Nucleotide sequence of the gene encoding the GMP reductase of *Escherichia coli* K12

Simon C. ANDREWS and John R. GUEST*
Department of Microbiology, University of Sheffield, Western Bank, Sheffield S10 2TN, U.K.

(1) The nucleotide sequence of a 1991 bp segment of DNA that expresses the GMP reductase (*guaC*) gene of *Escherichia coli* K12 was determined. (2) This gene comprises 1038 bp, 346 codons (including the initiation codon but excluding the termination codon), and it encodes a polypeptide of M_r 37437 which is in good agreement with previous maxicell studies. (3) The sequence contains a putative promoter 102 bp upstream of the translational start codon, and this is immediately followed by a (G+C)-rich discriminator sequence suggesting that *guaC* expression may be under stringent control (4) The GMP reductase exhibits a high degree of sequence identity (34%) with IMP dehydrogenase (the *guaB* gene product) indicative of a close evolutionary relationship between the salvage pathway and the biosynthetic enzymes, GMP reductase and IMP dehydrogenase, respectively. (5) A single conserved cysteine residue, possibly involved in IMP binding to IMP dehydrogenase, was located within a region that possesses some of the features of a nucleotide binding site. (6) The IMP dehydrogenase polypeptide contains an internal segment of 123 amino acid residues that has no counterpart in GMP reductase and may represent an independent folding domain flanked by (alanine+glycine)-rich interdomain linkers.

INTRODUCTION

GMP reductase (NADPH: GMP oxidoreductase; EC 1.6.6.8) catalyses the irreversible and NADPH-dependent reductive deamination of GMP to IMP:

$$\mathrm{GMP} + \mathrm{NADPH} + 2\mathrm{H}^+ \rightarrow \mathrm{IMP} + \mathrm{NH_4}^+ + \mathrm{NADP}^+$$

It functions in the conversion of nucleobase, nucleoside and nucleotide derivatives of G to A nucleotides, and in maintaining the intracellular balance of A and G nucleotides (Neuhard & Nygaard, 1987). GMP reductase has been purified from several sources, e.g. calf thymus (Stephens & Whittaker, 1973), human erythrocytes (Spector *et al.*, 1979), *Artemia salina* (Renart & Sillero, 1974), *Leishmania donovani* (Spector & Jones, 1982), as well as from *Escherichia coli* and *Salmonella typhimurium* (Mager & Magasanik, 1960; Neuhard & Nygaard, 1987). The bacterial enzymes are strongly inhibited by ATP and reactivated by GTP and the *S. typhimurium* enzyme is reported to be a tetramer of identical 45 kDa subunits (Neuhard & Nygaard, 1987).

The GMP reductase of *E. coli* is encoded by the *guaC* gene, which is located between the *mutT* and *nadC* genes at 2.6 min in the linkage map (Bachmann, 1983; Roberts *et al.*, 1988). Mutants lacking GMP reductase do not exhibit a purine requirement because the biosynthesis *de novo* of AMP and GMP is not affected. However, in purine auxotrophs that are blocked prior to the formation of IMP, *guaC* mutations prevent the use of G and X derivatives as sources of purine. The synthesis of GMP reductase in *E. coli* and *S. typhimurium* is increased by G, but this induction is blocked by A (Gots *et al.*, 1977; Nijkamp & DeHaan, 1967). The induction is also reported to require cyclic AMP in *E. coli*, but not in *S. typhimurium* (Benson *et al.*, 1971; Benson & Gots, 1975). In addition glutamine seems to act as a negative effector of *guaC* transcription because the synthesis of the enzyme increases during glutamine starvation and in the presence of glutamine analogues, but this is not related to the regulation of other nitrogen assimilatory enzymes (Garber *et al.*, 1980; Kessler & Gots, 1985). Thus it appears that the conversion of GMP to IMP is regulated by the ratio of G nucleotides to A nucleotides and that glutamine is involved in the regulation of *guaC* expression. Studies with *guaC* regulatory mutants have indicated that a *cis*-active operator or a closely-linked repressor is also involved in *guaC* expression (Kessler & Gots, 1985). The enzymes of the GMP biosynthetic pathway (IMP dehydrogenase and GMP synthetase) are encoded by the *guaBA* operon, which is located at 53.9 min in the *E. coli* linkage map, and is regulated independently of the *guaC* gene (Mehra & Drabble, 1981).

The *guaC* gene was originally cloned in several λ and pBR322 derivatives containing segments of the *nadC–aroP–aceEF–lpd* region during studies on the pyruvate dehydrogenase complex (Guest & Stephens, 1980; Guest *et al.*, 1983). The *guaC* gene was subsequently located in the 3.0 kbp *Eco*RI-*Bam*HI fragment of pGS89 (Fig. 1) that was derived by subcloning from the 10.5 kbp *Hin*dIII segment of the *nadC*$^+$–*aroP*$^+$ plasmid pGS15 (Roberts *et al.*, 1988). The GMP reductase activities of strains containing pGS89 were amplified 15-fold relative to untransformed strains under non-inducing conditions and the *guaC* gene product was identified as a polypeptide of M_r 37000 by maxicell analysis. The polarity of *guaC* transcription was also inferred from the properties of a truncated polypeptide that was expressed by a deletion

* To whom correspondence should be addressed.

These sequence data have been submitted to EMBL/Genbank Data Libraries under the accession number X07917.

derivative of pGS89, and from the β-galactosidase activity of a putative *guaC–lacZ* fusion. The *guaC* gene was independently isolated from an RP4::Mu cointegrate carrying the *leu–guaC* region by Moffat & Mackinnon (1985), but there are discrepancies between their restriction map and the map shown in Fig. 1. This could be due to a rearrangement of the bacterial DNA in their *guaC*$^+$ plasmid (pKGM71) because its construction involved subcloning from a partial *Sau*3A digest. Nevertheless, their *guaC*$^+$ plasmid expressed a polypeptide of M_r 36000 that was absent with *guaC*$^-$ plasmids that had deletions or Tn*5* insertions in the region adjoining a *Bgl*II site (presumed to be Bg_1 in Fig. 1).

The present paper reports the nucleotide sequence of a 1991 bp segment of *E. coli* DNA containing the *guaC* gene, the amino acid sequence of the GMP reductase monomer, and a high degree of homology between GMP reductase and IMP dehydrogenase.

EXPERIMENTAL

Strains of *E. coli*, plasmids and phages

The following strains of *E. coli* K12 were used for the purposes specified: ED8641 (*hsdR recA56*) from N. E. Murray, University of Edinburgh, Edinburgh, U.K., a transformable host for routine plasmid construction and preparation; GM242 (*dam-3 recA1*) for preparing *Bcl*I-susceptible plasmids (Marinus & Morris, 1973); JM101 (*Δlac-proAB* supE thi/F′ *traD36 proA*$^+$*B*$^+$ *lacI*^{q}*Z ΔM15*), for preparing M13 templates for sequence analysis (Messing, 1983); TX549 (*ΔguaC–aceE purD*::Tn*5 thi*) for testing plasmids for the presence of the *guaC* gene (Roberts *et al.*, 1988).

