Volume 11 Number 21 1983 Nucleic Acids Research

Nucleotide sequence of *thrC* and of the transcription termination region of the threonine operon in *Escherichia coli* K12

Claude Parsot, Pascale Cossart, Isabelle Saint-Girons and Georges N.Cohen

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

Received 19 September 1983; Accepted 14 October 1983

ABSTRACT

The entire threonine operon (thrABC) of Escherichia coli K12 was cloned, and the nucleotide sequence of the thrC gene and its 3' flanking region was determined. The translation initiation codon was identified by sequencing the N-terminal part of threonine synthase, the thrC gene product. Analysis of the deduced protein sequence (428 amino acid residues) revealed a region of homology, 35 amino acids long, between the three enzymes encoded by the threonine operon. During examination of the nucleotide sequence of the 1045 base pair fragments following the thrC gene, we detected some potential rho-independent and rho-dependent transcription termination signals.

INTRODUCTION

The threonine operon (1), located at 0 min on the Escherichia coli K12 genetic map, is composed of three genes, thrA, thrB, and thrC, which code for 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] respectively. These enzymes catalyse four of the five steps of the branched pathway leading from aspartate to threonine. Expression of the threonine operon is controlled both by threonine and isoleucine (2). Examination of the nucleotide sequence of the regulatory region of the operon led to a proposed attenuation mechanism for the regulation of the threonine operon (3). This proposal was supported by in vitro and in vivo analysis of the threonine transcripts (4,5), and by characterization of a deletion mutation in the attenuator region which resulted in constitutive expression of the operon (6).

The nucleotide sequences of thrA (7) and thrB (8) have been recently reported. In order to examine the complete structure of the threonine operon, we cloned the entire operon and determined the nucleotide sequence of the thrC gene and its 3' flanking region. In addition, we purified the thrC gene product, threonine synthase, and determined

© IRL Press Limited, Oxford, England.

its N-terminal amino acid sequence. We then compared the sequences of the three proteins encoded by the threonine operon and searched for regions of homology. We also analysed the nucleotide sequence of the transcription termination region and identified two possible transcription terminators, as well as a potential gene in opposite orientation of the operon.

MATERIALS AND METHODS

Bacterial and bacteriophage strains

The following strains of *Escherichia coli* K12 were used: AR1062 (9) *thr*, *leu*, *lac*, *gal*, *xyl*, *mal*, *mtl*, *hsdS*; C600 (10) *thrB1023*, *leu*, *recBC*; GT123 (10) *pyrA53*, *pro-1000*, Δ(*thr*), *metL1005*, *lysC1004*; and Tir8 (11,12) *ilvS*, *ilvU*, *ilvT*. Bacteriophage $\lambda dthr_c$, isolated by Schrenk and Weisberg (13), carries the *CI857*, *S7* mutations and transduces the whole threonine operon (14).

Media and chemicals

Growth media are described by Miller (15). T4 DNA ligase was a generous gift of O. Danos, all other enzymes were purchased from New England Laboratories or Boehringer Mannheim and used according to the manufacturers' instructions. [γ-^{32}P]ATP (3000 Ci/mmole), [α-^{32}P] dXTP (3000 Ci/mmole), [α-^{32}P] cordycepin triphosphate (3000 Ci/mmole) were from Amersham and [^{35}S] methionine (1000 Ci/mmole) was from New England Nuclear. Chemicals used for DNA sequencing were of the highest grade commercialy available. Chemicals used for protein sequencing were from Beckman.

Buffers: All buffers used for the purification of threonine synthase contained 2mM L-threonine, 0.05mM pyridoxal-5'-phosphate, 1mM dithiothreitol. Buffer A: 20mM potassium phosphate (pH 7.2), 2mM EDTA, 250mM KCl. Buffer B: 10mM potassium phosphate (pH 7.2), 150mM KCl. Buffer C: 20mM potassium phosphate (pH 7.2), 2mM EDTA, 150mM KCl.

Enzyme assay

Threonine synthase was assayed according to Daniel (16). Protein concentration was determined by the method of Bradford (17) using bovine serum albumin as a standard.

Purification of the threonine synthase

Threonine synthase was purified from 3 kg (wet weight) of *E.coli* K12, strain Tir8, which is a constitutive mutant for the enzymes encoded by the threonine operon (18). Culture conditions have been previously

Table 1 :
Purification of threonine synthase. The specific activity is expressed as nmoles of threonine synthesized per min per mg of protein, at 37°C.

Fraction	Volume (ml)	Total proteins	Specific activity	Purification (fold)	Recovery %
I	4920	70	220	1	100
II	870	24.4	490	2.2	80
III	175	1.4	4560	20.7	42
IV	91	1.3	3810	17.3	32
V	11.5	0.52	7390	33.6	25
VI	52	0.32	7750	35	19

described (19). The enzyme was purified using a streptomycin sulfate precipitation and ammonium sulfate fractionation as previously reported (19), followed by the purification steps described in table 1.

