The EMBO Journal vol.5 no.11 pp.3013–3019, 1986

Evolution of biosynthetic pathways: a common ancestor for threonine synthase, threonine dehydratase and D-serine dehydratase

Claude Parsot

Unité de Biochimie Cellulaire, Département de Biochimie et Génétique Moléculaire, Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15, France

Communicated by G.N.Cohen

The *Bacillus subtilis* genes encoding threonine synthase (*thrC*) and homoserine kinase (*thrB*) have been cloned via complementation of *Escherichia coli thr* mutants. Determination of their nucleotide sequences indicates that the *thrC* stop codon overlaps the *thrB* start codon; this genetic organization suggests that the two genes belong to the same operon, as in *E. coli*. However, the gene order is *thrC–thrB* in *B. subtilis* whereas it is *thrB–thrC* in the *thr* operon of *E. coli*. This inversion of the *thrC* and *thrB* genes between *E. coli* and *B. subtilis* is indicative of a possible independent construction of the *thr* operon in these two organisms. In other respects, comparison of the predicted amino acid sequences of the *B. subtilis* and *E. coli* threonine synthases with that of *Saccharomyces cerevisiae* threonine dehydratase and that of *E. coli* D-serine dehydratase revealed extensive homologies between these pyridoxal phosphate-dependent enzymes. This sequence homology, which correlates with similarities in the catalytic mechanisms of these enzymes, indicates that these proteins, catalyzing different reactions in different metabolic pathways, may have evolved from a common ancestor.

Key words: biosynthetic pathways/evolution/isoleucine/operon/pyridoxal phosphate enzyme

Introduction

In prokaryotes, aspartate is the precursor of numerous amino acids, namely lysine, methionine, threonine and isoleucine, as well as diaminopimelate, a component of the cell wall in Gram-negative bacteria and dipicolinate which is a component of the spore cortex of the Gram-positive bacteria. The threonine biosynthetic pathway itself is composed of five steps and produces the branch-point metabolites, aspartate semialdehyde which leads to diaminopimelate and lysine (and dipicolinate), homoserine which leads to methionine, and threonine which is the precursor of isoleucine (Figure 1).

In *E. coli*, four of the five steps involved in the threonine biosynthesis are catalyzed by enzymes encoded by genes belonging to the same operon: aspartokinase I-homoserine dehydrogenase I (EC 2.7.2.4) – (EC 1.1.1.3), homoserine kinase (EC 2.7.1.39) and threonine synthase (EC 4.2.99.2), encoded by the *thrA*, *thrB* and *thrC* genes respectively (Thèze and Saint-Girons, 1974). The expression of the *thr* operon is multivalently controlled by threonine and isoleucine via both attenuation (Gardner, 1979) and repression (Johnson and Somerville, 1983).

In the Gram-positive organism *Bacillus subtilis*, three genes involved in threonine biosynthesis are also closely linked on the chromosome: *hom*, *thrB* and *thrA*, coding for homoserine dehydrogenase, threonine synthase and homoserine kinase respectively (Skarstedt and Greer, 1973). Although classical genetic experiments did not establish the direction of transcription in this gene cluster, the gene order was apparently different from the *E. coli thr* operon. Another intriguing feature of the *B. subtilis thr* genes was that derepression of threonine synthase expression suppressed mutations affecting threonine dehydratase (EC 4.2.1.16) activity (leading to isoleucine auxotrophy), i.e. allowed growth on homoserine or threonine as well as isoleucine (Vapnek and Greer, 1971).

Here we report the cloning and the nucleotide sequence determination of the *B. subtilis* genes encoding threonine synthase and homoserine kinase, and present evidence for a *thr* operon in *B. subtilis*. The operon comprises at least these two genes and has an organization different from that of the *E. coli thr* operon. Extensive homologies were detected in the amino acid sequences of threonine synthase and threonine dehydratase, which catalyze consecutive steps in the isoleucine biosynthetic pathway (see Figure 1). This comparison was extended to the D-serine dehydratase amino acid sequence, which also seems to have evolved from the same common ancestor. This allowed us to predict the location of the binding site for the common co-factor, pyridoxal phosphate, in threonine synthase and threonine dehydratase. These sequence homologies are discussed in terms of construction and evolution of biosynthetic pathways.

Results and Discussion

The gene encoding homoserine kinase in *E. coli* is named *thrB*, whereas the *B. subtilis* gene has hitherto been called *thrA*. Similarly the gene encoding threonine synthase is named *thrC* in *E. coli* and *thrB* in *B. subtilis*. Since this could be an important source of misunderstanding, I propose renaming the *B. subtilis* genes in line with those of *E. coli*; I shall thus refer to *thrB* and *thrC* as the *B. subtilis* genes encoding homoserine kinase and threonine synthase respectively.

Cloning the thrB and thrC genes of B. subtilis

From a plasmid library constructed by inserting randomly sheared *B. subtilis* 168 DNA into the *Bam*HI site of the *B. subtilis/E. coli* shuttle vector pHV33 (Primrose and Ehrlich, 1981), Rapoport *et al.* (1979) previously obtained a recombinant plasmid, pBS02A, which was able to complement the *B. subtilis thr5* mutation. I used pBS02A to transform several *E. coli* strains carrying different *thr* mutations (*thrB1004*, *thrC1010* or *thrC1080*), and ampicillin resistant transformants were replica-plated to test their Thr phenotype. Complementation was obtained for the *thrB* but not for the *thrC* alleles, indicating that pBS02A carried and expressed the *B. subtilis thrB* gene, but not *thrC*.