The source of DNA for sequencing the *guaC* gene was the pBR322 derivative, pGS89 (Roberts *et al.*, 1988). It contains the 3.0 kb *Eco*RI/*Bam*HI fragment (E_1–B_1) from pGS15 (*guaC*$^+$*–aroP*$^+$) recloned between the corresponding sites in pBR322 (Fig. 1). A deletion derivative of pGS89, designated pGS235, was constructed by treatment with *Bcl*I plus *Bam*HI followed by religation (Fig. 1). The replicative forms of the M13 mp18 and M13 mp19 phages were used for subcloning and preparing templates for DNA sequencing (Yanisch-Peron *et al.*, 1985).

Growth tests and recombinant DNA techniques

The minimal and rich media used in growth tests and in the selection of transformants have been described previously (Guest *et al.*, 1983). For testing the Pur$^-$ GuaC$^{+/-}$ phenotypes of TX549 derivatives, minimal media were supplemented with (final concentration): thiamine (5 μg/ml); nicotinic acid (5 μg/ml); sodium acetate (2 mM); and A (35 μg/ml) or either guanosine (100 μg/ml) or G (35 μg/ml). Ampicillin (50 μg/ml) was added to rich media to select or maintain Amp^R transformants and kanamycin (25 μg/ml) was used to confirm the presence of Tn*5*. The methods used for constructing, preparing and analysing plasmids have been described elsewhere (Guest *et al.*, 1983; Maniatis *et al.*, 1982).

Cloning in M13 and DNA sequence analysis

The sequencing strategy involved cloning specific fragments of pGS89 into the corresponding sites of M13mp18 and M13mp19 (Fig. 1). The fragments included the *Eco*RI–*Sph*I (E_1–Sp_1), *Bgl*II–*Sph*I (Bg_1–Sp_1, in both orientations; Sp_1–Bg_2, in both orientations; and Bg_2–Sp_v), *Sst*I-*Bam*HI (St_1–B_1), *Sst*I-*Eco*RI (St_1–E_1), *Bgl*II–*Eco*RI (Bg_1–E_1) *Bcl*I–*Bgl*II (Bc–Bg_1), *Bcl*I–*Sal*I (Bc–S_v), and *Eco*RV–*Bgl*II (Ec–Bg_1 and Ec–Bg_2). Single-stranded M13 DNA templates were prepared and sequenced by the dideoxy chain-termination method using 'universal' primer, [α-^{35}S]thio-dATP and buffer-gradient gels (Sanger *et al.*, 1980; Biggin *et al.*, 1983). The amounts of sequence obtained from some of the clones (Bg_1–Sp_1, Bg_2–Sp_1, St_1–E_1 and St_1–B_1) were increased using four specific oligonucleotide primers, S70 (5′GTGATGGTTTCTACCGG 3′), S72 (5′CGGGGAAAAACACATGGC 3′), S73 (5′ TTTCGCAACGAACTGCA 3′) and S74 (5′ CAACAACCTGTAATCTC 3′), respectively. Nucleotide sequences were compiled and analysed with the aid of the Staden computer programs (Staden, 1979, 1980; Staden & McLachlan, 1982). Sequence comparisons were performed using the DIAGON program of Staden (1982).

Materials

Restriction endonucleases, T4-DNA ligase and DNA polymerase (Klenow fragment) were purchased from Gibco-BRL, Uxbridge, Middx., U.K., Boehringer Corp. and Northumbrian Biochemicals, Cramlington, Northumbria, U.K., respectively. The M13 mp18 and M13 mp19 replicative-form DNAs were from Pharmacia-PL Biochemicals and [α-[^{35}S]thio]-dATP was supplied by Amersham International. The specific primers (S70, 72, 73 and 74) were made with an Applied Biosystems 381A DNA Synthesizer.

RESULTS AND DISCUSSION

Location of the *guaC* gene

The *guaC* gene was traced to the 3.0 kb-*Eco*RI–*Bam*HI region of the 10.5 kb-*Hin*dIII fragment cloned in pGS15 during previous subcloning and complementation studies (Fig. 1; Roberts *et al.*, 1988). It was further concluded that the *guaC* gene spans the *Bgl*II and *Sph*I sites (Bg_1 and Sp_1 in Fig. 1), because plasmids containing smaller fragments (E_1–Sp_1, Bg_1–Bg_2 and Sp_1–B_1 in Fig. 1) failed to confer a GuaC$^+$ phenotype. GuaC$^+$ activity correlated with a polypeptide (M_r 37000) that was the only detectable product expressed from the 3.0 kb-*Eco*RI–*Bam*H fragment in pGS89. The polarity of *guaC* transcription (left to right in Fig. 1) was likewise deduced from the truncation of this polypeptide from M_r 37000 to M_r 28500) that accompanied the excision of the 2.4 kb-*Sph*I fragment from pGS89. The location of the *guaC* gene has now been confirmed following the discovery of a unique *Bcl*I site in pGS89, which allowed a facile deletion *in vitro* of a 1.6 kb-*Bcl*I–*Bam*HI fragment (Bc–B_1) and the creation of a plasmid designated pGS235 that contains a smaller (1.4 kb) insert (Fig. 1). This plasmid conferred a GuaC$^+$ phenotype upon Amp^R transformants of the deletion strain TX549 (*ΔguaC–aceE purD*::Tn*5*) indicating that the 1.4 kb-*Eco*RI–*Bcl*I region (E_1–Bc) encodes a functional GMP reductase (Fig. 1).

Nucleotide sequence and identification of the *guaC* coding region

The complete nucleotide sequence of the 1692 bp-*Eco*RI–*Bgl*II fragment (E_1–Bg_2) containing the *guaC* gene was determined from both strands using overlapping