The supernatant from the 40% saturation ammonium sulfate fractionation (Fraction I) was adjusted to 60% saturation by addition of solid ammonium sulfate and the precipitated protein was collected by centrifugation. The precipitated material was resuspended in buffer A, dialysed against the same buffer and loaded onto a DEAE-Sephadex A-50 (Pharmacia Fine Chemicals) column (45 x 5.5cm), preequilibrated with buffer A. Threonine synthase was not retained and fractions containing the activity were pooled (Fraction II), precipitated with ammonium sulfate (60% saturation), resuspended in buffer B, and dialysed against the same buffer. The dialysed material was loaded onto a hydroxylapatite (Biorad) column (50 x 3cm) preequilibrated with buffer B. The enzyme was not retained and fractions containing threonine synthase were pooled (Fraction III), precipitated with ammonium sulfate (60% saturation), resuspended and dialysed against buffer C. The protein was loaded onto a DEAE-Sephadex A-50 column (30 x 2cm) preequilibrated with buffer C. The column was

developed with a linear gradient (2 x 200ml) of 150mM to 400mM KCl in buffer C. Fractions containing activity were pooled (Fraction IV), solid ammonium sulfate was added to 50% saturation, and the precipitated material removed by centrifugation. The supernatant containing threonine synthase was adjusted to 60% saturation, and the precipitate was dialysed against buffer C (Fraction V). Fraction V was loaded onto an Ultrogel AcA44 (LKB) column (93 x 2.5cm), and fractions containing threonine synthase were analysed by polyacrylamide gel electrophoresis in the presence of sodium dodecyl sulfate (PAGE-SDS). Fractions containing the pure enzyme were pooled (Fraction VI) and the threonine synthase was stored at 4°C in buffer C containing ammonium sulfate (60% saturation).

Protein electrophoresis

Electrophoresis was performed either on a 7.5% to 15% exponential gradient polyacrylamide gel in the presence of SDS according to O'Farrell (20), or on a 10% polyacrylamide gel, containing SDS, according to Laemmli (21).

Amino acid sequence determination

The purified threonine synthase was subjected to automated Edman degradation using a Beckman 890C sequencer modified as described in (22). The PTH amino acid derivatives were identified by high pressure liquid chromatography on a Waters Associates Model 6000A equiped with a Model 440 UV detector.

Purification of plasmid DNA

Plasmid DNA, extracted by the cleared lysate method (23), was purified by centrifugation through an ethidium bromide-cesium chloride gradient (24). Plasmid DNA was further purified by sucrose gradient centrifugation (5-20%). For rapid screening of recombinant plasmids, plasmid DNA was extracted as described by Birnboim and Doly (25).

Nucleotide sequence determination

DNA fragments were 5'-end-labeled with [γ-^{32}P] ATP and polynucleotide kinase by the exchange reaction (26) as described by Maxam and Gilbert (27). Fragments were 3'-end-labeled either with [α-^{32}P] dXTP using DNA polymerase (large fragment) (28) or with [α-^{32}P] cordycepin triphosphate using deoxynucleotidyl terminal transferase (29).

Restriction fragments were separated on polyacrylamide gels (4%, 8% or 10%) as described by Maxam and Gilbert (27), and then electro-eluted. Strand separation was performed essentially as described in

(27), except that 50% dimethylsulfoxide was used instead of 30%.

End-labeled restriction fragments were subjected to base specific chemical cleavage according to Maxam and Gilbert (30). The G, A+G, A>C, C, C+T reactions were used, and the products were analysed on 20% and 8% polyacrylamide thin gels (31) containing 8M urea.

RESULTS AND DISCUSSION

Cloning of the threonine operon

The nucleotide sequence of thrA (7) and thrB (8) was previously determined using the recombinant plasmid pIP2 (10). This plasmid carries thrA, thrB and part of the thrC gene on a 4.2 Kb HindIII-EcoRI DNA fragment. To determine the complete nucleotide sequence of thrC, we cloned the entire threonine operon. The λdthr$_c$ bacteriophage DNA previously used for the construction of pIP2 (10) was cleaved by the HindIII restriction enzyme. The fragments mixture was inserted in the HindIII site of plasmid vector pBR322 (32), and used to transform E.coli strain C600 (thrB). The recombinant plasmid, pIP30, carries the threonine operon on a 12.8 Kb insert (Figure 1). The threonine operon was subcloned from pIP30 DNA to give the recombinant plasmid pIP3 containing the entire operon on a 6.3 Kb HindIII-BamHI insert (Figure 1).

In order to verify that pIP3 contained the complete threonine operon, this plasmid was used to transform E.coli strain GT123 (ΔthrABC, metL, lysC). All of the ampicillin resistant transformants were Thr$^+$, indicating that the thrA, thrB and thrC genes were present on pIP3 and were expressed in the recipient cell.

Analysis of the pIP3 encoded proteins

The pIP3, pIP2 and pBR322 plasmids were introduced by transformation into the AR1062 minicell producing strain (9). After purification of the minicells and incorporation of $[^{35}S]$ L-methionine, proteins were analysed by PAGE-SDS (see autoradiogram Figure 2). Since the cloning sites used to construct pIP2 and pIP3 were HindIII, EcoRI and HindIII, BamHI respectively, the tet gene product (a 33000 dalton polypeptide) was not expected with the recombinant plasmids. Strains harboring each of these plasmids are sensitive to tetracycline, confirming that the tet gene is not expressed. Two bands corresponding to molecular weights of 87000 and 34000 are observed in the case of pIP2 and pIP3 (but not with pBR322), in agreement with the known molecular weights of aspartokinase I-homoserine dehydrogenase I (89000) and homoserine kinase

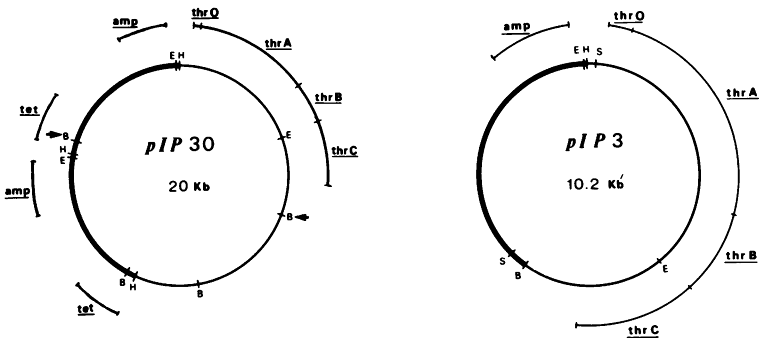

Figure 1 :

Restriction map of pIP30 and pIP3 plasmids. Recognition sites for EcoRI (E), HindIII (H), BamHI (B), and SalI (S) restriction endonucleases are indicated on the DNA of pIP30 and pIP3 plasmids. The heavy line represents the pBR322 part. The positions of the genes are indicated on the outer circles. Arrows indicate the BamHI sites used to construct pIP3 from pIP30.