Digestion of pBS02A with restriction endonuclease *Hin*dIII and religation with T4 DNA ligase produced pSU2, lacking the pC194 part of the pHV33 vector (Figure 2). The lack of complementation of the *E. coli thrB* mutant by pSU2 suggested that the promoter used to transcribe the *B. subtilis thrB* gene from pBS02A was in the pC194 part of the pHV33 vector (Figure 2). Subcloning of the 1.7-kb *Sal*I DNA fragment of pSU2 into plasmid vec-

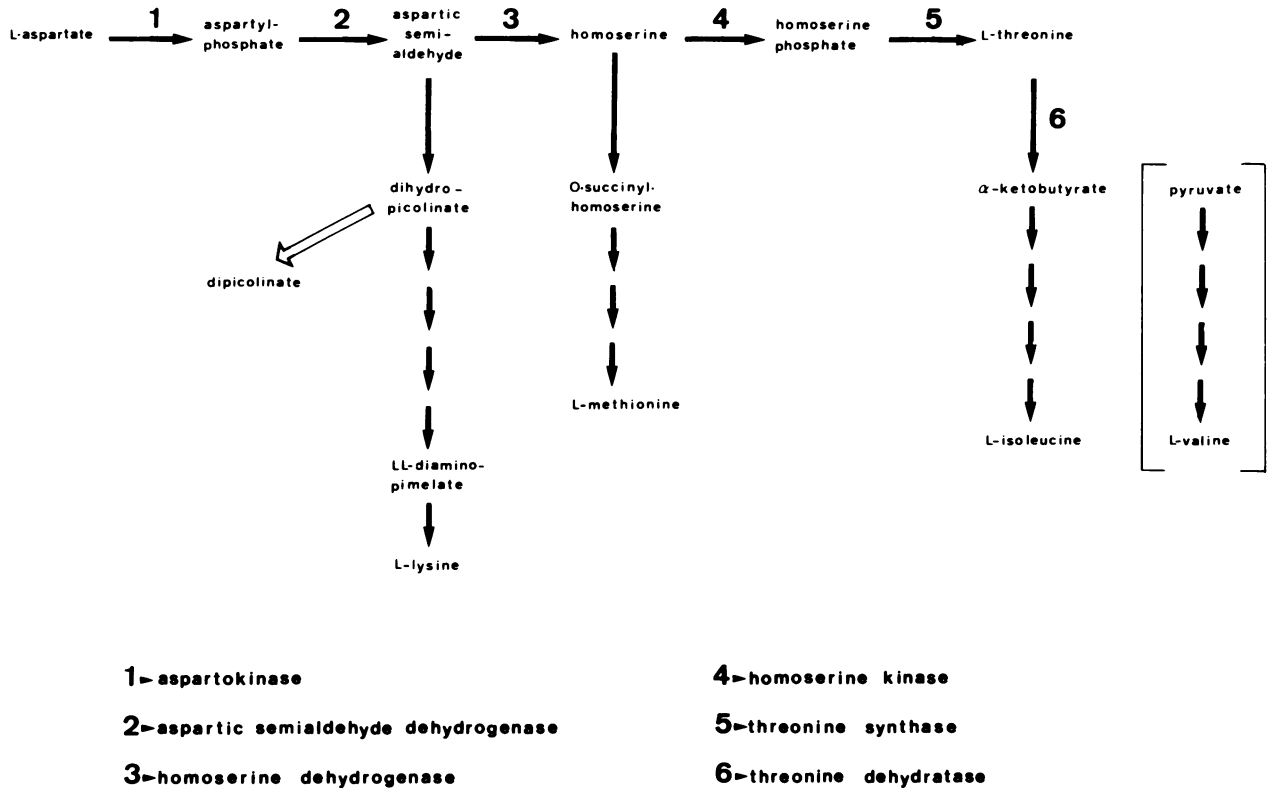

Fig. 1. The threonine biosynthetic pathway and its derivatives. The biosynthetic pathways leading to lysine, methionine and isoleucine are schematically represented from the branch-point metabolites of the threonine pathway presented on the top line. Numbers above the arrows refer to the enzymes, the names of which are indicated on the bottom part of the figure. The four steps of the valine biosynthetic pathway which are catalyzed by the same four enzymes involved in the isoleucine pathway are indicated in brackets. The open arrow refers to the synthesis of dipicolinate which takes place during the sporulation of the Gram-positive organisms.