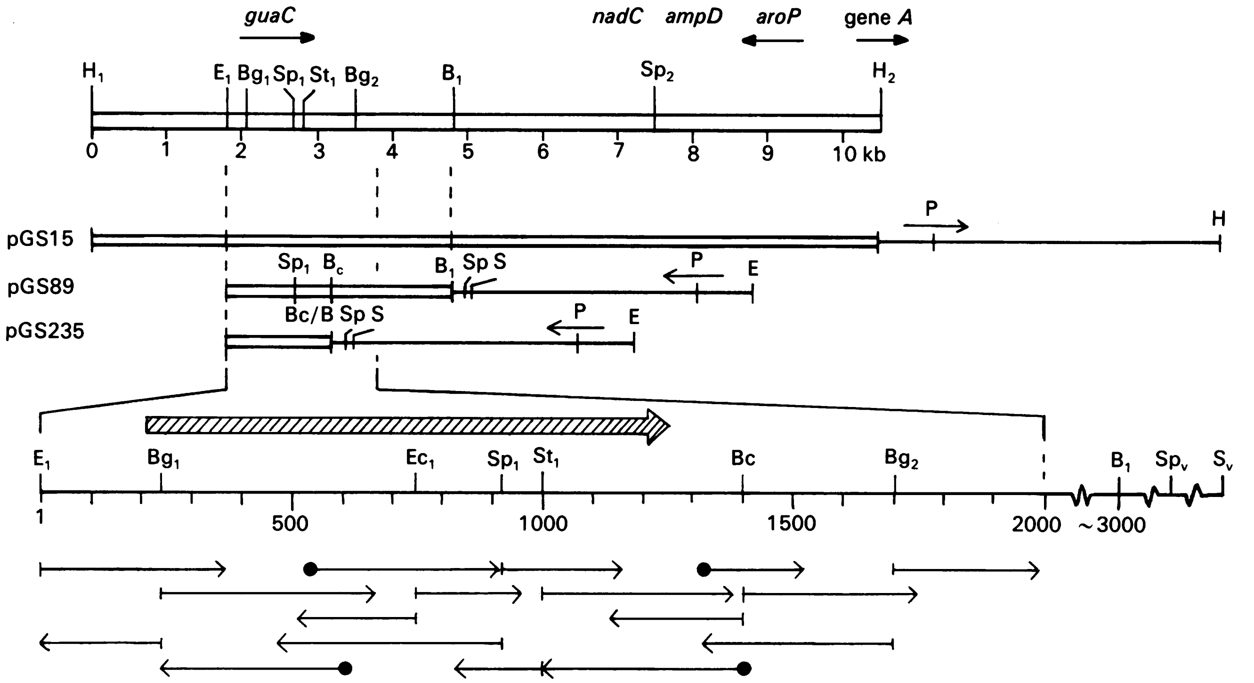

Fig. 1. Location of the *guaC* gene and summary of the DNA sequence data derived from M13 clones

Restriction map of the *guaC–aroP* region at 2.6 min in the *E. coli* linkage map showing the relative positions of the *guaC*, *nadC*, *ampD* and *aroP* genes according to Roberts *et al.* (1988), Guest *et al.* (1983), Chye *et al.* (1986) and Lindberg *et al.* (1987). Left to right corresponds to clockwise in the linkage map. The plasmid subclones (Amp^R Tet^S) are shown with open bars denoting bacterial DNA and lines denoting vector DNA, and the positions and polarities of the vector *bla* genes are indicated by horizontal arrows. The positions and extents of sequences obtained from M13 clones are indicated by the arrows in the expanded region. The nucleotide co-ordinates are numbered in bp from the first base of the *Eco*RI site (E$_1$) and the filled circles (●) identify sequences derived with the aid of specific primers. Relevant restriction sites are abbreviated and numbered according to Guest *et al.* (1983); *Bam*HI, B; *Bcl*I, Bc; *Bgl*II, Bg; *Eco*RI, E; *Eco*RV, Ec; *Hin*dIII, H; *Pst*I, P; *Sal*I, S; *Sph*I, Sp; and *Sst*I, St. Some of the vector sites are denoted by a v subscript, and the *guaC* coding region is indicated by the hatched arrow.

DNA fragments and several specific oligonucleotide primers (Figs. 1 and 2). The sequence extending a further 299 bp rightwards from the *Bgl*II site was also obtained on one strand. Only one potential coding region was detected using the FRAMESCAN program of Staden & McLachlan (1982) with the *E. coli* pyruvate dehydrogenase complex genes (Stephens *et al.*, 1983) as standards. It begins with an ATG codon at position 210 and extends for 1038 bp to a stop codon (TAA) at position 1248. The open reading frame encodes a polypeptide of 346 amino acid residues and M_r 37437 (including the initiating methionine), which closely matches the M_r reported by Roberts *et al.* (1988) and Moffat & Mackinnon (1985) for the *guaC* gene product. No significant coding region could be detected in either strand of the 700 bp segment distal to the *guaC* gene.

Features of the nucleotide sequence

The *guaC* coding region is preceded by a potential ribosome-binding site (Shine–Dalgarno sequence; Gold *et al.*, 1981) and the proposed translational start site gives a relatively high score when analysed using the PERCEPTION algorithm of Stormo *et al.* (1982). The codon usage of the *guaC* gene (Table 1) shows that a small proportion of modulatory codons (1.2%) are used, and the even distribution of optimal energy codons (50%) in the diagnostic set suggests that *guaC* is moderately expressed (Grosjean & Fiers, 1982).

A search for putative *E. coli* promoter sequences in the region upstream of the proposed *guaC* structural gene was made using the ANALYSEQ program (Staden, 1984) which utilizes a weight matrix derived from the promoter sequences compiled by Hawley & McClure (1983). Three relatively high-scoring promoters, P1, P2 and P3 in decreasing score order, were detected (Fig. 2).