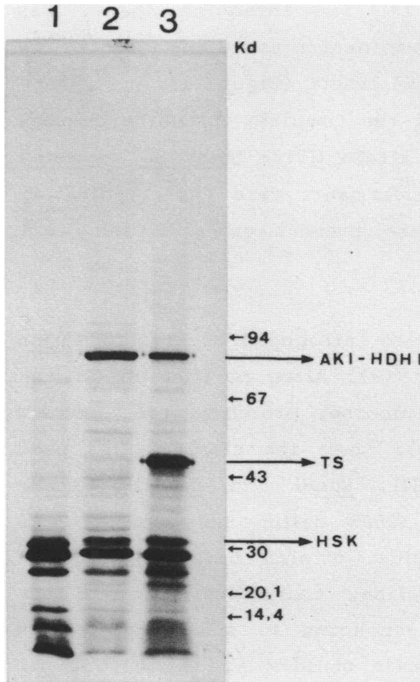

Figure 2 :
Analysis of polypeptides coded by the recombinant plasmids pIP2 and pIP3 in the minicell producing strain AR1062. After purification of the minicells and incorporation of [^{35}S] L-methionine, the polypeptides synthesized were analysed by PAGE-SDS and visualised by autoradiography; Lane 1: pBR322; Lane 2: pIP2; Lane 3: pIP3. The migration of protein standards is indicated in kilodaltons (kd). Identification of aspartokinase I-homoserine dehydrogenase I (AK I-HDH I), threonine synthase (TS), and homoserine kinase (HSK) is based on their electrophoretic migrations.

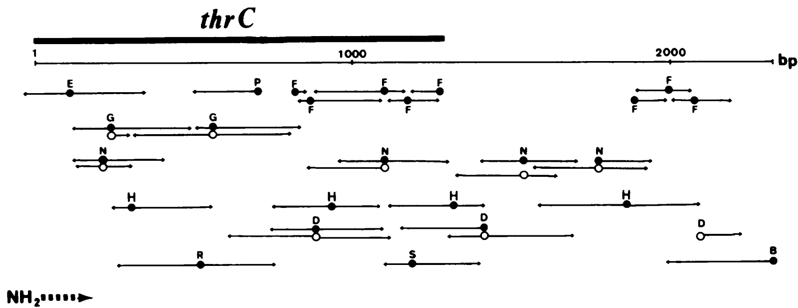

Figure 3 :

Sequencing strategy used for the determination of the nucleotide sequence of the thrC gene. Position and extent of thrC is indicated by the heavy line. The dashed arrow represents the extent of the amino acid sequence determined by automated Edman degradation on purified threonine synthase. The thin arrows indicate the sites (E: EcoRI; P: PvuII; F: HinfI; G: BglI; N: BstNI; H: HaeIII; D: DdeI; R: RsaI; S: Sau3AI; B: BamHI) used for 5' (●) or 3' (○) labeling as well as the direction and extent of the regions that were sequenced.

(34000), the thrA and thrB genes products respectively. In the case of pIP3, there is an additional band corresponding to a molecular weight of 48000, consistent with the molecular weight of 46000 estimated for purified threonine synthase by PAGE-SDS (data not shown).

Nucleotide sequence of the thrC gene

The nucleotide sequence of thrC was determined using the sequencing strategy shown in Figure 3. The sequence is presented in Figure 4 along with the deduced amino acid sequence of threonine synthase. The previously published sequence (8) of the 5' extremity of the gene, up to the EcoRI site, was corrected in positions 15-16 where CT has been changed to TC and between positions 38-39 where a T has been deleted.

In order to unambiguously identify the start codon of the thrC gene, we purified threonine synthase and determined the N-terminal sequence of the first 40 amino acid residues. This sequence was identical to that deduced from the nucleotide sequence of thrC, including the Met residue in position 1.

The thrC gene codes for a 428 amino acid polypeptide with a deduced molecular weight of 47060, in perfect agreement with the values of 46000 and 48000 estimated for threonine synthase as discussed above.

Amino acid sequence analysis

The amino acid sequences of the three proteins encoded by the

thr B ***thr C***

LEU GLU ASN MET LYS LEU TYR ASN LEU LYS ASP HIS ASN GLU GLN VAL SER PHE ALA GLN ALA VAL THR GLN GLY LEU GLY LYS ASN GLN GLY
CTG GAA AAC TAA ATG AAA CTC TAC AAT CTG AAA GAT CAC AAC GAG CAG GTC AGC TTT GCG CAA GCC GTA ACC CAG GGG TTG GGC AAA AAT CAG GGG

29
LEU PHE PHE PRO HIS ASP LEU PRO GLU PHE SER LEU THR GLU ILE ASP GLU MET LEU LYS LEU ASP PHE VAL THR ARG SER ALA LYS ILE LEU SER
CTG TTT TTT CCG CAC GAC CTG CCG GAA TTC AGC CTG ACT GAA ATT GAT GAG ATG CTG AAG CTG GAT TTT GTC ACC CGC AGT GCG AAG ATC CTC TCG

61
ALA PHE ILE GLY ASP GLU ILE PRO GLN GLU ILE LEU GLU GLU ARG VAL ARG ALA ALA PHE ALA PHE PRO ALA PRO VAL ALA ASN VAL GLU SER ASP
GCG TTT ATT GGT GAT GAA ATC CCA CAG GAA ATC CTG GAA GAG CGC GTG CGC GCG GCG TTT GCC TTC CCG GCT CCG GTC GCC AAT GTT GAA AGC GAT