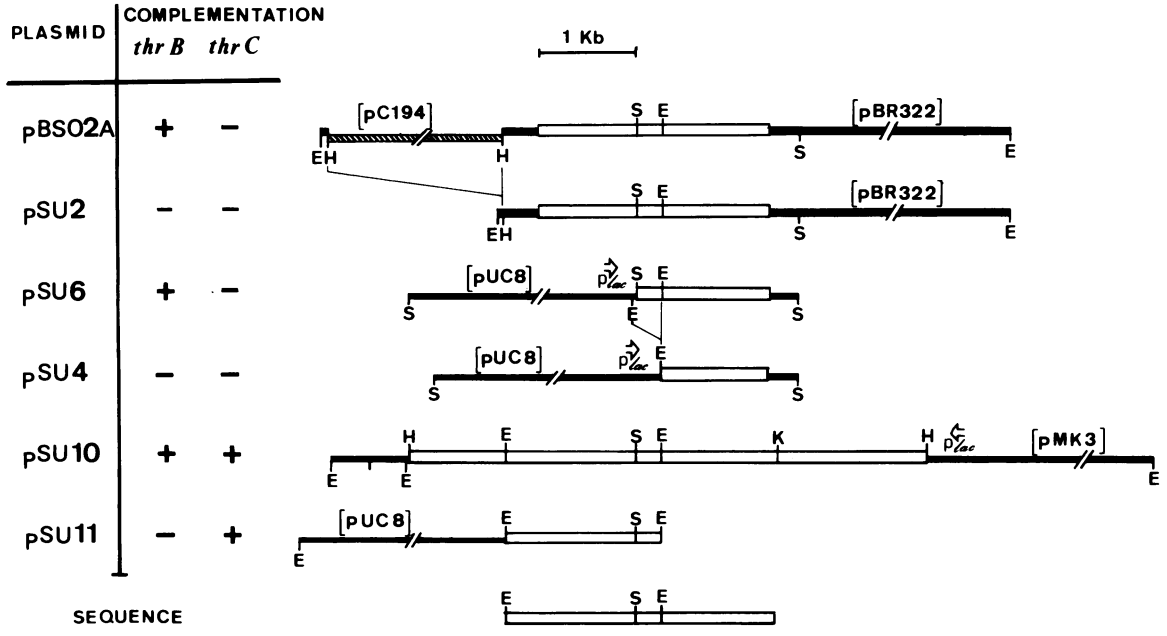

Fig. 2. Schematic representation of the *thr* plasmids. The plasmids are represented in a linear form with relevant restriction endonuclease sites being indicated above or under the lines for the insert (open bar) or the vector DNA (bold or dashed bars) respectively; symbols used for restriction sites are: E, *Eco*RI; H, *Hin*dIII; K, *Kpn*I; S, *Sal*I. Transcription from the *lac* promoter is indicated by an arrow. Complementation activities conferred by each plasmid in *E. coli thrB* or *thrC* strains are indicated in front of the plasmid names. The position of the fragment, the sequence of which is presented in Figure 3, is indicated on the bottom line.

tor pUC8 (Vieira and Messing, 1982) gave rise to pSU6 (Figure 2) which complemented the *thrB* mutation in *E. coli*, probably by transcription initiated at the vector *lac* promoter.

To clone *thrC*, which is closely linked to *thrB* on the *B. subtilis* chromosome (Skarstedt and Greer, 1973), hybridization analysis was performed to determine the restriction map of the *B. subtilis* chromosomal DNA around *thrB*. This led to the identification of a 5.3-kb-long *Hin*dIII chromosomal DNA fragment spanning the insert cloned in pBS02A. *B. subtilis* DNA was thus digested with *Hin*dIII and 5- to 7-kb DNA fragments were purified and inserted into the *Hin*dIII site of the *B. subtilis*/*E. coli* shuttle vector pMK3 (Sullivan *et al.*, 1984). Ampicillin resistant transformants of *E. coli* strain GT869 (*thrB1004*) were replica-plated to select the Thr^+ recombinants, from one which pSU10 was isolated (Figure 2). Complementation of the *thrC* mutations in *E. coli* strains GT28 (*thrC1010*) and GT121 (*thrC1080*) by pSU10 indicated the presence of the *B. subtilis thrC* gene in the insert of pSU10. The *B. subtilis thrC* gène was subsequently localized to a 1.6-kb *Eco*RI DNA fragment (see plasmid pSU11 in Figure 2).

Nucleotide sequence of the thrC and thrB genes

The nucleotide sequence of the pSU10 region encompassing the inserts carried by pSU11 and pSU6 is presented in Figure 3 along with the deduced amino acid sequence of the *thrC* and *thrB* gene products.

The identification of the *thrC* coding region is based on the presence of an open reading frame from nucleotide 188 to nucleotide 1303 in Figure 3, entirely included in the insert of pSU11 which complemented *E. coli thrC* mutations. The first potential start codon in this open reading frame is that ATG at position 248 which is preceded, three nucleotides upstream, by the sequence GGAG that could be used as a ribosome binding site for the translation of the *thrC* gene (Shine and Dalgarno, 1974). The *thrC* gene could thus code for a 351 amino acid long protein with a deduced molecular weight of 37 421.

The identification of the *thrB* coding region is based on the presence of an open reading frame from nucleotide 1279 to nucleotide 2226 (Figure 3), entirely included in the pSU6 insert which complemented an *E. coli thrB* mutant. In addition, deletion of the fragment upstream the *Eco*RI site at position 1466 (plasmid pSU4 in Figure 2) abolished this complementing activity. The first ATG codon in this open reading frame is at position 1303 and is preceded, 10 nucleotides upstream, by the sequence AAAGGAG which could be a ribosome binding site. This ATG codon overlaps the TGA stop codon of *thrC*. The *thrB* gene could thus code for a 298 amino acid long protein with a deduced molecular weight of 28 971.

Downstream from the *thrB* gene there is another reading frame open up to the extremity of the sequenced fragment (the 56 codon long ORF Y in Figure 3). This open reading frame has an ATG codon at position 2233 (three nucleotides downstream from the *thrB* stop codon) but it is not preceded by an obvious ribosome binding site. Whether or not ORF Y could represent the beginning of another gene is currently under investigation.

Comparison of the genetic organization of the E. coli and B. subtilis thr operons

In *E. coli*, the *thr* operon is composed of three structural genes in the following order: *thrA*−*thrB*−*thrC*. The nucleotide sequence of this operon (Katinka *et al.*, 1980; Cossart *et al.*, 1981; Parsot *et al.*, 1983) indicates that the intergenic regions are very short; there is only one nucleotide between *thrA* and *thrB*, and the stop codon of *thrB* is adjacent to the start codon of *thrC*.

In *B. subtilis* previous genetic studies demonstrated: (i) a close linkage between *hom* (encoding homoserine dehydrogenase), *thrC* and *thrB* (Skarstedt and Greer, 1973); and (ii) that mutations in the *sprA* locus (unlinked to the *thr* cluster) cause the simultaneous derepression of the *hom*, *thrC* and *thrB* genes (Vapnek and Greer, 1971). These data suggested that the three genes could be organized in an operon in *B. subtilis*. Our results indicate that the stop codon of *thrC* may overlap the start codon of *thrB* (Figure 3). In addition the *thrC* start codon also overlaps a stop codon (nucleotides 246−248 in Figure 3) ending a reading frame open from the 5′ end of the sequenced fragment (the 81 codon long ORF X in Figure 3). In view of the close linkage between *hom* and *thrC* (Skarstedt and Greer, 1973), ORF X could represent the 3′ end of the *hom* gene, although the amino acid sequence deduced from ORF X does not present any homology with the C-terminal amino acid sequence of the *E. coli* homoserine dehydrogenase I (Katinka *et al.*, 1980).

