary restriction analysis of the plasmid pLC33-5 (results not shown) indicated that the insert of *E. coli* DNA was about 8000 bp, compared with the ColE1 plasmid itself of 6646 bp. Given that the insert was thought to include the structural genes, *pgk* and *fda*, for two glycolytic enzymes of widespread interest (Thomson *et al.*, 1979), it was thought easier and more productive to determine the DNA sequence of the entire plasmid. The nucleotide sequence of ColE1 DNA became known during the course of this work (Chan *et al.*, 1985; a preliminary version was also kindly made available by Dr. J. Lebowitz in advance of publication) and sequences derived from ColE1 DNA could therefore easily be discarded at an early stage in the assembly of a consensus sequence for the DNA insert. It was expected that the *fda* gene could be recognized from what was already known of the structure of the *E. coli* Class II aldolase (Baldwin, 1977; Baldwin *et al.*, 1978).

In the event, this strategy proved very effective. A complete nucleotide sequence (8029 bp) was derived for the *E. coli* DNA insert in plasmid pLC33-5 and this sequence, together with a detailed analysis of all its open reading frames and potential control sequences, will be reported elsewhere (P. R. Alefounder & R. N. Perham, unpublished work). In Fig. 1 is shown the nucleotide sequence, determined in both senses of the DNA, of 1133

Fig. 1. DNA sequence of the *E. coli fda* gene and predicted amino acid sequence of Class II fructose 1,6-bisphosphate aldolase

The arrows (↓) mark the cleavage sites for the restriction enzyme *Hae*III.

bases of the DNA insert (between endonuclease *Hae*III cleavage sites) that incorporate an open reading frame thought likely to encode the Class II aldolase, together with its flanking sequences.

The identity of the putative *fda* gene was confirmed by several criteria. The gene encodes a polypeptide chain of 359 amino acid residues (including the initiating methionine) and the amino acid composition inferred from the DNA sequence (Table 1) is in good agreement (the only significant discrepancies are the values for arginine, alanine and leucine) with that of *E. coli* Class II aldolase reported by Baldwin *et al.* (1978). The calculated M_r of the polypeptide chain is 39146, also in good agreement with the value of 40000 estimated from SDS/polyacrylamide gel electrophoresis of the enzyme (Baldwin *et al.*, 1978).

Further evidence in support of the identification came from direct comparison of experimentally determined amino acid and DNA sequences. The *N*-terminal residue of *E. coli* Class II aldolase is serine (Baldwin *et al.*, 1978). Moreover, the sequence of the first 26 amino acid residues of *E. coli* Class II aldolase, kindly determined in a Beckman 890B sequencer by Dr. J. Bridgen (MRC Laboratory of Molecular Biology, Cambridge), is identical to that of the first 26 amino acids (excluding the

initiating methionine residue) inferred from the DNA sequence of the proposed *fda* gene (Fig. 1). There can be no doubt, therefore, that the correct initiation point has been selected in translating the DNA sequence. Similarly, the amino acid sequences of several tryptic peptides from *E. coli* Class II aldolase, previously determined (Baldwin, 1977), can be fitted to the DNA sequence at various points in the *fda* gene (see Fig. 1). These include the peptides Met-Ser-Lys (residues 165–167), Ile-Ser-Pro-Arg (residues 213–216), Asp-Ser-Val-Ser-Tyr-Gly-Val-Val-Lys (residues 277–285) and Ala-Gly-Gln-Thr-Ser-Met-Ile-Ala-Arg (residues 337–345). These peptides offer confirmation of the proposed DNA sequence throughout its length.

The codon usage in the *fda* gene is listed in Table 2. It has the characteristics of a strongly expressed gene (Grosjean & Fiers, 1982).

