*Gene*, 35 (1985) 217–222
Elsevier



GENE 1259

**Short Communication**

# Nucleotide sequence of the dihydrofolate reductase gene of methotrexate-resistant *Lactobacillus casei*

(Recombinant DNA; dideoxy sequencing; S1 mapping; *L. casei* promoter; codon usage)

**J. Andrews[a], G.M. Clore[b†], R.W. Davies[a†*], A.M. Gronenborn[b†*], B. Gronenborn[c†], D. Kalderon[b], P.C. Papadopoulos[b], S. Schäfer[c†], P.F.G. Sims[a], and R. Stancombe[a]**

[a]*University of Manchester Institute of Science and Technology, Department of Biochemistry and Applied Molecular Biology, P.O. Box 88, Manchester M60 10D (U.K.), Tel. 061 236 3311;* [b]*National Institute for Medical Research, Division of Biochemistry and Physical Biochemistry, The Ridgeway, Mill Hill, London NW7 1AA (U.K.), Tel. 01 959 3666, and* [c]*Institut für Genetik, Universität, Weyertal 121, D-5000 Köln 41 (F.R.G.) Tel. 0221 470 3407*

(Received December 23rd, 1984)
(Accepted January 4th, 1985)

SUMMARY

The nucleotide sequence of the dihydrofolate reductase (DHFR) gene of a methotrexate-resistant strain of *Lactobacillus casei*, which is the source of DHFR for nuclear magnetic resonance (NMR) studies, has been determined. The derived amino acid sequence differs from that obtained by protein sequencing by the presence of aspartic acid instead of asparagine at position 8 and proline instead of leucine at position 90. The nucleotide sequences of 320-bp 5′ and 335-bp 3′ flanking regions of this gene have also been determined.

* Authors to whom reprint requests and correspondence should be addressed.

† Present addresses: (R.W.D.) Allelix Inc., 6850 Goreway Drive, Mississauga, Ontario, (L4V 1P1 Canada) Tel. (416) 677-0831; (B.G.; S.S.) Max-Planck-Institut für Züchtungsforschung, Im Engelspfad, D-5000 Köln 30 (F.R.G.) Tel. 0221 5062 385; (G.M.C.; A.M.G.) Max-Planck-Institut für Biochemie, D-8033 Martinsried bei München (F.R.G.) Tel. 049 898 5782673.

Abbreviations: bp, base pair(s); DHFR, dihydrofolate reductase; TrMP, trimethoprim; MTX, methotrexate; NMR, nuclear magnetic resonance; ORF, open reading frame; $^{R}$, resistance; RBS, ribosome-binding site; $^{S}$, sensitivity; Tc, tetracycline.



## INTRODUCTION

DHFR plays a central role in one carbon unit metabolism (Blakely, 1967), and is the target of a number of clinically important drugs, such as trimethoprim and MTX. This enzyme has therefore been the subject of extensive physicochemical studies aimed at understanding the structure and function of the active site of the enzyme. Crystal structures of complexes of the *Escherichia coli* DHFR with MTX (Matthews et al., 1977; 1978) and TrMP (Baker et al., 1981), and of a ternary complex of the *L. casei* DHFR with MTX and NADPH, have been obtained, and *L. casei* DHFR has been studied

extensively by NMR (Birdsall et al., 1977; Feeney et al., 1977; Wyeth et al., 1980). Recent advances in recombinant DNA technology and chemical synthesis of oligodeoxynucleotides have made it possible to use oligonucleotide-directed in vitro mutagenesis to introduce specific amino acid exchanges into protein molecules, and several interesting alterations have recently been made in *E. coli* DHFR (Villafranca et al., 1983). A prerequisite to such studies is the accurate determination of the nucleotide sequence coding for the protein, and this is reported here for the *L. casei* enzyme.

The DHFR gene that we have sequenced was cloned from an $MTX^R$ strain of *L. casei* by Davies and Gronenborn (1982). It is not known whether the amino acid sequence of DHFR from $MTX^S$ *L. casei* is the same or different. The $MTX^R$ strain was chosen because the DHFR of this strain is the protein on which extensive NMR work has been carried out. The cloned sequence is of further interest since it has been shown to contain a region or regions to which *L. casei* DHFR binds specifically (Gronenborn and Davies, 1981; Clore et al., 1982; Gronenborn and Clore, 1983). The location of the DHFR gene within the cloned DNA is shown in Fig. 1. The plasmid pWDLcB1 contains two contiguous *L. casei* *Bam*HI fragments inserted into the *Bam*HI site in the $Tc^R$ gene of pBR322. The larger *L. casei* *Bam*HI fragment (1.8 kb) contains the DHFR gene between the *Eco*RI and *Pst*I sites (Fig. 1), and this is the region which has been sequenced.