93
VAL GLY CYS LEU GLU LEU PHE HIS GLY PRO THR LEU ALA PHE LYS ASP PHE GLY GLY ARG PHE MET ALA GLN MET LEU THR HIS ILE ALA GLY ASP
GTC GGT TGT CTG GAA TTG TTC CAC GGG CCA ACG CTG GCA TTT AAA GAT TTC GGC GGT CGC TTT ATG GCA CAA ATG CTG ACC CAT ATT GCG GGT GAT

125
LYS PRO VAL THR ILE LEU THR ALA THR SER GLY ASP THR GLY ALA ALA VAL ALA HIS ALA PHE TYR GLY LEU PRO ASN VAL LYS VAL VAL ILE LEU
AAG CCA GTG ACC ATT CTG ACC GCG ACC TCC GGT GAT ACC GGA GCG GCA GTG GCT CAT GCT TTC TAC GGT TTA CCG AAT GTG AAA GTG GTT ATC CTC

157
TYR PRO ARG GLY LYS ILE SER PRO LEU GLN GLU LYS LEU PHE CYS THR LEU GLY GLY ASN ILE GLU THR VAL ALA ILE ASP GLY ASP PHE ASP ALA
TAT CCA CGA GGC AAA ATC AGT CCA CTG CAA GAA AAA CTG TTC TGT ACA TTG GGC GGC AAT ATC GAA ACT GTT GCC ATC GAC GGC GAT TTC GAT GCC

189
CYS GLN ALA LEU VAL LYS GLN ALA PHE ASP ASP GLU GLU LEU LYS VAL ALA LEU GLY LEU ASN SER ALA ASN SER ILE ASN ILE SER ARG LEU LEU
TGT CAG GCG CTG GTG AAG CAG GCG TTT GAT GAT GAA GAA CTG AAA GTG GCG CTA GGG TTA AAC TCG GCT AAC TCG ATT AAC ATC AGC CGT TTG CTG

221
ALA GLN ILE CYS TYR TYR PHE GLU ALA VAL ALA GLN LEU PRO GLN GLU THR ARG ASN GLN LEU VAL VAL SER VAL PRO SER GLY ASN PHE GLY ASP
GCG CAG ATT TGC TAC TAC TTT GAA GCT GTT GCG CAG CTG CCG CAG GAG ACG CGC AAC CAG CTG GTT GTC TCG GTG CCA AGC GGA AAC TTC GGC GAT

253
LEU THR ALA GLY LEU LEU ALA LYS SER LEU GLY LEU PRO VAL LYS ARG PHE ILE ALA ALA THR ASN VAL ASN ASP THR VAL PRO ARG PHE LEU HIS
TTG ACG GCG GGT CTG CTG GCG AAG TCA CTC GGT CTG CCG GTG AAA CGT TTT ATT GCT GCG ACC AAC GTG AAC GAT ACC GTG CCA CGT TTC CTG CAC

285
ASP GLY GLN TRP SER PRO LYS ALA THR GLN ALA THR LEU SER ASN ALA MET ASP VAL SER GLN PRO ASN ASN TRP PRO ARG VAL GLU GLU LEU PHE
GAC GGT CAG TGG TCA CCC AAA GCG ACT CAG GCG ACG TTA TCC AAC GCG ATG GAC GTG AGT CAG CCG AAC AAC TGG CCG CGT GTG GAA GAG TTG TTC

317
ARG ARG LYS ILE TRP GLN LEU LYS GLU LEU GLY TYR ALA ALA VAL ASP ASP GLU THR THR GLN GLN THR MET ARG GLU LEU LYS GLU LEU GLY TYR
CGC CGC AAA ATC TGG CAA CTG AAA GAG CTG GGT TAT GCA GCC GTG GAT GAT GAA ACC ACG CAA CAG ACA ATG CGT GAG TTA AAA GAA CTG GGC TAC

349
THR SER GLU PRO HIS ALA ALA VAL ALA TYR ARG ALA LEU ARG ASP GLN LEU ASN PRO GLY GLU TYR GLY LEU PHE LEU GLY THR ALA HIS PRO ALA
ACT TCG GAG CCG CAC GCT GCC GTA GCT TAT CGT GCG CTG CGT GAT CAG TTG AAT CCA GGC GAA TAT GGC TTG TTC CTC GGC ACC GCG CAT CCG GCG

381
LYS PHE LYS GLU SER VAL GLU ALA ILE LEU GLY GLU THR LEU ASP LEU PRO LYS GLU LEU ALA GLU ARG ALA ASP LEU PRO LEU LEU SER HIS ASN
AAA TTT AAA GAG AGC GTG GAA GCG ATT CTC GGT GAA ACG TTG GAT CTG CCA AAA GAG CTG GCA GAA CGT GCT GAT TTA CCC TTG CTT TCA CAT AAT

413 428
LEU PRO ALA ASP PHE ALA ALA LEU ARG LYS LEU MET MET ASN HIS GLN
CTG CCC GCC GAT TTT GCT GCG TTG CGT AAA TTG ATG ATG AAT CAT CAG TAA

Figure 4 :

Nucleotide sequence of the thrC gene. The complete nucleotide sequence of thrC and amino acid sequence of the encoded threonine synthase, as well as the C-terminal extremity of homoserine kinase coded by thrB are presented. The amino acid sequence is numbered from the first residue of threonine synthase. The possible ribosome binding site for the translation of thrC is underlined.

threonine operon were compared to determine if they could have been derived from a common ancestor, in line with Horowitz's hypothesis (33,34). The only similarity we found was within a 35 amino acid region which is present twice in aspartokinase I-homoserine dehydrogenase I (residues 15 to 49 and 272 to 305) and once in both homoserine kinase (residues 273 to 305) and threonine synthase (residues 14 to 47). As