Over-expression of the *fda* gene

To facilitate future work on the enzyme, it was decided to subclone the *fda* gene into a suitable expression vector. The one chosen was pKK223-3, a pBR322-derived plasmid which carries the controllable *tac* promoter (de Boer *et al.*, 1983) and *rrnB* transcription terminators. As shown in Fig. 1, there exist cleavage sites for the

Table 1. Amino acid composition of Class II aldolase

Amino acid	Composition (mol/mol of polypeptide chain)	
	Measured*	Inferred from DNA sequence
Lys	25.6	25
His	12.1	13
Arg	8.6	6
Cys	4.2	4
Asx	39.4	39 (21 Asp, 18 Asn)
Thr	18.0	17
Ser	24.9†	25
Glx	36.7	37 (22 Glu, 15 Gln)
Pro	15.7	15
Gly	31.9	31
Ala	34.8	31
Val	29.5	30
Met	5.8	7
Ile	21.0	21
Leu	29.9	27
Tyr	12.9	13
Phe	12.9	13
Trp	3.8	4

Total number of residues = 359

Calculated M_r = 39147* Taken from Baldwin *et al.* (1978), based on a subunit M_r of 40000† N-terminal residue established by Baldwin *et al.* (1978)Table 2. Codon totals for the *E. coli* Class II fructose 1,6-bisphosphate aldolase gene

There are 360 codons in all, including initiation and termination codons.

F	TTT	2	S	TCT	11	Y	TAT	3	C	TGT	0
F	TTC	11	S	TCC	8	Y	TAC	10	C	TGC	4
L	TTA	0	S	TCA	0	*	TAA	1	*	TGA	0
L	TTG	1	S	TCG	1	*	TAG	0	W	TGG	4
L	CTT	0	P	CCT	1	H	CAT	2	R	CGT	4
L	CTC	0	P	CCC	0	H	CAC	11	R	CGC	2
L	CTA	0	P	CCA	1	Q	CAA	2	R	CGA	0
L	CTG	26	P	CCG	13	Q	CAG	13	R	CGG	0
I	ATT	1	T	ACT	7	N	AAT	0	S	AGT	0
I	ATC	20	T	ACC	10	N	AAC	18	S	AGC	5
I	ATA	0	T	ACA	0	K	AAA	19	R	AGA	0
M	ATG	8	T	ACG	0	K	AAG	6	R	AGG	0
V	GTT	14	A	GCT	12	D	GAT	8	G	GGT	23
V	GTC	1	A	GCC	2	D	GAC	13	G	GGC	8
V	GTA	12	A	GCA	9	E	GAA	19	G	GGA	0
V	GTG	3	A	GCG	8	E	GAG	3	G	GGG	0
Total: 360											

restriction enzyme *Hae*III to the 5'- and 3'-terminal sides of the *fda* gene, which would permit excision of the entire *fda* gene with only short flanking regions.

The plasmid pLC33-5 DNA was digested with restriction endonuclease *Hae*III and the products were separated by means of agarose gel electrophoresis. The appropriately-sized (1129 bp) band was excised and the DNA was extracted from the gel and ligated into *Sma*I-cut and phosphatase-treated vector pKK223-3, as

Fig. 2. Expression of the *fda* gene in *E. coli* strain JM101

Samples of cell extracts of *E. coli* strain JM101 were prepared and analysed by means of SDS/polyacrylamide-gel electrophoresis, as described in the Materials and methods section. Protein bands were stained with Coomassie Brilliant Blue. Track *a*, cell extract of *E. coli* JM101 transformed with plasmid pKfda and induced with IPTG; track *b*, cell extract of *E. coli* JM101; track *c*, cell extract of *E. coli* JM101 transformed with plasmid pKK223-3.

described in Materials and methods. The products of ligation were transformed into *E. coli* JM101 cells grown on 2×TY medium supplemented with 200 µg/ml ampicillin. Ampicillin-resistant colonies were isolated and plasmid DNA was prepared. Restriction analysis (data not shown) revealed whether the *fda* gene fragment was inserted into the vector 5'- to 3'- with respect to the *tac* promoter. The plasmid with the insert in this orientation was designated pKfda.