Overlapping of the *thrC* stop codon and the *thrB* start codon most probably indicates that the two genes are part of an operon in *B. subtilis*. Although the *thrC* and *thrB* genes are adjacent in *E. coli* and overlapping in *B. subtilis*, their relative order is different in the two organisms, i.e. *thrC*−*thrB* in *B. subtilis* and *thrB*−*thrC* in *E. coli*. This is in marked contrast to the *trp* operon, involved in the tryptophan biosynthetic pathway, whose genetic organization is very similar in *E. coli* and in *B. subtilis* (Yanofsky *et al.*, 1981; Henner *et al.*, 1984). This similarity suggests that the construction of the *trp* operon was achieved before separation of the two species. On the other hand, the inversion of the *thrB* and *thrC* genes between *E. coli* and *B. subtilis* probably indicates that the *thr* operons were constructed independently in these two species.

Comparison of the B. subtilis and E. coli thrB and thrC gene products

The *B. subtilis* and *E. coli thrB* genes encode 298 and 308 amino acid long proteins respectively. Alignment of the two homoserine kinase sequences (Figure 4) indicates that, with eight gaps representing 22 positions introduced to maximize homology, the score of identity is 26% (73 identical residues) and rises to 46% when classes of homologous amino acid residues (Schwartz and Dayhoff, 1978) are taken into account. This homology encompasses the entire length of the two proteins.

Comparison of the *B. subtilis* and *E. coli thrC* gene products (Figure 5) reveals that *E. coli* threonine synthase has a 48 amino acid long N-terminal extension which accounts for most of the size difference between the two proteins (351 and 428 amino acid residues for *B. subtilis* and *E. coli* threonine synthase respectively). The N-terminal region of *B. subtilis* threonine synthase (amino acids 15−191) is quite homologous to the central region of the *E. coli* threonine synthase (amino acids 63−255). When five gaps representing 15 positions are introduced, the two protein segments are 34% homologous (51 identical residues), and this score rises to 42% when conservative changes are included. However, the degree of homology between the C-termini of the two proteins (i.e. amino acids 192−351 and 256−428 of *B. subtilis* and *E. coli* threonine synthase respectively) is much lower, with only 13 identical residues (Figure 5).

Similar levels of homology (~30%) have also been detected between functionally equivalent proteins encoded by the *E. coli*