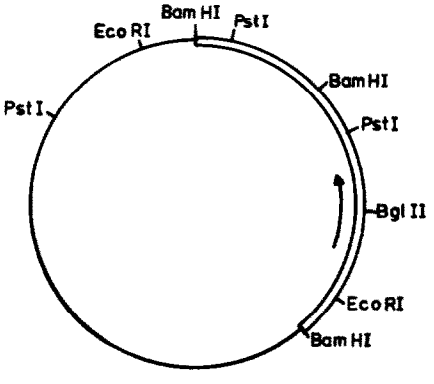


Fig. 1. Restriction maps and location of the *L. casei* DHFR gene in plasmid pWDLcB1. The single line represents sequences derived from plasmid pBR322, whilst the double line corresponds to the inserted *L. casei* DNA. The location and orientation of the ORF corresponding to the gene for *L. casei* DHFR are shown by the arrow.

## EXPERIMENTAL AND DISCUSSION

### (a) Nucleotide sequencing

The nucleotide sequence was determined using primarily the chain-termination method of Sanger et al. (1976) as described by Davies (1982). All of the sequence was determined in both directions, and some of it was determined by the method of Maxam and Gilbert (1980) as well. Furthermore, a second independent clone was obtained and sequenced, and the nucleotide sequence was found to be identical to that of pWDLcB1. The sequence comprising the DHFR gene and the 320-bp 5′ and 335-bp 3′ flanking regions is shown in Fig. 2.

### (b) Coding

The DHFR gene, the first sequenced gene from *L. casei*, has standard initiation (ATG) and termination (TAA) codons. The initiation codon is separated by 8 bp from a sequence that would be a likely RBS in *E. coli* (Shine and Dalgarno, 1974). Thus the translation signals for DHFR in this species closely resemble those of *E. coli*. Since the amino acid sequence of *L. casei* DHFR has been determined (Freisheim et al., 1978), the ORF and codon assignments can be made. Codon usage (Table I) is very

TABLE I

Codon usage for the DHFR gene of *L. casei*

| Amino acid | Codon | Frequency of use[a] |
|---|---|---|
| Ala | GCA | 3 |
| | U | 6 |
| | C | 1 |
| | G | 5 |
| Arg | AGA | 0 |
| | G | 0 |
| | CGA | 1 |
| | U | 2 |
| | C | 3 |
| | G | 2 |
| Asn | AAU | 2 |
| | C | 1 |
| Asp | GAU | 14 |
| | C | 1 |
| Cys | UGU | 0 |
| | C | 0 |

| Amino acid | Codon | Frequency of use[a] |
|---|---|---|
| Gln | CAA | 3 |
| | G | 6 |
| Glu | GAA | 5 |
| | G | 2 |
| Gly | GGA | 0 |
| | U | 5 |
| | C | 5 |
| | G | 0 |
| His | CAU | 6 |
| | C | 1 |
| Ile | AUU | 3 |
| | C | 2 |
| | A | 0 |
| Leu | UUA | 7 |
| | G | 3 |
| | CUA | 0 |
| | U | 0 |
| | C | 0 |
| | G | 3 |
| Lys | AAA | 5 |
| | G | 4 |
| Met | AUG | 3 |
| Phe | UUU | 7 |
| | C | 1 |
| Pro | CCA | 2 |
| | U | 3 |
| | C | 1 |
| | G | 2 |
| Ser | UCA | 0 |
| | U | 0 |
| | C | 1 |
| | G | 0 |
| | AGU | 2 |
| | C | 1 |
| Thr | ACA | 2 |
| | U | 1 |
| | C | 7 |
| | G | 4 |
| Trp | UGG | 4 |
| Tyr | UAU | 5 |
| | C | 0 |
| Val | GUA | 1 |
| | U | 8 |
| | C | 5 |
| | G | 2 |

[a] Numbers indicate the number of occurrences of a particular codon in the DHFR gene of *L. casei*.

similar to that in *E. coli* with the single exception that all five tyrosine codons are UAU, which is usually the rare codon in *E. coli* and related species (Ikemura, 1982).