E. coli JM101 cells transformed with plasmid pKfda were grown up on 2×TY medium supplemented with 200 µg/ml ampicillin, in the presence and absence of the inducer IPTG (1 mM). The cells were harvested by centrifugation and sonicated to generate cell-free extracts. The extracts were assayed for Class II aldolase activity (Baldwin *et al.*, 1978). The level of aldolase activity was found to be raised about 68-fold above that of untransformed *E. coli* JM101 cells in the absence of IPTG and about 77-fold greater in the presence of the inducer. Samples of the extracts were also submitted to

SDS/polyacrylamide gel electrophoresis (Fig. 2). This revealed the existence of a highly-expressed protein, apparent M_r 40000, in the plasmid-transformed cells, which was attributed to the subunit polypeptide chain of the dimeric Class II aldolase. On the basis of the intensity of bands in Coomassie-stained gels, it was estimated that pKfda-transformed cells grown in the presence of IPTG were synthesizing approx. 70% of their soluble protein as the *fda*-gene product. These experiments offer additional proof that the product of the cloned *fda* gene is the Class II fructose 1,6-bisphosphate aldolase.

DISCUSSION

From the plasmid pLC33-5 of the Clarke & Carbon (1976) collection, we have cloned, sequenced and over-expressed the *fda* gene encoding the Class II fructose 1,6-bisphosphate aldolase of *E. coli*. The initial identification of the gene rests on direct comparison of the amino acid composition and sequence of the polypeptide chain inferred from the DNA sequence (Fig. 1) with the known properties (Table 1) and amino acid sequences of peptides of the *E. coli* enzyme (Baldwin, 1977; Baldwin *et al.*, 1978). It is abundantly supported by the high levels of Class II aldolase activity observed in *E. coli* JM101 cells transformed with the plasmid pKfda, a plasmid carrying a 1129 bp insert from the plasmid pLC33-5 incorporating the *fda* gene placed under the control of the plasmid pKK223-3 *tac* promoter (Fig. 2).

In the presence of the inducer IPTG, the *fda* gene was expressed from plasmid pKfda at levels more than 70 times greater than those from the chromosomal gene in *E. coli* JM101 cells, raising the *E. coli* Class II aldolase polypeptide chain to about 70% of the soluble protein of the transformed cells (Fig. 2). Even in the absence of the inducer IPTG, the gene was still strongly expressed, suggesting that the plasmid copy number was sufficiently high to overcome the *lac* repressor in the cells, even though *E. coli* JM101 is *lacI^a*. These values are in broad agreement with an earlier estimate (Baldwin *et al.*, 1978) that the Class II fructose 1,6-bisphosphate aldolase accounts for about 1% of the soluble protein in wild-type *E. coli* cells. *E. coli* appears to have a very high tolerance for this enzyme.

The amino acid sequence of the *E. coli* Class II fructose 1,6-bisphosphate aldolase inferred from the DNA

sequence of the *fda* gene (Fig. 1) is the first primary structure of a Class II aldolase to be established. There is no immediately apparent sequence homology with any of the Class I fructose 1,6-bisphosphate aldolases which have been widely studied (Lai *et al.*, 1974; Sajgo & Hajos, 1974; Benfield *et al.*, 1979; Besmond *et al.*, 1983; Paoletta *et al.*, 1984; Tolan *et al.*, 1984; Tsutsumi *et al.*, 1984; Clayton, 1985). Thus, as originally suspected (Rutter, 1964), the two classes of aldolase are most unlikely to have arisen by divergent evolution from a common ancestor. Likewise, there is no obvious homology with the sequences of the 2-keto-3-deoxy-6-phosphogluconate aldolase (EC 4.1.2.14) of *Pseudomonas putida* (Suzuki & Wood, 1980) or the 2-deoxy-ribose 5-phosphate aldolase (EC 4.1.2.4) of *E. coli* (Valentin-Hansen *et al.*, 1982), both of which function by the Class I imine-forming mechanism and both of which are smaller (polypeptide chain M_r approx. 30000). It would appear that these enzymes are all unrelated.