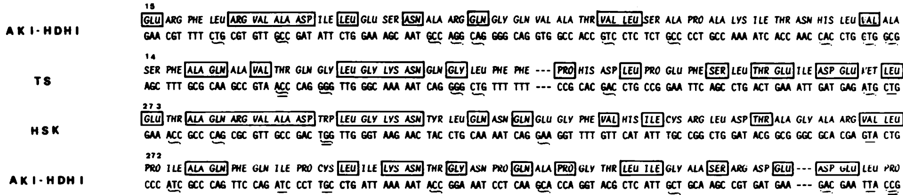

Figure 5 :

Sequence homologies in the primary structure of aspartokinase I-homoserine dehydrogenase I (AK I-HDH I), homoserine kinase (HSK) and threonine synthase (TS). Numerals above the amino acid residues indicate the position of the residue in the sequence of each protein. The corresponding nucleotide sequences are shown below the amino acid sequences. Two gaps have been introduced into the sequences at positions indicated by dashed lines. Identical amino acid residues are boxed and codons which code for different amino acid residues but differ by only one nucleotide are underlined by wavy or straight lines.

shown in Figure 5, the homology is also detectable at the level of the corresponding nucleotide sequence, even for amino acids which are not identical.

The region of homology is too short to conclude that the genes have a common ancestor, and its significance remains to be determined. One possible explanation is that these repeated sequences are the result of some genetic exchanges between genes belonging to the same operon, as proposed by Ornston and Yeh (35). Another possibility is that this homology is related to a common function of these 3 proteins, such as the binding of threonine. Threonine is an allosteric effector for aspartokinase I-homoserine dehydrogenase I with 2 sites per monomer (36), a competitive inhibitor for homoserine kinase (37), and the product of the reaction catalysed by threonine synthase, as well as the substrate of the minor threonine dehydratase activity of this enzyme (38).

Nucleotide sequence of the DNA fragment located downstream from the thrC gene

The nucleotide sequence of the 1 Kb DNA fragment located downstream from the thrC gene is shown Figure 6, along with the particular features which have been detected and are discussed below. Examination of this sequence did not reveal any open reading frame that could stand for a fourth gene in the same orientation as the threonine operon.

Potential transcription termination sites: Twenty one base pairs after the stop codon of thrC, there is a GC rich region of dyad symmetry (see Figure 6). When transcribed, this region could form a stable stem

```
▲▲▲        10         20         30         40         50         60
TAAAATCTAT TCATTATCTC AATCAGGCCG GGTTTGCTTT TATGCAGCCC GGCTTTTTTA
ATTTTAGATA AGTAATAGAG TTAGTCCGGC CCAAACGAAA ATACGTCGGG CCGAAAAAAT

        70         80         90        100        110        120
TGAAGAAATT ATGGAGAAAA ATGACAGGGA AAAAGGAGAA ATTCTCAATA AATGCGGTAA
ACTTCTTTAA TACCTCTTTT TACTGTCCCT TTTTCCTCTT TAAGAGTTAT TTACGCCATT

       130        140        150        160        170        180
CTTAGAGATT AGGATTGCGG AGAATAACAA CCGCCGTTCT CATCGAGTAA TCTCCGGATA
GAATCTCTAA TCCTAACGCC TCTTATTGTT GGCGGCAAGA GTAGCTCATT AGAGGCCTAT

       190        200        210        220        230        240
TCGACCCATA ACGGGCAATG ATAAAAGGAG TAACCTGTGA AAAAGATGCA ATCTATCGTA
AGCTGGGTAT TGCCCGTTAC TATTTTCCTC ATTGGACACT TTTTCTACGT TAGATAGCAT

       250        260        270        280        290        300
CTCGCACTTT CGCTGGTTCT GGTCGCTCCC ATGGCAGCAC AGGCTGCGGA AATTACGTTA
GAGCGTGAAA GCGACCAAGA CCAGCGAGGG TACCGTCGTG TCCGACGCCT TTAATGCAAT

       310        320        330        340        350        360
GTCCCGTCAG TAAAATTACA GATAGGCGAT CGTGATAATC GTGGCTATTA CTGGGATGGA
CAGGGCAGTC ATTTTAATGT CTATCCGCTA GCACTATTAG CACCGATAAT GACCCTACCT

       370        380        390        400        410        420
GGTCACTGGC GCGACCACGG CTGGTGGAAA CAACATTATG AATGGCGAGG CAATCGCTGG
CCAGTGACCG CGCTGGTGCC GACCACCTTT GTTGTAATAC TTACCGCTCC GTTAGCGACC

       430        440        450        460        470        480
CACCTACACG GACCGCCGCC ACCGCCGCGC CACCATAAGA AAGCTCCTCA TGATCATCAC
GTGGATGTGC CTGGCGGCGG TGGCGGCGCG GTGGTATTCT TTCGAGGAGT ACTAGTAGTG

       490        500        510        520        530        540
GGCGGTCATG GTCCTGGCAA ACATCACCGC TAAATGACAA ATGCCGGGTA ACAATCCGGC
CCGCCAGTAC CAGGACCGTT TGTAGTGGCG ATTTACTGTT TACGGCCCAT TGTTAGGCCG

       550 UP2    560        570        580        590 UP3    600
ATTCAGCGCC TGATGCGACG CTGGCGCGTC TTATCAGGCC TACGTTAATT CTGCAATATA
TAAGTCGCGG ACTACGCTGC GACCGCGCAG AATAGTCCGG ATGCAATTAA GACGTTATAT

       610        620        630        640        650        660
TTGAATCTGC ATGCTTTTGT AGGCAGGATA AGGCGTTCAC GCCGCATCCG GCATTGACTG
AACTTAGACG TACGAAAACA TCCGTCCTAT TCCGCAAGTG CGGCGTAGGC CGTAACTGAC

       670        680        690        700 UP1    710        720
CAAACTTAAC GCTGCTCGTA GCGTTTAAAC ACCAGTTCGC CATTGCTGGA GGAATCTTCA
GTTTGAATTG CGACGAGCAT CGCAAATTTG TGGTCAAGCG GTAACGACCT CCTTAGAAGT
     *** R  Q  E  Y   R  K  F    V  L  E  G   N  S  S    S  D  E
       730        740        750        760        770        780
TCAAAGAAGT AACCTTCGCT ATTAAAACCA GTCAGTTGCT CTGGTTTGGT CAGCCGATTT
AGTTTCTTCA TTGGAAGCGA TAATTTTGGT CAGTCAACGA GACCAAACCA GTCGGCTAAA
D  F  F  Y   G  E  S   N  F  G   T  L  Q  E   P  K  T   L  R  N
       790        800        810        820        830        840
TCAATAATGA AACGACTCAT CAGACCGCGT GCTTTCTTAG CGTAGAAGCT GATGATCTTA
AGTTATTACT TTGCTGAGTA GTCTGGCGCA CGAAAGAATC GCATCTTCGA CTACTAGAAT
E  I  I  F   R  S  M   L  G  R   A  K  K  A   Y  F  S   I  I  K
       850        860        870        880        890        900
AATTTGCCGT TCTTCTCATC GAGGAACACC GGCTTGATAA TCTCGGCATT CAATTTCTTC
TTAAACGGCA AGAAGAGTAG CTCCTTGTGG CCGAACTATT AGAGCCGTAA GTTAAAGAAG
F  K  G  N   K  E  D   L  F  V   P  K  I  I   E  A  N   L  K  K
       910        920        930        940        950        960
GGCTTCACCG ATTTAAAATA CTCATCTGAC GCCAGATTAA TCACCACATT ATCGCCTTGT
CCGAAGTGGC TAAATTTTAT GAGTAGACTG CGGTCTAATT AGTGGTGTAA TAGCGGAACA
P  K  V  S   K  F  Y   E  D  S   A  L  N  I   V  V  N   D  G  Q
       970        980        990       1000       1010       1020
GCTGCGAGCG CCTCGTTCAG CTTGTTGGTG ATGATATCTC CCCAGAATTG ATACAGATCT
CGACGCTCGC GGAGCAAGTC GAACAACCAC TACTATAGAG GGGTCTTAAC TATGTCTAGA
A  A  L  A   E  N  L   K  N  T   I  I  D  G   W  F  Q   Y  L  D
      1030       1040
TTCCCTCGGG CATTCTCAAG ACGGATCC-
AAGGGAGCCC GTAAGAGTTC TGCCTAGG-   orf 127
K  G  R  A   N  E  L   R  I
```