Stringently controlled promoters are characterized by a conserved G+C-rich sequence known as the discriminator, which is situated between the Pribnow box (−10 sequence) and the transcriptional start site. Consensus sequences for the discriminator regions of stable RNA promoters (GCGCC-C; −7 to −1) and ribosomal protein promoters ($G{}^{C}/_{G}{}^{C}/_{G}{}^{C}/_{G}$-${}^{C}/_{G}$-T; −5 to +4) have been defined (Travers, 1984). Discriminator sequences matching those of stable RNA genes are associated with two of the putative *guaC* promoters, P1 and P3 (Fig. 2). It is interesting that a discriminator-like sequence has also been detected for the *guaBA* promoter (Thomas & Drabble, 1985). This indicates that both the *guaC* and *guaBA* operons may be subject to the stringent response, as are other genes involved in the biosynthesis of nucleic acid precursors (Turnbough, 1983; Stayton & Fromm, 1977; Bouvier *et al.*, 1984). It may also be significant that the putative *guaBA* and *guaC*(P1) transcripts contain a common sequence at or near their 5′ ends: ATT${}^{G}/_{T}$ATTA (co-ordinates 135–142 in Fig.

```
EcoRI                                               P2
GAATTCATCATGATTATCAAAACGTTAAAAATGAGTGCACGAAAGCGAAATTGATGAAACGTTCGCTCACTATTTACCAGGTAAATTTAT
        10        20 P1     30              D1            60                  80        90
                                               P3                  D3
GGGATTGTAGCGTAAAAAAAGACAATTTCGCAGTCTTGCGCCGCATTGATTAGTGCGTATGATAGCGTCACTGGAGTTGCGCTCTTACCC
       100       110       120       130       140       150       160       170       180
                              1             BglII       10                            20
                 RBS         fM  R  I  E  E  D  L  K  L  G  F  K  D  V  L  I  R  P  K  R
TTATAGCCATTAACCCCAGGAATCCGCACATGCGTATTGAAGAAGATCTGAAGTTAGGTTTTAAAGACGTTCTCATCCGCCCTAAACGCT
       190       200       210       220       230       240       250       260       270
                          30                            40                            50
S  T  L  K  S  R  S  D  V  E  L  E  R  Q  F  T  F  K  H  S  G  Q  S  W  S  G  V  P  I  I
CCACTCTTAAAAGCCGTTCCGATGTTGAACTGGAACGTCAATTCACCTTCAAACATTCAGGTCAGAGCTGGTCCGGCGTGCCGATTATCG
       280       290       300       310       320       330       340       350       360
                          60                            70                            80
A  A  N  M  D  T  V  G  T  F  S  M  A  S  A  L  A  S  F  D  I  L  T  A  V  H  K  H  Y  S
CCGCAAATATGGACACCGTAGGCACATTTTCTATGGCCTCTGCGCTGGCTTCTTTTGATATTTTGACTGCTGTGCATAAACACTATTCTG
       370       380       390       400       410       420       430       440       450
                          90                            100                           110
V  E  E  W  Q  A  F  I  N  N  S  S  A  D  V  L  K  H  V  M  V  S  T  G  T  S  D  A  D  F
TCGAAGAGTGGCAAGCGTTTATCAACAATTCTTCCGCTGATGTGCTGAAACATGTGATGGTTTCTACCGGTACGTCTGATGCGGATTTCG
       460       470       480       490       500       510       520       530       540
                          120                           130                           140
E  K  T  K  Q  I  L  D  L  N  P  A  L  N  F  V  C  I  D  V  A  N  G  Y  S  E  H  F  V  Q
AAAAAACTAAACAGATTCTCGACCTGAACCCGGCATTAAACTTCGTTTGTATTGACGTGGCGAATGGTTATTCCGAACACTTCGTGCAGT
       550       560       570       580       590       600       610       620       630
                          150                           160                           170
F  V  A  K  A  R  E  A  W  P  T  K  T  I  C  A  G  N  V  V  T  G  E  M  C  E  E  L  I  L
TCGTTGCGAAAGCGCGTGAAGCGTGGCCGACCAAAACCATTTGTGCTGGTAACGTAGTGACTGGTGAAATGTGTGAGGAGCTTATCCTCT
       640       650       660       670       680       690       700       710       720
        EcoRV             180                           190                           200
S  G  A  D  I  V  K  V  G  I  G  P  G  S  V  C  T  T  R  V  K  T  G  V  G  Y  P  Q  L  S
CAGGTGCCGATATCGTTAAAGTTGGCATTGGCCCAGGTTCTGTTTGTACAACTCGCGTCAAAACAGGCGTCGGTTATCCGCAACTTTCTG
       730       740       750       760       770       780       790       800       810
                          210                           220                           230
A  V  I  E  C  A  D  A  A  H  G  L  G  G  M  I  V  S  D  G  G  C  T  T  P  G  D  V  A  K
CGGTAATCGAATGTGCCGATGCTGCGCACGGTCTGGGCGGAATGATCGTCAGCGATGGTGGCTGCACCACGCCGGGCGATGTGGCGAAAG
       820       830       840       850       860       870       880       890       900
                          240   SphI                    250                           260
A  F  A  R  A  D  F  V  M  L  G  G  M  L  A  G  H  E  E  S  G  G  R  I  V  E  E  N  G  E
CCTTTGCGCGTGCCGATTTCGTCATGCTTGGCGGCATGCTGGCGGGCCACGAAGAGAGCGGCGGTCGCATCGTTGAGGAGAACGGCGAGA
       910       920       930       940       950       960       970       980       990
                      SstI                              280                           290
K  F  M  L  F  Y  G  M  S  S  E  S  A  M  K  R  H  V  G  G  V  A  E  Y  R  A  A  E  G  K
AATTTATGCTGTTCTACGGCATGAGCTCCGAGTCTGCGATGAAACGTCACGTTGGCGGCGTTGCGGAATATCGCGCAGCAGAAGGTAAAA
      1000      1010      1020      1030      1040      1050      1060      1070      1080
                          300                           310                           320
T  V  K  L  P  L  R  G  P  V  E  N  T  A  R  D  I  L  G  G  L  R  S  A  C  T  Y  V  G  A
CCGTTAAGCTGCCGCTGCGAGGCCCGGTTGAAAATACCGCGCGAGATATTTTGGGCGGCCTGCGTTCAGCTTGTACATACGTTGGGGCTT
      1090      1100      1110      1120      1130      1140      1150      1160      1170
                          330                           340            346
S  R  L  K  E  L  T  K  R  T  T  F  I  R  V  Q  E  Q  E  N  R  I  F  N  N  L  *
CACGCCTGAAAGAGCTGACCAAGCGCACCACGTTTATTCGTGTGCAGGAACAAGAAAACCGCATCTTCAACAACCTGTAATCTCCCAACG
      1180      1190      1200      1210      1220      1230      1240      1250      1260

CTGGCGTGGAGCAACACGCCACGGTTATCCCATCCCACTCATCGCATCGCCTAAATGGAAAATTGGCAGATACATTGCCACCACCAGCGT
      1270      1280      1290      1300      1310      1320      1330      1340      1350
                    BclI
ACCAATAATTCCTCCCGTTATGATCAGCAACGCGGTTCAGTAAGGCTGCGAGGTTATCCGCCAGCGCCATTGTGTTTTCCCGATGATGAT
      1360      1370      1380      1390      1400      1410      1420      1430      1440

GGGCGAGGTTGTCTAACATGAGATCCAGAGAGCCGGATGCCTCTCCTGTTCTCACTAATTGCAAACAGAGCGGGCTAAACTCACCGGTAT
      1450      1460      1470      1480      1490      1500      1510      1520      1530

TTTTTAGCGCCAGCCAGATGGGTTGACCGTTACTGATATCGTGCTGGATTTGTGTCAGAAGTTGCACCCAGTACGGGCAGCGCATTGTTT
      1540      1550      1560      1570      1580      1590      1600      1610      1620
                                                                  BglII
CTCTGACGCTCTCTACGCCCTGTAAAAAAGTAATGCCTGCACTTTGTGTCAGCGCCAGAATCGTAAAGATCTGCGTGAGTTTTTGTCCCC
      1630      1640      1650      1660      1670      1680      1690      1700      1710

GCATCAGTGAACCCATAATCGGGATGCGTAACAGCAATTTCTGCCGCACTATAAGCCAGGTCGGTCGGCGCATCAGCAACTTATTGGCTA
      1720      1730      1740      1750      1760      1770      1780      1790      1800

TCGCCAGCAGAAAGCCGAACACACCAGCAGCCAGCTCCATTCGCCACTAAAGTCTGCCAGCGTCATGATCCCCTGCGTTAGTGCCGGTAG
      1810      1820      1830      1840      1850      1860      1870      1880      1890

TGGGGTGTTGAAGGTCTTATAGATAGCGGCAAACTCCGGCAGACACAAAATGCAGCATTGCCACAACCACCATGATTAGCCATCGCTAAA
      1900      1910      1920      1930      1940      1950      1960      1970      1980

ATGATGATGGG
      1990
```

Table 1. Codon usage in the *guaC* gene

The codon pairs enclosed in boxes are those whose use varies between strongly and weakly expressed genes, and asterisks denote potential modulatory codons (Grosjean & Fiers, 1982).

UUU Phe 7	UCU Ser 10	UAU Tyr 4	UGU Cys 6
UUC Phe 9	UCC Ser 6	UAC Tyr 2	UGC Cys 1
UUA Leu 2	UCA Ser 4	UAA End 1	UGA End 0
UUG Leu 2	UCG Ser 0	UAG End 0	UGG Trp 3
CUU Leu 4	CCU Pro 1	CAU His 3	CGU Arg 8
CUC Leu 3	CCC Pro 0	CAC His 5	CGC Arg 8
*CUA Leu 0	CCA Pro 1	CAA Gln 4	*CGA Arg 2
CUG Leu 14	CCG Pro 7	CAG Gln 4	*CGG Arg 0
AUU Ile 9	ACU Thr 5	AAU Asn 4	AGU Ser 0
AUC Ile 9	ACC Thr 10	AAC Asn 8	AGC Ser 5
*AUA Ile 0	ACA Thr 4	AAA Lys 17	*AGA Arg 0
AUG Met 11	ACG Thr 3	AAG Lys 3	*AGG Arg 0
GUU Val 14	GCU Ala 7	GAU Asp 12	GGU Gly 13
GUC Val 5	GCC Ala 6	GAC Asp 4	GGC Gly 19
GUA Val 3	GCA Ala 4	GAA Glu 16	*GGA Gly 1
GUG Val 9	GCG Ala 15	GAG Glu 9	*GGG Gly 1

2). No CRP-binding site (Busby, 1986) is apparent in the region upstream of the *guaC* structural gene, which is consistent with previous observations that *guaC* is not subject to catabolic repression, even though induction is thought to require cyclic AMP (Benson *et al.*, 1971). This contrasts with the *guaBA* operon where there is some evidence for catabolite repression (Nijkamp, 1969) and there is a good cyclic AMP receptor protein (CRP)-binding site in the promoter region (ACATGTGA-GCGAGATCAAATTC, co-ordinates 126–147; Thomas & Drabble, 1985), although this was not reported previously.