On the other hand, considerable homology is revealed when the *N*-terminal sequence of yeast aldolase (Jack, 1973; J. I. Harris, personal communication) is compared with that of *E. coli* Class II aldolase (Fig. 3). Almost half the residues in the first 40 or so amino acids are identical, with only two gaps (in the yeast enzyme sequence) being needed to maximize the homology. It is reasonable to infer that these two Class II aldolases, one from a eukaryote and the other from a prokaryote, have evolved from a common ancestor. The only other sequence available for a Class II aldolase is that of the *N*-terminal 25 amino acids of the enzyme from *Bacillus stearothermophilus* (Jack, 1973; J. I. Harris, personal communication). This sequence is also displayed in Fig. 3, from which it is evident that, with several gaps introduced, some similarity to the *E. coli* and yeast enzymes is possible. However, the homology between the two prokaryote aldolases is much less clear-cut than that between the *E. coli* and yeast enzymes and a verdict on the relationship of the *B. stearothermophilus* aldolase to the other Class II enzymes must be deferred until a more extended sequence becomes available.

We thank the Science and Engineering Research Council for financial support (Grants GR/C 13187 and GR/D 21127) and the Wellcome Trust for a grant to purchase DNA sequencing equipment. We are grateful to the Medical Research Council

Fig. 3. *N*-Terminal sequences of the Class II fructose 1,6-bisphosphate aldolases of yeast, *E. coli* and *B. stearothermophilus*

The amino acid sequence of the *E. coli* enzyme is taken from the present paper (Fig. 1). The amino acid sequences of the yeast and *B. stearothermophilus* enzymes are from Jack (1973) and J. I. Harris (personal communication). Identical residues in the yeast and *E. coli* enzymes are boxed. The numbering refers to the sequence of the yeast enzyme. Two gaps have been left in the yeast sequence, after residues 36 and 37, to maximize the homology with the *E. coli* sequence. Two gaps in the yeast and *E. coli* sequences have been left after position 27 in the yeast sequence to maximize any homology with the *B. stearothermophilus* enzyme. Gaps in the sequence of the latter enzyme have been left, after positions 5 and 11 in that sequence, for the same reason.

for the award of a Research Studentship (to S.A.B.) and the Australian National University for the award of a Postbachelor Travelling Scholarship (to N.J.S.). We are indebted to Dr. R. S. Jack for permission to quote from his Ph.D. thesis, to Dr. D. McKay for discussion of results with the *B. stearothermophilus* enzyme and to Dr. P. Lazo for his collaboration in the initial experiments.