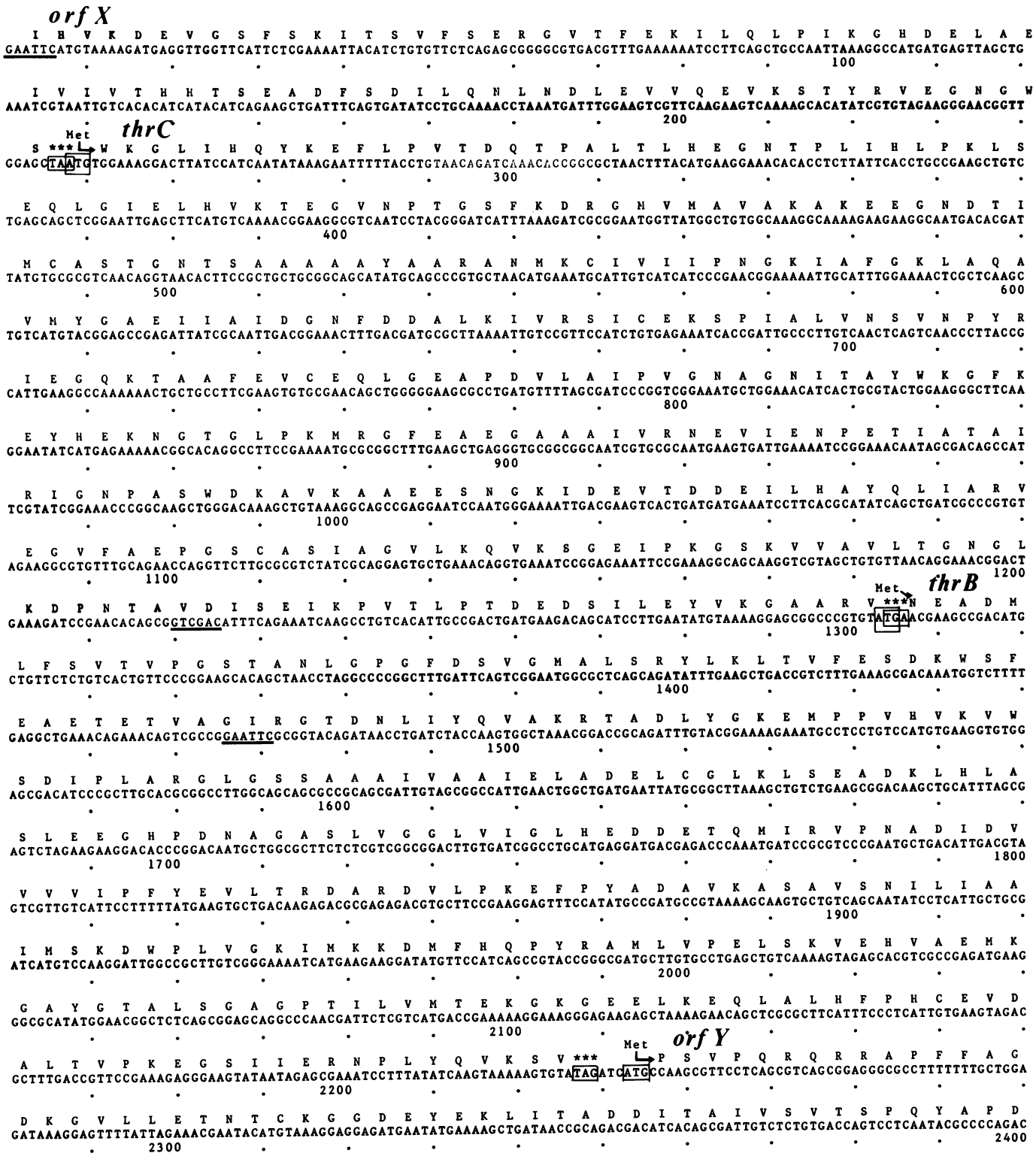

Fig. 3. Nucleotide sequence of the *B. subtilis thrC* and *thrB* genes and their flanking regions. The deduced amino acid sequences of the *thrC* and *thrB* encoded proteins as well as those of ORF X and ORF Y are presented in the one letter code above the nucleotide sequence of the non-coding strand. The stop codons of the relevant reading frames are boxed, as are the proposed start codon, above which is indicated the N-terminal Met residue. The *Eco*RI and *Sal*I sites used in subcloning experiments are underlined.

and *B. subtilis trp* operons (Henner *et al.*, 1984), and probably reflect the divergence between these two organisms.

Comparison of threonine synthase, threonine dehydratase and D-serine dehydratase

Knowledge of the *B. subtilis thrC* nucleotide sequence allowed further consideration of the intriguing results of Greer and collaborators, who demonstrated that *B. subtilis* threonine synthase can exhibit a low threonine dehydratase activity both *in vivo* and *in vitro* (Vapnek and Greer, 1971; Skarstedt and Greer, 1973). A possible relationship between threonine synthase and threonine dehydratase was also suggested by studies on their catalytic

mechanisms: the synthesis of threonine from phosphohomoserine, catalyzed by *Neurospora crassa* threonine synthase, proceeds by formation of an α,β unsaturated intermediate, α-aminocrotonate (see Figure 6), to which water is added to form threonine (Flavin and Slaughter, 1960). The same intermediate, α-aminocrotonate, has been postulated in the conversion of threonine to α-ketobutyrate catalyzed by threonine dehydratase (Figure 6); in the threonine dehydratase reaction, α-aminocrotonate is eliminated from the enzyme and converted into α-iminobutyrate which then spontaneously hydrolyzes to α-ketobutyrate and ammonia (Phillips and Wood, 1965).

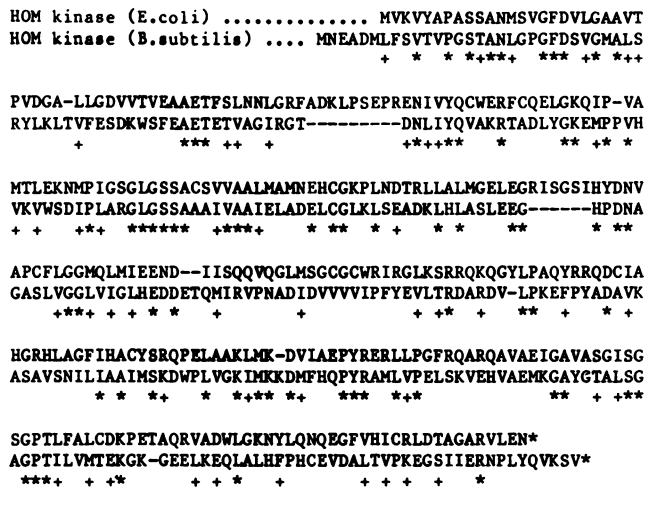

Fig. 4. Comparison of the *E. coli* and *B. subtilis* homoserine kinase sequences. The entire amino acid sequences of *E. coli* (Cossart *et al.*, 1981) and *B. subtilis* (this work) homoserine kinases are presented in the one letter code. They have been aligned by introducing gaps (−) to maximize the homology. Stars below the sequence indicate identical residues and + signs indicate chemically similar residues: I-L-V-M, D-E, R-K, S-T, F-Y (Schwartz and Dayhoff, 1978).

The amino acid sequence of *B. subtilis* threonine synthase was therefore compared with the sequence of threonine dehydratase deduced from the *Saccharomyces cerevisiae* ILV 1 gene (Kielland-Brandt *et al.*, 1984), the only known sequence of a threonine dehydratase. Despite the difference in the sizes of the two proteins (351 and 576 amino acid residues for *B. subtilis* threonine synthase and *S. cerevisiae* threonine dehydratase respectively), the two protein sequences revealed extensive homology. The alignment presented in Figure 5 involves nine gaps representing 15 positions for the whole sequence of *B. subtilis* threonine synthase compared to amino acids 52−408 of *S. cerevisiae* threonine dehydratase: 72 amino acid residues (21%) are conserved and 40 are accepted replacements (as defined in legend of Figure 5), which gives a total of 33% homology. Additional identities were revealed when the yeast enzyme was compared with *E. coli* threonine synthase, especially in the 50 amino acid N-terminal region which is not present in *B. subtilis* threonine synthase (Figure 5). This similarity between the threonine synthases and threonine dehydratase amino acid sequences most probably indicates a common ancestor for these proteins, which catalyze consecutive steps in isoleucine biosynthesis (see Figures 1 and 6).