There are several regions of hyphenated dyad symmetry that could form stable stem-and-loop structures in RNA transcripts and the most significant of these [$\Delta G < -5.0$ kcal/mol (-21 kJ/mol); Tinoco *et al.*, 1973] are indicated in Fig. 2. A strong potential hairpin structure [co-ordinates 1262–1281; ΔG -19.0 kcal/mol (-80 kJ/mol)] is located 7 bp downstream of the stop codon for the *guaC* structural gene, where it could function as a transcriptional terminator. However, it lacks the typical run of T(U) nucleotides associated with *rho*-independent terminators (Rosenberg & Court, 1979), but it could function as a stabilization structure protecting the 3′ end of the transcript. The *guaBA* promoter region possesses a sequence of imperfect dyad symmetry centred within the proposed discriminator and extending over 18–24 bp (Thomas & Drabble, 1985). The *guaC* promoter region contains a comparable 29 bp segment centred within the discriminator associated with the putative promoter P1. These sequences possess very little sequence homology, but they could represent operators at which independent *guaC* and *guaBA* repressors bind.

Table 2. The predicted amino acid composition of GMP reductase

The predicted amino acid composition of GMP reductase including the initiating methionine is compared with that for IMP dehydrogenase. The IMP dehydrogenase composition is derived from the *guaB* sequence of Tiedeman & Smith (1985) except that the translational initiation start identified by Thomas & Drabble (1985) is used.

	GMP reductase		IMP dehydrogenase	
Amino acid	Number of residues	% by wt.	Number of residues	% by wt.
Asp	16	4.92	23	5.11
Asn	12	3.66	11	2.42
Thr	22	5.94	33	6.44
Ser	25	5.81	27	4.54
Glu	25	8.62	40	9.97
Gln	8	2.74	15	3.71
Pro	9	2.33	16	3.00
Gly	34	5.18	52	5.73
Ala	32	6.07	52	7.13
Val	31	8.21	48	9.18
Met	11	3.85	13	3.29
Ile	18	5.44	32	6.99
Leu	25	7.56	37	8.08
Tyr	6	2.62	9	2.83
Phe	16	6.29	9	2.56
Lys	20	6.85	23	5.69
His	8	2.93	10	2.65
Arg	18	7.51	32	9.65
Cys	7	1.93	5	1.00
Trp	3	1.49	0	0.00
Total	346		487	

The primary structure of GMP reductase and homology with IMP dehydrogenase

The primary structure of GMP reductase deduced from the nucleotide sequence is shown in Fig. 2 and the amino acid composition is summarized in Table 2. The composition resembles that of a typical soluble protein, but it has significantly more histidine and phenylalanine and less alanine than an average *E. coli* protein (Schulz & Schirmer, 1979). The sequence has been compared with four databases (GenBank, Claverie PGTrans, PIR and Doolittle) using the PEPSCAN and PEPSCORE programs (Bishop, 1984), and specifically with enzymes involved in G and purine nucleotide

Fig. 2. Nucleotide sequence of the *guaC* gene and primary structure of its product

The nucleotide sequence of 1991 bp of the non-transcribed strand of the *guaC* gene plus flanking regions is presented in the 5′–3′ direction. The nucleotide co-ordinates are assigned relative to the first base of the *Eco*RI site (E_1). The primary structure of the *guaC* gene product is shown above the nucleotide sequence. The region labelled RBS represents a potential ribosome-binding site. Putative promoter sites are denoted P1, P2 and P3, and the corresponding -35, -10 and transcriptional start sites are indicated by open arrows, filled arrows and open boxes (respectively) above the nucleotide sequence. The boxed regions marked D1 and D2 identify potential discriminator sequences associated with P1 and P3. The translational initiation site is underscored, and the stop site is denoted by an asterisk. Regions of hyphenated dyad symmetry capable of forming stable stem–loop structures (see text) are underscored by converging arrows. Key restriction sites are indicated.

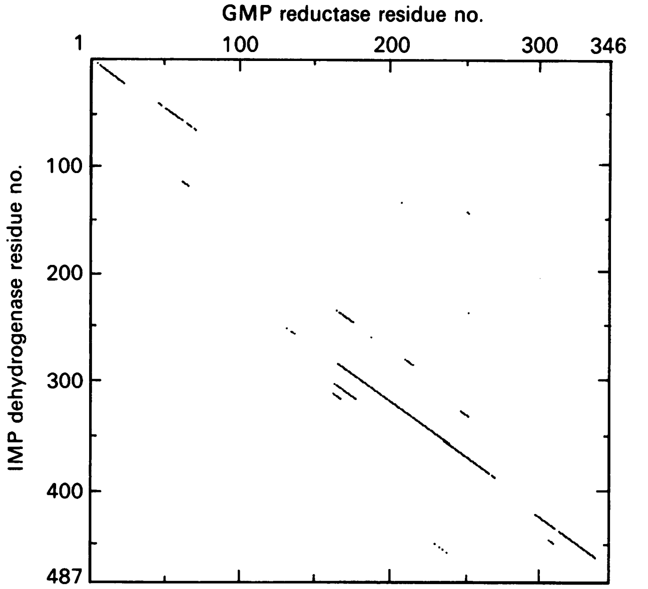

Fig. 3. Comparison matrices for the amino acid sequences of the *guaC* and *guaB* gene products

The matrices show pairwise comparisons for the *guaC* (GMP reductase) and *guaB* (IMP dehydrogenase) gene products of *E. coli* K12. The proportional option of the DIAGON program (Staden, 1982) was used and the dots correspond to the mid-points of 25-residue spans giving a double matching probability of $\leqslant 0.0005$ (McLachlan, 1971).

metabolism (IMP dehydrogenase, GMP synthetase, guanine phosphoribosyltransferase and CTP synthetase) and with enzymes containing NAD(P)- and nucleotide-binding sites (lipoamide dehydrogenase, glutathione reductase, succinyl-CoA synthetase, adenylate kinase and glutamate dehydrogenase).