and loop secondary structure (ΔG = -25 kcal/mole) followed by a stretch of 6 uridine residues. Such a structure is characteristic of rho-independent transcription termination sites (39) and may be a terminator for transcription at the end of the operon.

We also located some possible rho-dependent transcription termination signals based on the following criteria (39): 1) an AT rich sequence in the transcription termination region, 2) a sequence related to CAATCAA, and 3) regions of dyad symmetry preceding this sequence. Downstream from the potential rho-independent terminator, there is an AT rich region (nucleotides 60 to 240 in Figure 6) with several stretches of A residues and 2 sequences similar to CAATCAA preceded by small regions of dyad symmetry.

The biological significance of these features is currently under investigation and it will be of particular interest to determine if the association of different types of terminator is a common characteristic of bacterial operons, as already found for the tryptophan operon (40) and for the tyrT locus (41).

Presence of some palindromic units: A novel genetic element common to several intercistronic regions of bacterial operons has been recently described by Higgins et al. (42). This element consists of a long dyad symmetry composed of repeats of a palindromic unit (P.U.). The function(s) of this element is still unknown, but it has been proposed (42) that it could act as a transcription termination site or as a processing site for the transcripts. A systematic computer search in data banks, performed by Clément and Hofnung (manuscript in preparation), detected 37 occurrences of this palindromic unit.

Such a sequence was also found three times (P.U. 1, 2, 3 in Figure 6), 520 base pairs after the thrC gene. In addition we found that the DNA sequence located in between P.U. 1 and P.U. 2 shows a strong

Figure 6 :
Nucleotide sequence downstream from the thrC gene. The nucleotide sequence is numbered from the stop codon of thrC, indicated by ▲ in positions 1-3. Arrows indicate regions of dyad symmetry relevent to the possible (see text) rho-independent termination signal (nucleotides 25 to 55) and rho-dependent signals (nucleotides 87 to 112, and 216 to 235) for which the sequences similar to CAATCAA are indicated by brackets. The sequences designated UP 1, UP 2, UP 3, are homologous to the repeated element described by Higgins et al. (42). The amino acid sequence deduced from the nucleotide sequence of orf127 (see text) is indicated in the one letter code below the corresponding DNA sequence. The BamHI site is indicated by a dotted line.

```
orf127  ( 44)  CTAC AAAAgCATGC AgATTCAA TATATT-G CAGaatTaAC GTAG  ( 86)
gly A   ( 70)  CTAC AAAActtTGC AAATTCAA TATATT-G CAAtctccgt GTAG  (112)
gln S   (166)  CTgC AAAAgCAcGg ---gcTgg TgTGTT-G CAGagaTcAt GTAG  (113)
met J   ( 99)  CTAC AAgttCgTGC AAATTCAA TAaATT-G CAAta-TgAC GTAG  ( 54)
met L   (120)  CaAt AAgtaCATGg ttAgTttA TATATTTG CAGtccggtt tgct  (163)
trp R   ( 80)  CTAC AAAtaCcgGt -AATTCAA TATGTTTG ---------- GTAG  (112)
                                            A          A
CONSENSUS      CTAC AAAAnCATGC AAATTCAA TAT TTTG CA nnnTnAC GTAG
                                            G          G
his G   ( 65)  CTAC AgAAcCc--- ----aaAA TATcaacG CA----TtAC GTAG  ( 33)
lam B   ( 82)  CTAC AAcggCtgtC AAAT---- -------- ---------- GTAG  (103)
```