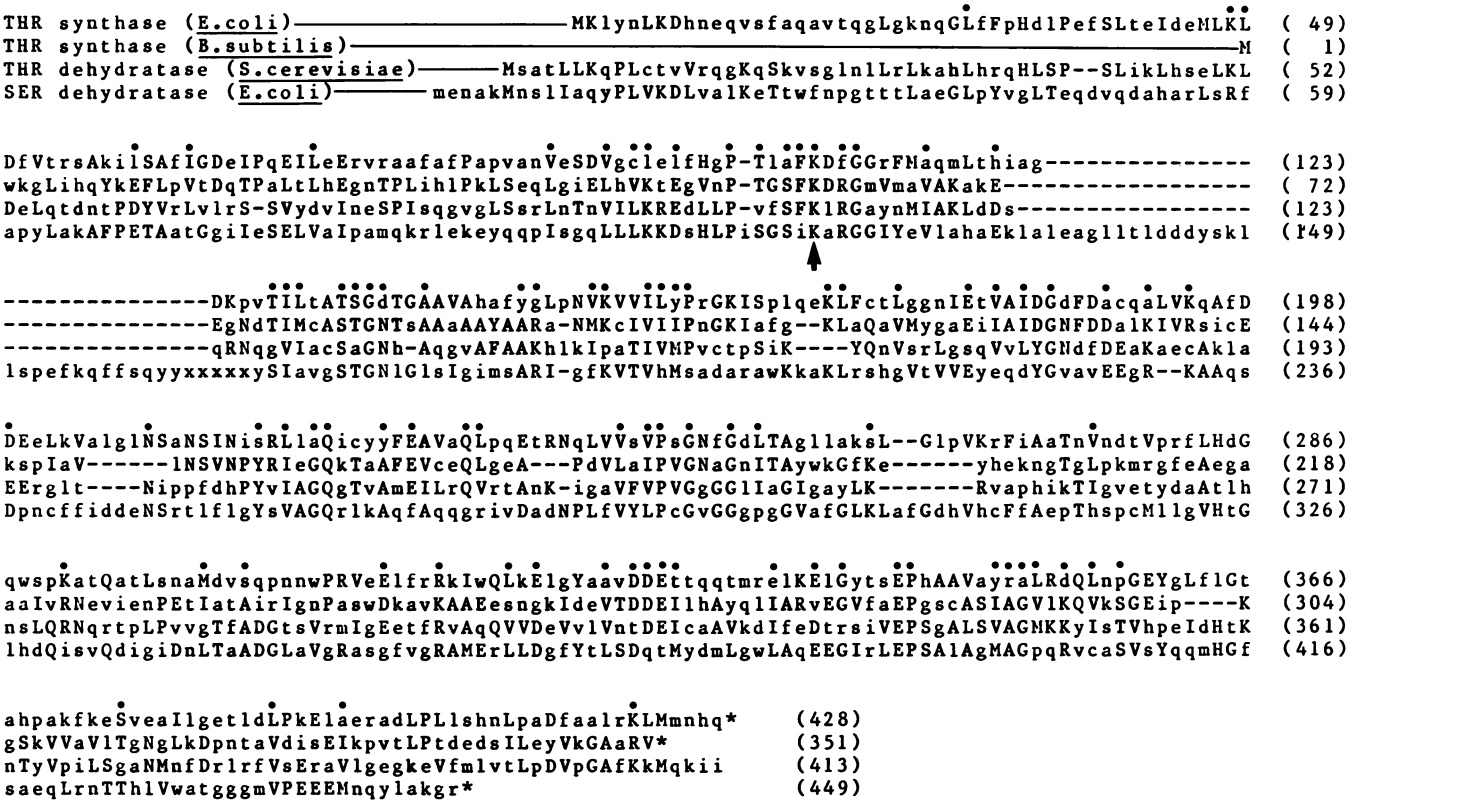

Fig. 5. Comparison of threonine synthase, threonine dehydratase and D-serine dehydratase amino acid sequences. The entire amino acid sequences of *E. coli* threonine synthase (Parsot *et al.*, 1983), *B. subtilis* threonine synthase (this work), *E. coli* D-serine dehydratase (Schiltz and Schmitt, 1981) and part of *S. cerevisiae* threonine dehydratase (576 amino acid residues, Kielland-Brandt *et al.*, 1984) are presented in the one letter code. Gaps (−) have been introduced to maximize homology. Identical and homologous residues are presented in upper case, others in lower case (accepted alternatives are I-L-V-M, D-E, R-K, S-T, F-Y). Positions where three or more residues are identical or similar are indicated by dots above the sequences. The arrow indicates the lysine residue of D-serine dehydratase which is bound to the pyridoxal phosphate co-factor (Schiltz and Schnackerz, 1976). Numbers in brackets refer to the position of the last presented amino acid residue of each line in the original sequences.