No sustained homologies were detected except with the sequence of IMP dehydrogenase (EC 1.2.1.14) that was deduced from the nucleotide sequence of the *guaB* gene (Tiedeman & Smith, 1985; Thomas & Drabble, 1985). This is apparent from the comparison matrix obtained with the computer program DIAGON which detects good homology in the *N*-terminal regions and throughout the *C*-terminal halves of the two sequences (Fig. 3). An alignment based on the DIAGON comparisons is shown in Fig. 4. Apart from one large insertion of 123 amino acid residues, which could be placed anywhere between residues 82 and 119 in GMP reductase, very few insertions or deletions were needed to optimize the alignment. In the alignment shown, some 34% of the 335 equivalenced residues are identical, and the homology increases to 54% when conservative substitutions at the scoring limit $\geqslant 0.1$ in MDM$_{78}$ (Schwartz & Dayhoff, 1978) are included. A comparison of the hydropathy profiles confirms that GMP reductase and IMP dehydrogenase have homologous *N*-terminal and *C*-terminal segments of approx. 110 and 220 residues, respectively (Fig. 5). However, there is an internal 123-residue segment in IMP dehydrogenase that has no counterpart in GMP reductase. Secondary structure predictions using a combination of methods (Eliopoulos *et al.*, 1982) further indicates that the homologous regions are based on similar structural elements, and that the minor insertions/deletions occur where turns or coils are predicted.

IMP dehydrogenase catalyses the NAD$^+$-dependent conversion of IMP to XMP and the reaction is inhibited by GMP (Magasanik *et al.*, 1957; Gilbert *et al.*, 1979). It resembles GMP reductase in having affinities for the same nucleotides, IMP and GMP, and in using an analogous pyridine nucleotide coenzyme (NAD$^+$ not NADPH) so the relatively high degree of homology is not surprising. IMP dehydrogenase is known to have a cysteine residue at its IMP-binding site (Gilbert & Drabble, 1980), and it may be significant that only one of the seven or five cysteine residues is conserved in both sequences (Fig. 4). Furthermore, these residues (Cys-186 in GMP reductase and Cys-304 in IMP dehydrogenase) are located in the most highly conserved segments of the two polypeptide chains (positions 175 to 203 in GMP reductase, Fig. 4). It is tempting to speculate that these segments represent at least part of the GMP- and IMP-binding sites of GMP reductase and IMP dehydrogenases, and that the conserved cysteine residue is required for IMP-binding in IMP dehydrogenase. It may also be significant that the conserved cysteine residues are located in sequences, GS$^{V}/_{I}$ CT, which resemble one that contains the essential cysteine of pig lactate dehydrogenase (GSGCN) and others that are conserved in the *E. coli* enzyme, GSSCI and GGICN (Campbell *et al.*, 1984). Other potentially-important residues are two conserved histidine residues in the *N*-terminal segments, five conserved methionine residues, and the histidine, methionine and cysteine residues in the unique segment of IMP dehydrogenase (Fig. 4).

The AMP-binding sites of NAD(P) and FAD enzymes are often associated with the β_A–α_A–β_B segment of a Rossman fold and a G-X-G-X$_2$-$^{G}/_{A}$-X$_{10}$-G consensus (Rice *et al.*, 1984). Using the DIAGON program good homology is detected between the corresponding segment of the FAD-binding fold of human glutathione reductase (residues 23–50) and the most highly conserved and cysteine-containing segments of GMP reductase and IMP dehydrogenase (*a* in Fig. 4). This region may thus form part of a nucleotide-binding site, although the desired structural elements are not strongly predicted. An adjoining segment (*b* in Fig. 4) likewise shows good homology with a segment of the NADP-binding site of the *E. coli* glutamate dehydrogenase (residues 193–216; McPherson & Wootton, 1983) that specifies another part of the Rossman fold (D. W. Rice, personal communication). The AMP-binding pockets of adenylate kinase and related enzymes are associated with a $^{G}/_{A}$-X$_4$-G-K-$^{T}/_{G}$ consensus (Buck *et al.*, 1985), but apart from an unconserved GKT motif in GMP reductase (positions 289–291, Fig. 4) no such consensus can be detected. It is therefore difficult to identify the nucleotide- and coenzyme-binding sites from the primary structures; indeed, they may be formed by contributions from different subunits of a multimeric protein. In this context, the GMP reductase of *S. typhimurium* and the IMP dehydrogenase of *E. coli* are tetrameric enzymes containing identical subunits (Neuhard & Nygaard, 1987; Gilbert *et al.*, 1979), and thus it would appear that the observed conservation of primary and secondary structures may extend to the quaternary level.

```
         1       10        20        30        40        50
GMPR:    MRIEEDLKLGFKDVLIRPKRSTLKSRSDVELERQFTFKHSGQSWSGVPIIAANMDTVGT
         |**  |  * * ***| * |**|   |  |*  *|*          |*|||* ****
IMPD:   MLRIAKEA-LTFDDVLLVPAHSTVLPNT-ADLSTQLTKTIR----LNIPMLSAAMDTVTE
        1        10        20         30            40        50

       60        70        80        90       100            110
        FSMASALASFDILTAVHKHYSVEEWQAFINNSSADVLKHVMVS-----TGTSDADFEKTK
          |* ***    |  |**| *|*     |      * **         *     | |  *
        ARLAIALAQEGGIGFIHKNMSIERQAEEV----RRVKKHESGVVTDPQTVLPTTTLREVK
            60        70        80            90       100       110

              <-- c -->
        QILD--------------------------------------------------------
        || |
        ELTERNGFAGYPVVTEENELVGIITGRDVRFVTDLNQPVSVYMTPKERLVTVREGEAREV
               120       130       140       150       160       170

                                                         <-- d -->
        ------------------------------------------------------------

        VLAKMHEKRVEKALVVDDEFHLIGMITVKDFQKAEAKPNACKDEQGRLRVGAAVGAGAGN
               180       190       200       210       220       230

              120       130       140       150       160       170
        -------LNPALNFVCIDVANGYSEHFVQFVAKAREAWPTKTICAGNVVTGEMCEELILS
               | |||| | ** ||* **  |* |  |* * *   * |*** *|     *  |
        EERVDALVAAGVDVLLIDSSHGHSEGVLQRIRETR-AKPDLQIIGGNVATAAGARALAEA
               240       250       260        270       280
          <-------------- a -------------->          <--------- b --
              180       190       200       210       220       230
        GADIVKVGIGPGSVCTTRVKTGVGYPQLSAVIECADAAHGLGGMIVSDGGCTTPGDVAKA
        *   *********|****| **** **||** |  |*  * *  |||***    **|***
        GCSAVKVGIGPGSICTTRIVTGVGVPQITAVADAVEALEGTGIPVIADGGIRFSGDIAKA
      290       300       310       320       330       340
        --------->
               240       250       260       270       280
        FAR-ADFVMLGGMLAGHEESGGRIVEENGEKFMLFYGMSSESAMKRHV-GGVAEYRAAEG
        |*  *  **|*|**** *** * *    *  |  | **|* |** |       |   *
        IAAGASAVMVGSMLAGTEESPGEIELYQGRSYKSYRGMGSLGAMSKGSSDRYFQSDNAAD
      350       360       370       380       390       400

      290             300       310       320       330       340
        KTVKLPLR------GPVENTARDILGGLRSACTYVGASRLKELTKRTTFIRVQEQENRIF
        * *   |       * | |  || |*****     *   | **  || *|*|     |
        KLVPEGIEGRVAYKGRLYEIIHQQMGGLRSCMGLTGCGTIDELRTKAEFVRISGAGIQES
      410       420       430       440       450       460

        346
        NNL                  :GMPR
        |
        HVHDVTITKESPNYRLGS   :IMPD
      470       480    487
```

Fig. 4. Alignment of the amino acid sequences of GMP reductase (GMPR) and IMP dehydrogenase (IMPD)

The sequences have been aligned for maximum homology based on the DIAGON comparisons (Fig. 4). The asterisks signify absolutely conserved residues between the sequences whereas vertical bars indicate conserved substitutions scoring ⩾ 0.1 in the MDM_{78} mutation data matrix (Schwartz & Dayhoff, 1978). The conserved cysteine, histidine and methionine residues are boxed, components of a putative nucleotide binding site(s) are marked *a* and *b* (see text), and two potential interdomain linkers (*c* and *d*) are underlined in the IMPD sequence.