The amino acid sequence derived from the nucleotide sequence differs from the amino acid sequence published by Freisheim et al. (1978) in two places; Asp for Asn at position 8 and Pro for Leu at position 90. The residue at position 8 has already been proposed to be Asp, not Asn, by Batley and Morris (1977) on the basis of a comparison of the N-terminal part of the protein purified from a different $MTX^R$ strain of *L. casei*. We have checked the nucleotide sequences in these regions very carefully, and they are completely unambiguous. Thus the finding of a proline at position 90 may either be due to a leucine having been identified by error, or to a genuine difference in the protein produced in these two strains. The latter possibility seems less likely, in our view, since a comparison of the amino acid sequence of the three bacterial enzymes from *E. coli*, *L. casei* and *Streptococcus faecium* implies a strong conservation of amino acid sequences and both the *E. coli* and *S. faecium* proteins possess a proline in this position. Furthermore, inspection of the crystal structure of the *L. casei* DHFR complex places amino acid 90 at the outside of the protein structure in the region of a turn similar in both the *E. coli* and *L. casei* structures. It is very likely that in both enzymes the proline residue in position 90 is involved in such a turn. Therefore, residue 90 is probably proline, and the assignment of leucine to this position was a mistake.

### (c) Initiation of transcription

There is no clearly recognizable sequence resembling a promoter, of the type found in *E. coli*, close to the putative RBS for the *L. casei* DHFR gene. However, a sequence with almost consensus −35 and −10 (Rosenberg and Court, 1979) elements of *E. coli* promoters occurs 189 bp upstream of the start codon. Transcript mapping experiments (Fig. 3) (Berk and Sharp, 1977) show that this is indeed the sequence used as a promoter by *E. coli* RNA polymerase.

The 625 bp *Eco*RI-*Bgl*II fragment (Fig. 1) gave rise to a 455-bp DNA-RNA hybrid fragment, which places the 5′ end of the mRNA at position 174 ± 2 in