REFERENCES

- Bai, N. J., Pai, M. R., Murthy, P. S. & Venkitesubramanian, T. A. (1975) Arch. Biochem. Biophys. **168**, 230–234
- Baldwin, S. A. (1977) Ph.D. Thesis, University of Cambridge
- Baldwin, S. A. & Perham, R. N. (1978) Biochem. J. **169**, 643–652
- Baldwin, S. A., Perham, R. N. & Stribling, D. (1978) Biochem. J. **169**, 633–641
- Benfield, P. A., Forcina, B. G., Gibbons, I. & Perham, R. N. (1979) Biochem. J. **183**, 429–444
- Besmond, C., Dreyfus, J.-C., Gregori, C., Frain, M., Zakin, M. M., Sala Trepas, J. & Kahn, A. (1983) Biochem. Biophys. Res. Commun. **117**, 601–609
- Biggin, M. D., Gibson, T. J. & Hong, G. F. (1983) Proc. Natl. Acad. Sci. U.S.A. **80**, 3963–3965
- Chan, P. T., Ohmori, H., Tomizawa, J. & Lebowitz, J. (1985) J. Biol. Chem. **260**, 8925–8935
- Clarke, L. & Carbon, J. (1976) Cell **9**, 91–99
- Clayton, C. E. (1985) EMBO J. **4**, 2997–3003
- de Boer, H. A., Comstock, L. J. & Vassar, M. (1983) Proc. Natl. Acad. Sci. U.S.A. **80**, 21–25
- Deininger, P. L. (1983) Anal. Biochem. **129**, 216–223
- Greer, S. & Perham, R. N. (1986) Biochemistry **25**, 2736–2742
- Grosjean, H. & Fiers, W. (1982) Gene **18**, 199–209
- Horecker, B. L., Tsolas, O. & Lai, C. Y. (1972) in The Enzymes (Boyer, P. D., ed.) 3rd edn., vol. 7, pp. 213–258, Academic Press, New York
- Ingram, J. M. (1969) Can. J. Biochem. **47**, 595–601
- Jack, R. S. (1973) Ph.D. Thesis, University of Cambridge
- Jack, R. S. & Harris, J. I. (1971) Biochem. J. **124**, 68P–69P
- Lai, C. Y., Nakai, N. & Chang, D. (1974) Science **183**, 1204–1206
- Lin, Y., Kobes, R. D., Norton, I. L. & Hartman, F. C. (1971) Biochem. Biophys. Res. Commun. **45**, 34–42
- London, J. (1974) J. Biol. Chem. **249**, 7977–7983
- Maniatis, T., Fritsch, E. F. & Sambrooke, J. (1982) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York
- Messing, J., Crea, R. & Seeburg, P. H. (1981) Nucleic Acids Res. **9**, 309–322
- Mildvan, A. S., Kobes, R. D. & Rutter, W. J. (1971) Biochemistry **10**, 1191–1204
- Millar, J. R., Shaw, P. J., Stammers, D. K. & Watson, H. C. (1981) Philos. Trans. R. Soc. London Ser. B **293**, 209–214
- Morse, D. E. & Horecker, B. L. (1968) Adv. Enzymol. Relat. Areas Mol. Biol. **31**, 125–181
- Packman, L. C., Hale, G. & Perham, R. N. (1984) EMBO J. **3**, 1315–1319
- Paolella, G., Santamaria, R., Izzo, P., Constanzo, P. & Salvatore, F. (1984) Nucleic Acids Res. **12**, 7401–7410
- Rutter, W. J. (1964) Fed. Proc. Fed. Am. Soc. Exp. Biol. **23**, 1248–1257
- Sajgo, M. & Hajos, G. (1974) Acta Biochim. Biophys. Acad. Sci. Hung. **9**, 239–241
- Sanger, F., Coulson, A. R., Barrell, B. G., Smith, A. J. H. & Roe, B. A. (1980) J. Mol. Biol. **143**, 161–178
- Scrutton, N. S., Berry, A. & Perham, R. N. (1987) Biochem. J. **245**, 875–880
- Staden, R. (1986) Nucl. Acids. Res. **14**, 217–231
- Stribling, D. & Perham, R. N. (1973) Biochem. J. **131**, 833–841
- Suzuki, N. & Wood, W. A. (1980) J. Biol. Chem. **255**, 3427–3435
- Thomson, J., Gerstenberger, P. D., Goldberg, D. E., Gociar, E., De Silva, A. O. & Fraenkel, D. G. (1979) J. Bacteriol. **137**, 502–506
- Tolan, D. R., Amsden, A. B., Putney, S. D., Urdea, M. S. & Penhoet, E. E. (1984) J. Biol. Chem. **259**, 1127–1131
- Tsutsumi, K., Mukai, T., Tsutsumi, R., Mori, M., Daimon, M., Tanaka, T., Yatsuki, H., Hori, K. & Ishikawa, K. (1984) J. Biol. Chem. **259**, 14572–14575
- Valentin-Hansen, P., Boëtius, F., Hammer-Jespersen, K. & Svendsen, I. (1982) Eur. J. Biochem. **125**, 561–566

Received 9 May 1988/12 September 1988; accepted 16 September 1988