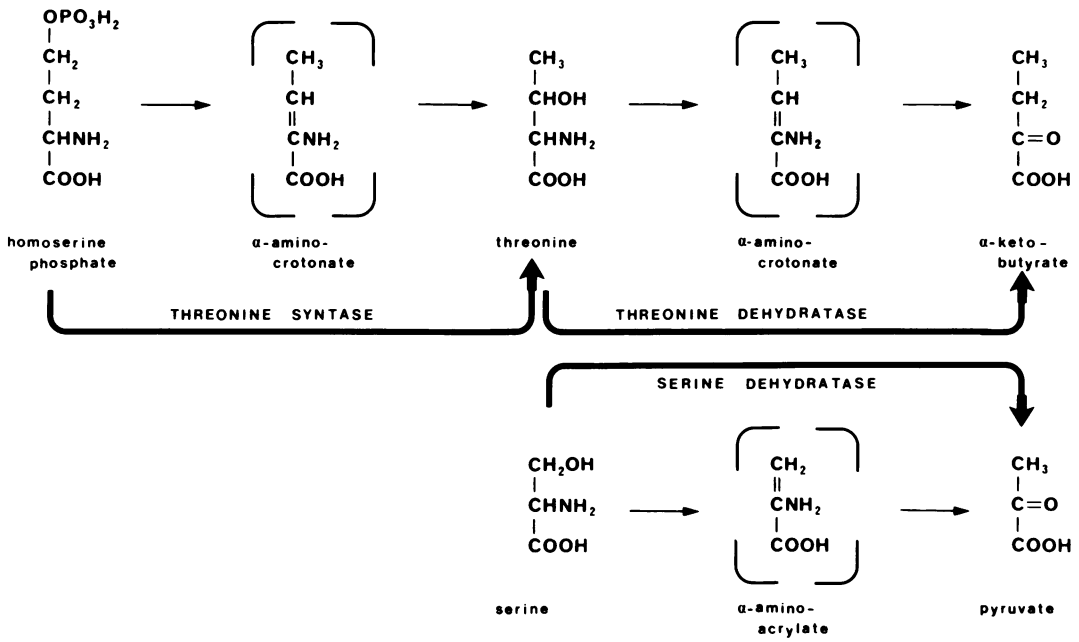

Fig. 6. Structure of the substrates, intermediates and products of the reactions catalyzed by threonine synthase, threonine dehydratase and serine dehydratase. The structure of α-aminocrotonate, one of the intermediates in the reactions catalyzed by threonine synthase (Flavin and Slaughter, 1960) and threonine dehydratase (Phillips and Wood, 1965), as well as the structure of α-aminoacrylate, an intermediate postulated in the reaction catalyzed by serine dehydratase (Chargaff and Sprinson, 1943) are presented in brackets. Inorganic phosphate, ammonia and water — the by-products of the reactions — have not been indicated for clarity.

This amino acid sequence comparison was then extended to serine dehydratase, as both mammalian and microbial degradative threonine dehydratases are able to deaminate L-serine as well as L-threonine (Nishimura and Greenberg, 1961; Shizuta *et al.*, 1969). A reactional intermediate homologous to α-aminocrotonate, α-aminoacrylate, has also been postulated in the reaction catalyzed by serine dehydratase (Figure 6) leading to pyruvate and ammonia (Chargaff and Sprinson, 1943). The sequence of *E. coli* D-serine dehydratase (Schiltz and Schmitt, 1981), the only known sequence of a serine dehydratase, was thus compared to those of threonine synthases and threonine dehydratase. With a gap of about 30 amino acid residues introduced at the same place in the threonine synthases and in the threonine dehydratase sequences, the whole sequence of D-serine dehydratase can easily be aligned with those of the three other enzymes (Figure 5): there are 61 identical residues and 38 homologous ones between *S. cerevisiae* threonine dehydratase and *E. coli* D-serine dehydratase. In addition, conserved amino acids were often present in three out of the four sequences (see dots above the alignment presented in Figure 5).

These enzymes (threonine synthase, threonine dehydratase and serine dehydratase) have the same co-factor, pyridoxal phosphate, which forms a Schiff base with the ϵ-NH_2 moiety of a lysine residue. Interestingly, the lysine residue of D-serine dehydratase identified as being covalently bound to pyridoxal phosphate co-factor (Schiltz and Schnackerz, 1976) is matched with the only lysine residue conserved in the alignment of threonine synthases and threonine dehydratase. This suggests that the pyridoxal phosphate co-factor of the threonine synthases and threonine dehydratase could be bound to this lysine residue, indicated by an arrow on Figure 5.

Conclusion

From (i) the threonine dehydratase activity exhibited by *B. subtilis* threonine synthase (Skarstedt and Greer, 1973) and by *E. coli* threonine synthase (B.Burr, personal communication); (ii) the similarities in the reaction mechanisms of threonine synthase and threonine dehydratase (Flavin and Slaughter, 1960; Phillips and Wood, 1965); (iii) the serine dehydratase activity exhibited by threonine dehydratase (Nishimura and Greenberg, 1961; Shizuta *et al.*, 1969); and (iv) the comparison of the amino acid sequences presented here, I conclude that threonine synthase, threonine dehydratase and D-serine dehydratase may have a common evolutionary origin. The variety of organisms from which the protein sequence data were obtained (*E. coli, S. cerevisiae* and *B. subtilis*) may have little impact, if any, on this proposal since it is likely that these proteins diverged from a common ancestor early in evolution, prior to the separation of these organisms. Considering the properties of these enzymes, it is reasonable to speculate that the primitive enzyme could have been a pyridoxal phosphate enzyme (due to the conservation of the lysine residue bound to the co-factor), with a wide substrate specificity (phosphohomoserine, threonine, serine) and able to produce related α-amino acid intermediates (α-aminocrotonate or α-aminoacrylate). Evolution could then have been achieved by gene duplications and subsequent mutations leading to specialization in the substrate specificity of the encoded proteins and to different controls of their activity and expression.

This example strongly supports the hypothesis that ancestral cells may have produced a relatively small number of enzymes, albeit unregulated and with a substrate ambiguous specificity, able to react with a wide range of related substrates (reviewed by Jensen, 1976). Although modern enzymes may be very specific, substrate ambiguity remains in some cases, as noted here for the threonine dehydratase activity exhibited by threonine synthase and the serine dehydratase activity exhibited by threonine dehydratase. Substrate ambiguity is, in some cases, readily exploited to perform a series of reactions, as in the parallel pathways leading to isoleucine and valine: each enzyme catalyzing one of the four steps of the pathways has dual specificity for the precur-

sors of valine and isoleucine (Figure 1). In that manner, a plausible scheme for the acquisition of isoleucine pathway itself could be the evolution of threonine dehydratase from threonine synthase (documented here by the homology between these two enzymes). The α-ketobutyrate could then be used to synthesize isoleucine solely by recruiting enzymes involved in valine biosynthesis.

Materials and methods

Bacterial strains

E. coli K12 derivatives GT28 *thrC1010* and GT121 *thrC1080 lacZU239 metLM1000 lysC1004* (Thèze and Saint-Girons, 1974) were used to test the various plasmids for *thrC* complementing activity. *E. coli* strain GT869 [*thrB1004 pro thi strA hsdS lacZΔM15* (F′ *lacZΔM15 lacI^q traD36 proA$^+$ proB$^+$*)] was used for the cloning experiments and to select plasmids able to complement *thrB*. This strain was constructed as follows: RR1ΔM15 [*leu pro thi strA hsdS lacZΔM15* (F′ *lacZΔM15 lacI^q pro$^+$*)] (Rüther, 1982) was cured of the F′ episome, and used as recipient for the conjugal transfer first of the *thrB1004* allele from strain GT22 (*thrB1004 HfrH*) (Thèze and Saint-Girons, 1974), and then of the F′ episome from strain JM101 (Yanisch-Perron *et al.*, 1985). Media and classical genetic experiments were according to Miller (1972). *B. subtilis* 168 derivative QB39 *trpC2 sacT30* (Lepesant *et al.*, 1972) was used for chromosomal DNA preparation.