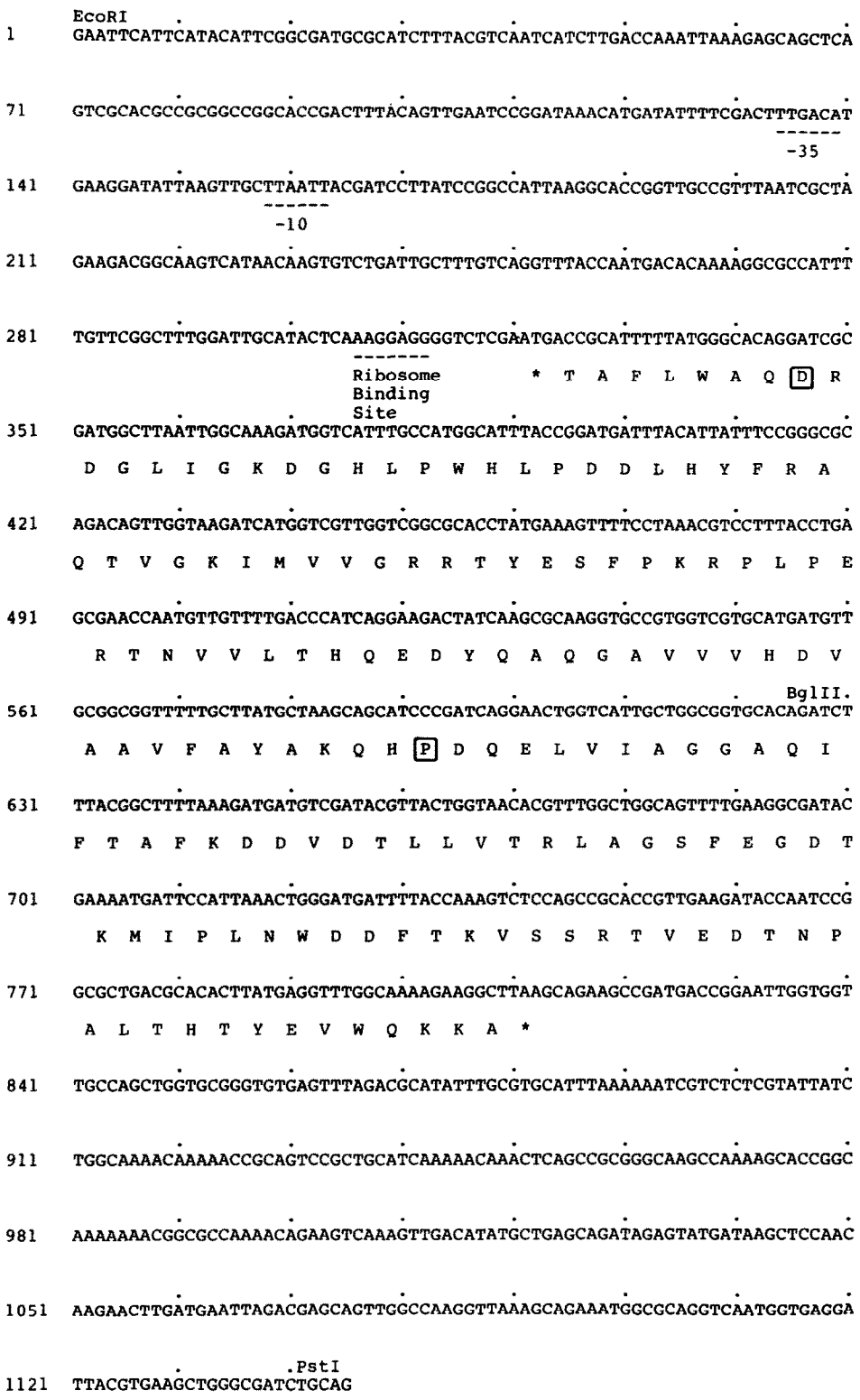


Fig. 2. Nucleotide sequence of the *L. casei* DHFR gene from an MTX$^R$ strain of *L. casei*. The nucleotide sequence of the *Eco*RI-*Pst*I fragment from pWDLcB1 is shown. The derived amino acid sequence of the coding region is shown, and the two residues that differ between this and the previously published (Freisheim et al., 1978) amino acid sequence are boxed. Also indicated by dashed underlining are the RBS and the −10 and −35 components of the promoter that resembles analogous sites used by *E. coli* RNA polymerase. Asterisks represent start and stop codons; the methionine of the start codon is not found in the mature DHFR protein.