The homology between the reductase and dehydrogenase suggests that they share common ancestors, but it is not clear whether the ancestral *guaC* gene has suffered a deletion or whether an insertion has occurred in the *guaB* precursor. Structural domains within proteins are often composed of amino acid residues that are consecutively arranged in the polypeptide chain and ancestrally-related proteins often possess amino acid homologies that correspond to entire structural domains rather than parts thereof. Consequently, it is possible that the *N*- and *C*-terminal regions represent independently-folding domains, and that the extra 123-residue segment of IMP dehydrogenase forms an additional domain that is unique to this enzyme (see Fig. 6). It may also be significant that the central segment of the IMP dehydrogenase sequence is flanked by two relatively hydrophobic segments of polypeptide that are unusually rich in glycine and alanine (*c* and *d* in Fig. 4). These segments resemble the four (alanine + proline)-rich interdomain linkers that occur in the dihydrolipoamide acetyltransferase component of the pyruvate dehydrogenase complex (Radford *et al.*, 1987), and the polypeptide sequence that links individual domains in the tryptophan synthetase *β*-subunit (Crawford *et al.*,

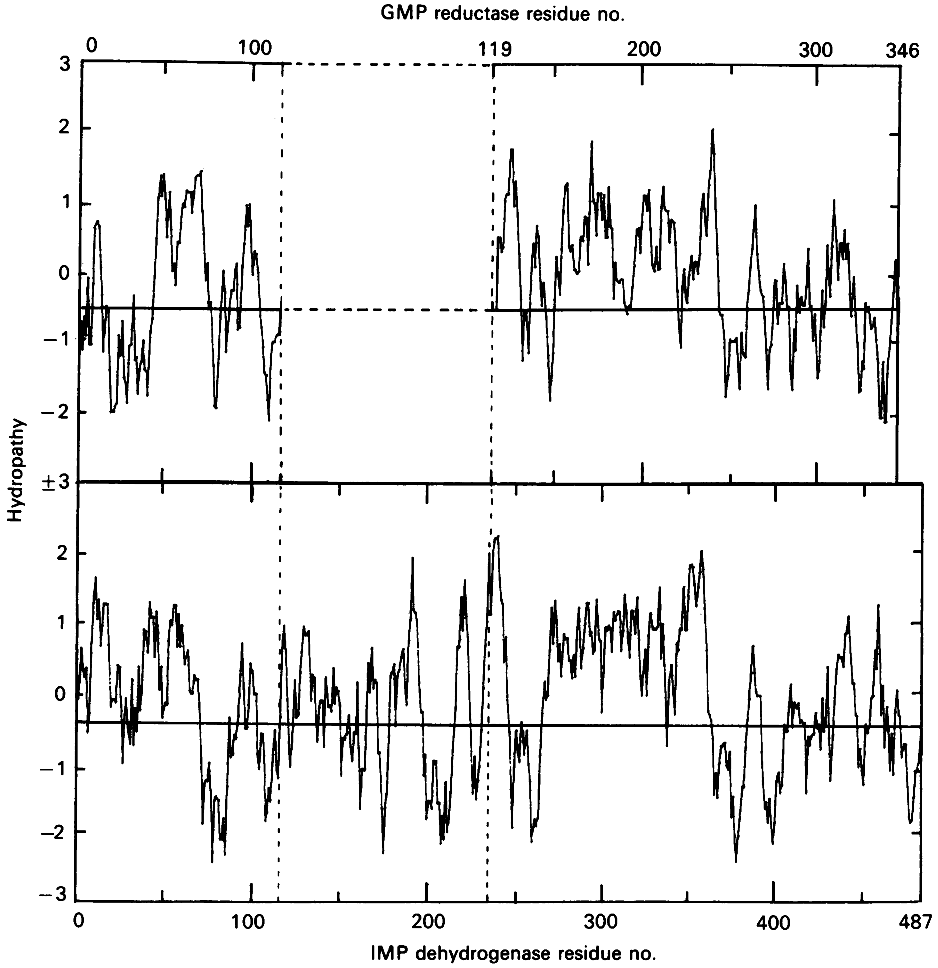

Fig. 5. Hydropathy profiles of GMP reductase and IMP dehydrogenase

The hydropathy profile of GMP reductase (upper panel) is interupted at residue 118 to illustrate its homology with the *N*- and *C*-terminal segments of the IMP dehydrogenase profile (lower panel). Consecutive hydropathy averages are plotted at the mid-points of a 9-residue segment as it advances from *N*- to *C*-terminus. Relative hydrophobicity and hydrophilicity are recorded in the range +3 to −3 and a horizontal line representing the average for most sequenced proteins is included (Kyte & Doolittle, 1982).

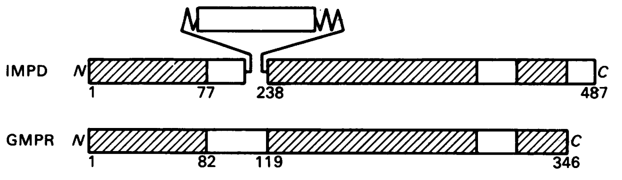

Fig. 6. Schematic representations of the structures of GMP reductase (GMPR) and IMP dehydrogenase (IMPD)

The homologous segments (hatched regions) and the domain structures of the two enzymes that have been inferred from the amino acid sequences are illustrated. The putative (alanine+glycine)-rich interdomain linkers (zig-zag lines) flank the domain that is unique for IMP dehydrogenase.

1980) and the connectors in multifunctional enzymes (Zalkin *et al.*, 1984). It is therefore conceivable that the (alanine+glycine)-rich sequences define the boundaries of the additional domain in IMP dehydrogenase (Fig. 6). The function of this putative domain is not known, but the dehydrogenase differs from the reductase in catalysing the first unique step in the GMP biosynthetic pathway, so it could be a regulatory domain that mediates the allosteric inhibition of IMP dehydrogenase by the end-product, GMP (Buzzee & Levin, 1968). The proposed domain structure and multimeric quaternary structure are consistent with the observed enzyme complementation of a *guaB* mutant by a 96-residue *N*-terminal polypeptide (Thomas & Drabble, 1985), which could specify a functional *N*-terminal domain.

These studies provide the first amino acid sequence of a GMP reductase from any source. As a result, a very interesting structural relationship between GMP reductase and IMP dehydrogenase has emerged, and this could well merit further studies to interpret the relationship at the functional level.

We thank Ian K. Duckenfield for assistance with computer analysis and Dr. D. W. Rice for helpful discussions. This work was supported by the Wellcome Trust.