Molecular cloning procedures

DNA preparation, DNA restriction and separation by agarose gel electrophoresis, hybridization of DNA digests, ligation and bacterial transformation were done according to the standard methods described in Maniatis *et al.* (1982). DNA sequences were determined by the dideoxy chain termination method (Sanger *et al.*, 1980) after cloning of the DNA fragments into the M13 mp18 and mp19 derivatives (Yanisch-Perron *et al.*, 1985). All the sites used for cloning were overlapped and most of the sequence (80%) was determined on both strands; the detailed sequencing strategy is available from the author on request. Restriction enzymes, T4 DNA ligase were from Boehringer (FRG), [α-^{35}S]dATP and M13 cloning and sequencing kits were from Amersham (UK).

Acknowledgements

I am deeply indebted to M.Arnaud and to G.Rapoport for the gift of pBS02A and *B. subtilis* DNA which have allowed the development of this work, and to I.Saint-Girons for the gift of *E. coli thr* strains and for helpful discussions. I am grateful to I.Crawford for comments and advice at the beginning of this work and to D.Mazel for his contribution in subcloning experiments. I wish to thank A.Pugsley for critical reading of the manuscript and L.Girardot for typing it. I am pleased to acknowledge G.N.Cohen for his encouragement and support. This work was supported by the Centre National de la Recherche Scientifique (U.A. 1129).

References

Chargaff,E. and Sprinson,D.B. (1943) *J. Biol. Chem.*, **151**, 273–280.
Cossart,P., Katinka,M. and Yaniv,M. (1981) *Nucl. Acids Res.*, **9**, 339–347.
Flavin,M. and Slaughter,C. (1960) *J. Biol. Chem.*, **235**, 1112–1118.
Gardner,J.F. (1979) *Proc. Natl. Acad. Sci. USA*, **76**, 1706–1710.
Henner,D.J., Band,L. and Shimotsu,H. (1984) *Gene*, **34**, 169–177.
Jensen,R.A. (1976) *Ann. Rev. Microbiol.*, **30**, 409–425.
Johnson,D.I. and Somerville,R.L. (1983) *J. Bacteriol.*, **155**, 49–55.
Katinka,M., Cossart,P., Sibilli,L., Saint-Girons,I., Chalvignac,M.A., Le Bras,G., Cohen,G.N. and Yaniv,M. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 5730–5733.
Kielland-Brandt,M.C., Holmberg,S., Petersen,J.G.L. and Nilsson-Tillgren,T. (1984) *Carlsberg Res. Commun.*, **49**, 567–575.
Lepesant,J.A., Kuntz,F., Lepesant-Keszlarova,J. and Dedonder,R. (1972) *Mol. Gen. Genet.*, **118**, 135–160.
Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning, A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Miller,J.H. (1972) *Experiments in Molecular Genetics*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Nishimura,J.S. and Greenberg,D.M.(1961) *J. Biol. Chem.*, **236**, 2684–2691.
Parsot,C., Cossart,P., Saint-Girons,I. and Cohen,G.N. (1983) *Nucl. Acids Res.*, **11**, 7331–7345.
Phillips,A.T. and Wood,W. (1965) *J. Biol. Chem.*, **240**, 4703–4709.
Primrose,S.B. and Ehrlich,S.D. (1981) *Plasmid*, **6**, 193–201.
Rapoport,G., Klier,A., Billault,A., Fargette,F. and Dedonder,R. (1979) *Mol. Gen. Genet.*, **176**, 239–245.
Rüther,U. (1982) *Nucl. Acids Res.*, **10**, 5765–5772.
Sanger,F., Coulson,A.R., Barrell,B.G., Smith,A.J.H. and Roe,B.A. (1980) *J. Mol. Biol.*, **143**, 161–178.
Schiltz,E. and Schmitt,W. (1981) *FEBS Lett.*, **134**, 57–62.
Schiltz,E. and Schnackerz,D. (1976) *Eur. J. Biochem.*, **71**, 109–116.
Schwartz,R.M. and Dayhoff,M.O. (1978) In Dayhoff,M.L. (ed.), *Atlas of Protein Sequence and Structure*. National Biomedical Research Foundation, Washington DC, Vol.5, pp. 353–358.
Shine,J. and Dalgarno,L. (1974) *Proc. Natl. Acad. Sci. USA*, **71**, 1342–1346.
Shizuta,Y., Nakazawa,A., Tokushige,M. and Hayaishi,O. (1969) *J. Biol. Chem.*, **244**, 1883–1889.
Skarstedt,M.T. and Greer,S.B. (1973) *J. Biol. Chem.*, **248**, 1032–1044.
Sullivan,M.A., Yasbin,R.E. and Young,F.E. (1984) *Gene*, **29**, 21–26.
Thèze,J. and Saint-Girons,I. (1974) *J. Bacteriol.*, **118**, 990–998.
Vapnek,D. and Greer,S.B. (1971) *J. Bacteriol.*, **106**, 983–993.
Vieira,J. and Messing,J. (1982) *Gene*, **19**, 259–268.
Yanisch-Perron,C., Vieira,J. and Messing,J. (1985) *Gene*, **33**, 103–119.
Yanofsky,C., Platt,T., Crawford,I.P., Nichols,B.P., Christie,G.E., Horowitz,H., Van Cleemput,M. and Wu,A.M. (1981) *Nucl. Acids. Res.*, **9**, 6647–6668.

Received on 16 June 1986; revised on 11 August 1986