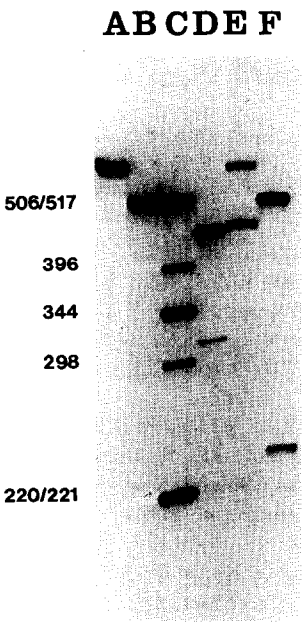


Fig. 3. Autoradiogram of an S1 mapping experiment of the transcript of the *L. casei* DHFR gene produced by *E. coli*. Total mRNA from a 100-ml culture of pWDLcB1 containing *E. coli* was prepared according to the method of Ikemura and Dahlberg (1973). DNA fragments for S1 mapping were prepared from purified plasmid DNA by digestion with restriction endonucleases and preparative agarose gel electrophoresis. The 625-bp *Eco*RI-*Bgl*II fragment covering the protein start was labelled at the 5′ end of its noncoding strand by phosphatase treatment followed by treatment with T4 polynucleotide kinase and [$\gamma$-$^{32}$P]ATP. The 520-bp *Bgl*II-*Pst*I fragment, covering the protein stop signal, was labelled at the 3′ end of its noncoding strand by filling in the recessed 3′ end with the large fragment of DNA polymerase I (Klenow) using [$\alpha$-$^{32}$P]dGTP. S1 mapping was carried out as follows: 300 ng of labelled DNA (520 bp or 625 bp fragment) was dissolved with 30 $\mu$g of total mRNA in 10 $\mu$l of 40 mM PIPES (pH 6.4), 1 mM EDTA (pH 8.0), 400 mM NaCl, 80% formamide, and the mixture was shaken for 30 min. The tube was incubated at 90°C for 5 min and then immediately transferred to a 50°C waterbath and left overnight. S1 nuclease was added (300 units) at 50°C in 300 $\mu$l of 280 mM NaCl, 30 mM Na · acetate (pH 4.4), 4.5 mM Zn · acetate and the tube was transferred to 20°C for 2 h. The reaction was stopped with 6 $\mu$l 0.5 mM EDTA and the solution phenol extracted and ethanol precipitated. The resulting DNA was analyzed on a 6% polyacrylamide gel containing 8 M urea with appropriate markers. The gel was autoradiographed for 2 days using a preflashed Fuji RX X-ray film with an intensifying screen. Lane A, 625-bp *Eco*RI-*Bgl*II fragment; lane B, 520-bp *Bgl*II-*Pst*I fragment; lane C, *Hin*fI digest of pBR322 with the fragment sizes specified on the left margin; lane D, marker fragments, 441-bp *Bgl*I-*Bgl*II fragment, partially cut with *Taq*I (minor fragment is 308 bp long); lane E, S1 mapping of 625-bp fragment; lane F, S1 mapping of 520-bp fragment.

the sequence in Fig. 2, and thus possibly at A173. This site is 9 bp downstream from the *E. coli*-type promoter sequences. The 520-bp *Bgl*II-*Pst*I fragment (Fig. 1) gave rise to a major band corresponding to a 242-bp DNA-RNA hybrid fragment after hybridization to mRNA followed by S1 digestion. This places the 3′ end of the mRNA at position 866 ± 2 bp in Fig. 2, which is 54 ± 2 bp downstream from the termination codon of the DHFR gene. While it is possible that *L. casei* promoters are completely different and therefore unrecognizable structures, it does seem statistically very improbable that both −35 and −10 elements with correct spacing should occur in the 320 bp upstream of the *L. casei* DHFR gene by chance alone. There is, therefore, a strong possibility that *L. casei* promoters resemble *E. coli* promoters closely, and this may also apply to transcription termination signals in the two organisms. There are several precedents in *E. coli* for promoters being located considerable distances 5′ to genes, including the case of the *crp* gene promoter, which is under autoregulation by the gene product, CRP (Aiba, 1983). It is interesting that *L. casei* DHFR may also regulate its own synthesis (Gronenborn and Davies, 1981). The precise sequence elements involved in this remain to be determined.

ACKNOWLEDGEMENTS

This work was supported by a Medical Research Council project grant (to R.W.D.), by the Medical Research Council through the National Institute for Medical Research (to A.M.G. and G.M.C.), by the Lister Institute for Preventive Medicine (to G.M.C.), and by the Deutsche Forschungsgemeinschaft through SFB 74 (to B.G.).

REFERENCES

Aiba, H.: Autoregulation of the *E. coli crp* gene: CRP is a transcriptional repressor of its own gene. Cell 32 (1983) 141–149.

Baker, D.J., Beddell, C.R., Champness, T.N., Goodford, D.J., Norrington, F.E.H., Smith, D.R. and Stammes, D.K.: The binding of trimethoprim to bacterial dihydrofolate reductase. FEBS Lett. 126 (1981) 49–52.

Batley, K.E. and Morris, H.R.: Dihydrofolate reductase from *Lactobacillus casei*: N-terminal sequence and comparison with the substrate binding region of other reductases. Biochem. Biophys. Res. Commun. 75 (1977) 1010–1014.

Berk, A.J. and Sharp, P.A.: Sizing and mapping of early adenovirus messenger RNA by gel electrophoresis of S1 endonuclease digested hybrids. Cell 12 (1977) 721–732.

Birdsall, B., Griffiths, D.V., Roberts, G.C.K., Feeney, J. and Burgen, A.S.V.: $^1$H nuclear magnetic resonance studies of *Lactobacillus casei* dihydrofolate reductase: effects of substrate and inhibitor binding on the histidine residues. Proc. R. Soc. Lond. B. 196 (1977) 251–265.

Blakely, R.L.: The Biochemistry of Folic Acid and Related Pteridines. Elsevier, Amsterdam, 1967.