Figure 7 :

Comparison of the nucleotide sequences located in between P.U.s detected after glyA (43), glnS (47), trpR (48,49), metL (44), hisG (42), lamB (50), orf127 (see text) and metJ (Zakin and Duchange, personal communication). Numerals in brackets indicate the position of the first and last nucleotides of the presented sequence distal to the relevent gene. Residues identical to the "consensus" are presented in upper cases. Homologies were optimised by introducing gaps, indicated by dashed lines.

homology with the sequences located between the P.U.s detected after trpR (42), glnS and glyA (43), metL (44) and metJ (Zakin and Duchange, personal communication) (see Figure 7). This homology, although weaker, was also found in the sequences flanked by P.U.s located in the intercistronic regions (42) after lamB and hisG (but not after hisJ, for which this sequence is only 18 nucleotides long). This suggests that the common

Table 2 : Codon usage in orf127.

UUU *Phe* 5	UCU *Ser* 0	UAU *Tyr* 2	UGU *Cys* 0
UUC *Phe* 5	UCC *Ser* 2	UAC *Tyr* 3	UGC *Cys* 0
UUA *Leu* 0	UCA *Ser* 1	UAA 1	UGA 0
UUG *Leu* 1	UCG *Ser* 1	UAG 0	UGG *Trp* 1
CUU *Leu* 1	CCU *Pro* 0	CAU *His* 0	CGU *Arg* 3
CUC *Leu* 2	CCC *Pro* 0	CAC *His* 0	CGC *Arg* 2
CUA *Leu* 0	CCA *Pro* 1	CAA *Gln* 3	CGA *Arg* 1
CUG *Leu* 7	CCG *Pro* 2	CAG *Gln* 1	CGG *Arg* 1
AUU *Ile* 4	ACU *Thr* 1	AAU *Asn* 7	AGU *Ser* 1
AUC *Ile* 6	ACC *Thr* 2	AAC *Asn* 3	AGC *Ser* 3
AUA *Ile* 0	ACA *Thr* 0	AAA *Lys* 7	AGA *Arg* 0
AUG *Met* 1	ACG *Thr* 0	AAG *Lys* 7	AGG *Arg* 0
GUU *Val* 0	GCU *Ala* 1	GAU *Asp* 7	GGU *Gly* 3
GUC *Val* 0	GCC *Ala* 2	GAC *Asp* 0	GGC *Gly* 3
GUA *Val* 0	GCA *Ala* 3	GAA *Glu* 4	GGA *Gly* 1
GUG *Val* 5	GCG *Ala* 2	GAG *Glu* 7	GGG *Gly* 1

element was originally composed of two convergent palindromic units bordering a sequence similar to the "consensus" sequence indicated in Figure 7. As this sequence is not symmetrical, the entire element (about 110 base pairs long) can be oriented with respect to the transcription direction, and is found in both orientations.

Presence of an open reading frame in opposite orientation to the threonine operon: Since these palindromic units have always been detected very close to genes (42, Clément and Hofnung, manuscript in preparation), we looked for evidence of a gene distal to the threonine operon. There is an open reading frame 127 codons long in opposite orientation to the threonine operon, from nucleotide 1048 to nucleotide 669 on Figure 6 (nucleotide 1048 is part of the BamHI site limiting the bacterial insert on pIP3 DNA). This open reading frame, designated orf127, ends 9 nucleotides before P.U. 1. In orf127 the codon usage is not random (see table 2) and shows a positive correlation between the choice of codons and the relative tRNA abundance, as found for most E.coli genes (45).

These data (codon usage and position relative to P.U.s) suggest the existence of a gene and thus of a transcription unit located downstream from and in opposite orientation to the threonine operon. The size and function of the protein coded by orf127 are still unknown, but the nearest identified locus clockwise to the threonine operon on the E.coli genetic map is tolJ, which has been located approximately at 0.1 min (46).

This work completes the determination of the nucleotide sequence of the threonine operon, and provides a solid basis for future biochemical and genetic studies on the transcription of the 3 genes encoded by the operon.

ACKNOWLEDGMENTS

We are very grateful to B. Burr and J. Daniel for communication of unpublished results and protocols on purification of threonine synthase, P. Parouteau and D. Strosberg for the rapid and expert determination of the N-terminal sequence of the protein, G. Rapoport for invaluable help in monitoring the minicell experiments, and B. Caudron for assistance and advice in the computer analysis of sequence data. We thank J.M. Clément and M. Hofnung for helpful discussions and communication of results prior to publication, N. Duchange and M. Zakin for kindly providing unpublished results. Also we thank D. Margarita for technical assistance and L. Girardot for typing the manuscript. Construction of pIP30 recombinant plasmid was done during the Cours de Microbiologie Générale de l'Institut Pasteur, in autumn 1980,

and all of the students are gratefully acknowledged. We are indebted to R. Hohman for critical reading of the manuscript.

This work was supported by grants of the Centre National de la Recherche Scientifique (G.R. n° 30). Claude Parsot was recipient of a long term fellowship from the Délégation Générale à la Recherche Scientifique et Technique.