REFERENCES

Bachmann, B. J. (1983) Microbiol. Rev. **47**, 180–230

Benson, C. E. & Gots, J. S. (1975) Biochim. Biophys. Acta **403**, 47–57

Benson, C. E., Brehmeyer, B. A. & Gots, J. S. (1971) Biochem. Biophys. Res. Commun. **43**, 1089–1094

Biggin, M. D., Gibson, T. J. & Hong, G. F. (1983) Proc. Natl. Acad. Sci. U.S.A. **80**, 3963–3965

Bishop, M. (1984) BioEssays **1**, 29–31

Bouvier, J., Patte, J.-C. & Stragier, P. (1984) Proc. Natl. Acad. Sci. U.S.A. **81**, 4139–4143

Buck, D., Spencer, M. E. & Guest, J. R. (1985) Biochemistry **24**, 6245–6252

Busby, S. (1986) Symp. Soc. Gen. Microbiol. **39**, 51–77

Buzzee, D. H. & Levin, A. P. (1968) Biochem. Biophys. Res. Commun. **30**, 673–677

Campbell, H. D., Rogers, B. L. & Young, I. G. (1984) Eur. J. Biochem. **144**, 367–373

Chye, M.-L., Guest, J. R. & Pittard, J. (1986) J. Bacteriol. **167**, 749–753

Crawford, I. P., Nichols, B. P. & Yanofsky, C. (1980) J. Mol. Biol. **142**, 489–502

Eliopoulos, E. E., Geddes, A. J., Brett, M., Pappin, D. J. C. & Findlay, J. B. C. (1982) Int. J. Biol. Macromol. **4**, 263–268

Garber, B. B., Jochimsen, B. U. & Gots, J. S. (1980) J. Bacteriol. **167**, 749–753

Gilbert, H. J. & Drabble, W. T. (1980) Biochem. J. **191**, 533–541

Gilbert, H. J., Lowe, C. R. & Drabble, W. T. (1979) Biochem. J. **183**, 481–494

Gold, L., Pribnow, D., Schneider, T., Shinedling, S., Singer, B. S. & Stormo, G. (1981) Annu. Rev. Microbiol. **35**, 365–403

Gots, J. S., Benson, C. E., Jochimsen, B. U. & Koduri, K. R. (1977) Ciba Found. Symp. **48**, 23–41

Grosjean, H. & Fiers, W. (1982) Gene **18**, 199–209

Guest, J. R. & Stephens, P. E. (1980) J. Gen. Microbiol. **121**, 277–292

Guest, J. R., Roberts, R. E. & Stephens, P. E. (1983) J. Gen. Microbiol. **129**, 671–680

Hawley, D. K. & McClure, R. (1983) Nucleic Acids Res. **11**, 2237–2255

Kessler, A. I. & Gots, J. S. (1985) J. Bacteriol. **164**, 1288–1293

Kyte, J. & Doolittle, R. F. (1982) J. Mol. Biol. **157**, 105–132

Lindberg, F., Lindquist, S. & Normark, S. (1987) J. Bacteriol. **169**, 1923–1928

Magasanik, B., Moyed, H. S. & Gehring, L. B. (1957) J. Biol. Chem. **226**, 339–350

Mager, J. & Magasanik, B. (1960) J. Biol. Chem. **235**, 1474–1478

Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor

Marinus, M. G. & Morris, M. R. (1973) J. Bacteriol. **141**, 1143–1150

McLachlan, A. D. (1971) J. Mol. Biol. **61**, 409–424

McPherson, M. J. & Wootton, J. C. (1983) Nucleic Acids Res. **11**, 5257–5266

Mehra, R. K. & Drabble, W. T. (1981) J. Gen. Microbiol. **123**, 27–37

Messing, J. (1983) Methods Enzymol. **101**, 20–78

Moffat, K. G. & Mackinnon, G. (1985) Gene **40**, 141–143

Neuhard, J. & Nygaard, P. (1987) in *Escherichia coli* and *Salmonella typhimurium* Cellular and Molecular Biology (Neidhardt, F. C., Ingraham, J. L., Low, K. B., Magasanik, B., Schaechter, M. & Umbarger, E., eds.), vol. 1, pp. 445–473, American Society for Microbiology, Washington

Nijkamp, H. J. J. (1969) J. Bacteriol. **100**, 585–593

Nijkamp, H. J. J. & DeHaan, P. G. (1967) Biochim. Biophys. Acta **145**, 31–40

Radford, S. E., Laue, E. D., Perham, R. N., Miles, J. S. & Guest, J. R. (1987) Biochem. J. **247**, 641–649

Renart, M. F. & Sillero, A. (1974) Biochim. Biophys. Acta **341**, 178–186

Rice, D. W., Schulz, G. E. & Guest, J. R. (1984) J. Mol. Biol. **174**, 483–496

Roberts, R. E., Lienhard, C. I., Gaines, C. G., Smith, J. M. & Guest, J. R. (1988) J. Bacteriol. **170**, 463–467

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

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

Schulz, G. E. & Schirmer, R. H. (1979) Principles of Protein Structure, pp. 1–17, Springer-Verlag, New York

Schwartz, R. M. & Dayhoff, M. O. (1978) in Atlas of Proteins Sequence and Structure (Dayhoff, M. O., ed.), vol. 5, supplement 3, pp. 353–358, National Biomedical Research Foundation, Washington

Spector, T. & Jones, T. E. (1982) Biochem. Pharmacol. **31**, 3891–3897

Spector, T., Jones, T. E. & Miller, R. J. (1979) J. Biol. Chem. **254**, 2308–2315

Staden, R. (1979) Nucleic Acids Res. **6**, 2601–2611

Staden, R. (1980) Nucleic Acids Res. **8**, 3673–3694

Staden, R. (1982) Nucleic Acids Res. **10**, 2951–2961

Staden, R. (1984) Nucleic Acids Res. **12**, 505–519

Staden, R. & McLachlan, A. D. (1982) Nucleic Acids Res. **10**, 141–156

Stayton, M. M. & Fromm, H J. (1977) J. Biol. Chem. **254**, 2579–2581

Stephens, P. E., Lewis, H. M., Darlison, M. G. & Guest, J. R. (1983) Eur. J. Biochem. **135**, 519–527

Stephens, R. W. & Whittaker, V. K. (1973) Biochem. Biophys. Res. Commun. **53**, 975–981

Stormo, G. D., Schneider, T. D., Gold, L. M. & Ehrenfeucht, A. (1982) Nucleic Acids Res. **10**, 2997–3011

Thomas, M. S. & Drabble, W. T. (1985) Gene **36**, 45–53

Tiedeman, A. A. & Smith, J. M. (1985) Nucleic Acids Res. **13**, 1303–1316

Tinoco, I., Jr., Borer, P. N., Dengler, B., Levine, M. D., Uhlenbeck, O., Crothers, D. M. & Gralla, J. (1973) Nature (London) New Biol. **246**, 40–41

Travers, A. A. (1984) Nucleic Acids Res. **6**, 2605–2618

Turnbough, C. L. (1983) J. Bacteriol. **153**, 998–1007

Yanisch-Peron, C., Vieira, J. & Messing, J. (1985) Gene **33**, 103–119

Zalkin, H., Paluh, J. L., van Cleemput, M., Moye, W. S. & Yanofsky, C. (1984) J. Biol. Chem. **259**, 3985–3992

Received 12 January 1988; accepted 4 March 1988