Clore, G.M., Gronenborn, A.M. and Davies, R.W.: Theoretical aspects of specific and non-specific equilibrium binding of proteins to DNA as studied by the nitrocellulose filter binding assay. Cooperative and non-cooperative binding to a one-dimensional lattice. J. Mol. Biol. 155 (1982) 447–466.

Davies, R.W.: DNA sequencing, in Rickwood, D. and Hames, B.D. (Eds.), Gel Electrophoresis of Nucleic Acids: A Practical Approach. IRL Press, Oxford, 1982, pp. 117–172.

Davies, R.W. and Gronenborn, A.M.: Molecular cloning of the gene for dihydrofolate reductase from *Lactobacillus casei*. Gene 17 (1982) 229–233.

Feeney, J., Roberts, G.C.K., Birdsall, B., Griffiths, D.V., King, R., Scudder, P. and Burgen, A.S.V.: $^1$H nuclear magnetic resonance studies of the tyrosine residues of selectively deuterated *Lactobacillus casei* dihydrofolate reductase. Proc. Roy. Soc. Lond. B 196 (1977) 267–290.

Freisheim, J.M., Bitan, K.G., Reddy, A.G. and Blankenship, D.T.: Dihydrofolate reductase from amethopterin-resistant *Lactobacillus casei*. J. Biol. Chem. 253 (1978) 6437–6444.

Gronenborn, A.M. and Clore, G.M.: Characterization of the DNA binding region recognized by dihydrofolate reductase from *Lactobacillus casei*. J. Biol. Chem. 258 (1983) 11256–11259.

Gronenborn, A.M. and Davies, R.W.: DNA binding by dihydrofolate reductase from *Lactobacillus casei*. J. Biol. Chem. 256 (1981) 12152–12155.

Ikemura, T.: Correlation between the abundance of yeast transfer RNA, and the occurrence of the respective codons in protein genes: differences in synonymous codon choice patterns of yeast and *Escherichia coli* with reference to the abundance of coaccepting transfer RNAs. J. Mol. Biol. 158 (1982) 573–597.

Ikemura, T. and Dahlberg, J.E.: Small RNA of *Escherichia coli*. Part 1. Characterization by polyacrylamide gel electrophoresis and fingerprint analysis. J. Biol. Chem. 248 (1973) 5024–5032.

Matthews, D.A., Alden, R.A., Bolin, J.T., Filman, D.J., Freer, S.T., Hamlin, R., Hol, W.G.J., Kisliuk, R.L., Pastore, E.J., Plante, L.T., Xuong, N. and Kraut, J.: Dihydrofolate reductase from *Lactobacillus casei*: X-ray structure of the enzyme-methotrexate-NADPH complex. J. Biol. Chem. 253 (1978) 6946–6954.

Matthews, D.A., Alden, R.A., Bolin, J.T., Freer, S.T., Hamlin, R., Xuong, N., Kraut, J., Poe, M., Williams, M. and Hoogsteen, K.: Dihydrofolate reductase: X-ray structure of the binary complex with methotrexate. Science 197 (1977) 452–455.

Maxam, A. and Gilbert, W.: Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. 65 (1980) 499–560.

Rosenberg, M. and Court, D.: Regulatory sequences involved in the promotion and termination of RNA transcription. Annu. Rev. Genet. 13 (1979) 319–353.

Sanger, F., Nicklen, S. and Coulson, A.R.: DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74 (1976) 5463–5467.

Shine, T. and Dalgarno, L.: The 3′-terminal sequence of *E. coli* 16S ribosomal RNA: Complementarity to nonsense triplets and ribosome binding sites. Proc. Natl. Acad. Sci. USA 71 (1974) 1342–1346.

Villafranca, J.E., Howell, E.E., Voet, D.H., Strobel, M.S., Ogden, R.C., Abelson, J.N. and Kraut, J.: Directed mutagenesis of dihydrofolate reductase. Science 222 (1983) 782–788.

Wyeth, P., Gronenborn, A.M., Birdsall, B., Roberts, G.C.K., Feeney, J. and Burgen, A.S.V.: Histidine residues of *Lactobacillus casei* dihydrofolate reductase: paramagnetic relaxation and deuterium-exchange studies and partial assignments. Biochemistry 19 (1980) 2608–2615.

Communicated by A.J. Podhajska.