REFERENCES

1. Thèze,J. and Saint-Girons,I. (1974) J. Bact. 118, 990-998
2. Freundlich,M. (1963) Biochem. Biophys. Res. Comm. 10, 277-282
3. Gardner,J.F. (1979) Proc. Natl. Acad. Sci. USA 76, 1706-1710
4. Gardner,J.F. (1982) J. Biol. Chem. 257, 3896-3904
5. Lynn,S.P., Gardner,J.F. and Reznikoff,W.S. (1982) J. Bact. 152, 363-371
6. Parsot,C., Saint-Girons,I. and Cossart,P. (1982) Molec. Gen. Genet. 188, 455-458
7. 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
8. Cossart,P., Katinka,M. and Yaniv,M. (1981) Nucl. Acids Res. 9, 339-347
9. Rambach,A. and Hogness,D.S. (1977) Proc. Natl. Acad. Sci. USA 74, 5041-5045
10. Cossart,P., Katinka,M., Yaniv,M., Saint-Girons,I. and Cohen,G.N. (1979) Molec. Gen. Genet. 175, 39-44
11. Szentirmai,A., Szentirmai,M. and Umbarger,H.E. (1968) J. Bact. 95, 1672-1679
12. Brenchley,J.E. and Williams,L.S. (1975) Ann. Rev. Microbiol. 29, 251-274
13. Schrenk,W.J. and Weisberg,R. (1975) Molec. Gen. Genet. 137, 101-107
14. Thèze,J. and Margarita,D. (1974) Molec. Gen. Genet. 132, 41-48
15. Miller,J.H. (1972) in Experiments in Molecular Genetics, Cold Spring Harbor Laboratory, New York
16. Daniel,J. (1976) Nature 264, 90-92
17. Bradford,M. (1976) Anal. Biochem. 72, 248-254
18. Dwyer,S.B. and Umbarger,H.E. (1968) J. Bact. 95, 1680-1684
19. Falcoz-Kelly,F., Janin,J., Saari,J.C., Véron,M., Truffa-Bachi,P. and Cohen,G.N. (1972) Eur. J. Biochem. 28, 507-519
20. O'Farrell,P.H. (1975) J. Biol. Chem. 250, 4007-4021
21. Laemmli,U.K. (1970) Nature 227, 680-685
22. Wittman-Liebold,B., Graffunder,H. and Kohls,H. (1976) Anal. Biochem. 75, 621-633
23. Katz,L., Kingsbury,D.K. and Helinski,D.R. (1973) J. Bact. 114, 577-591
24. Radloff,R., Bauer,W. and Vinograd,J. (1967) Proc. Natl. Acad. Sci. USA 57, 1514-1521
25. Birnboim,H.C. and Doly,J. (1979) Nucl. Acids Res. 7, 1513-1524
26. Berkner,K.L. and Folk,W.L. (1977) J. Biol. Chem. 252, 3176-3184
27. Maxam,A.M. and Gilbert,W. (1980) in Methods in Enzymology, Grossman,L. and Moldave,K. Eds, Vol 65, pp. 499-560, Academic Press, New York
28. Klenow,H., Overgaard-Hansen,K. and Patkar,S.A. (1971) Eur. J. Biochem. 22, 371-381
29. Tu,C.P.D. and Cohen,S.N. (1980) Gene 10, 177-183

30. Maxam,A.M. and Gilbert,W. (1977) Proc. Natl. Acad. Sci. USA 74, 560-564
31. Sanger,F. and Coulson,N.A. (1978) FEBS Lett. 87, 107-110
32. Bolivar,F., Rodriguez,R.L., Greene,P.J., Betlach,M.L., Heynecker, H.L. and Boyer,H.W. (1977) Gene 2, 95-113
33. Horowitz,N.H. (1945) Proc. Natl. Acad. Sci. USA 31, 153-157
34. Horowitz,N.H. (1965) in Evolving Genes and Proteins, Bryson,V. and Vogel,H.J. Eds, pp. 15-23, Academic Press
35. Ornston,L.N. and Yeh,W.K. (1979) Proc. Natl. Acad. Sci. USA 76, 3996-4000
36. Véron,M., Saari,J.C., Villar-Palasi,C. and Cohen,G.N. (1973) Eur. J. Biochem. 38, 325-335
37. Burr,B., Walker,J., Truffa-Bachi,P. and Cohen,G.N. (1976) Eur. J. Biochem. 62, 519-526
38. Skarstedt,M.T. and Greer,S.B. (1973) J. Biol. Chem. 248, 1032-1044
39. Rosenberg,M. and Court,D. (1979) Ann. Rev. Genet. 13, 319-353
40. Wu,A.M., Christie,G.E. and Platt,T. (1981) Proc. Natl. Acad. Sci. USA 78, 2913-2917
41. Rossi,J., Egan,J., Hudson,L. and Landy,A. (1981) Cell 26, 305-314
42. Higgins,C.F., Ferro-Luzzi Ames,G., Barnes,W.M., Clément,J.M. and Hofnung,M. (1982) Nature 298, 760-762
43. Plamann,M.D., Stauffer,L.T., Urbanowski,M.L. and Stauffer,G.V. (1983) Nucl. Acids Res. 11, 2065-2075
44. Duchange,N., Zakin,M.M., Ferrara,P., Saint-Girons,I., Park,I., Tran,V.S., Py,M.C. and Cohen,G.N. (1983) J. Biol. Chem., in press
45. Ikemura,T. (1981) J. Mol. Biol. 146, 1-21
46. Davies,J.K. and Reeves,P. (1975) J. Bact. 123, 102-117
47. Yamao,F., Inokuchi,H., Cheung,A., Ozeki,H. and Söll,D. (1982) J. Biol. Chem. 257, 11639-11643
48. Gunsalus,R.P. and Yanofsky,C. (1980) Proc. Natl. Acad. Sci. USA 77, 7117-7121
49. Singleton,C.K., Roeder,W.D., Bogosian,G., Somerville,R.L. and Weith, H.L. (1980) Nucl. Acids Res. 8, 1551-1560
50. Clément,J.M. and Hofnung,M. (1981) Cell 27, 507-514.