Biochem. J. (1984) **222**, 519–534
Printed in Great Britain

Nucleotide sequence encoding the flavoprotein and hydrophobic subunits of the succinate dehydrogenase of *Escherichia coli*

David WOOD, Mark G. DARLISON,* Robin J. WILDE and John R. GUEST†
Department of Microbiology, Sheffield University, Western Bank, Sheffield S10 2TN, U.K.

(*Received 11 April 1984/Accepted 15 May 1984*)

The nucleotide sequence of a 3614 base-pair segment of DNA containing the *sdhA* gene, encoding the flavoprotein subunit of succinate dehydrogenase of *Escherichia coli*, and two genes *sdhC* and *sdhD*, encoding small hydrophobic subunits, has been determined. Together with the iron–sulphur protein gene (*sdhB*) these genes form an operon (*sdhCDAB*) situated between the citrate synthase gene (*gltA*) and the 2-oxoglutarate dehydrogenase complex genes (*sucAB*): $\overleftarrow{gltA}$-$\overrightarrow{sdhCDAB}$-$\overrightarrow{sucAB}$. Transcription of the *gltA* and *sdhCDAB* gene appears to diverge from a single intergenic region that contains two pairs of potential promoter sequences and two putative CRP (cyclic AMP receptor protein)-binding sites. The *sdhA* structural gene comprises 1761 base-pairs (587 codons, excluding the initiation codon, AUG) and it encodes a polypeptide of M_r 64268 that is strikingly homologous with the flavoprotein subunit of fumarate reductase (*frdA* gene product). The FAD-binding region, including the histidine residue at the FAD-attachment site, has been identified by its homology with other flavoproteins and with the flavopeptide of the bovine heart mitochondrial succinate dehydrogenase. Potential active-site cysteine and histidine residues have also been indicated by the comparisons. The *sdhC* (384 base-pairs) and *sdhD* (342 base-pairs) structural genes encode two strongly hydrophobic proteins of M_r 14167 and 12792 respectively. These proteins resemble in size and composition, but not sequence, the membrane anchor proteins of fumarate reductase (the *frdC* and *frdD* gene products).

Succinate dehydrogenase (SDH; EC 1.3.99.1) is a membrane-bound flavoenzyme that is present in all aerobic organisms. It functions in the tricarboxylic acid cycle, catalysing the oxidation of succinate to fumarate and donating reducing equivalents to the respiratory chain. Succinate dehydrogenases have been purified from several mitochondrial and bacterial sources and they invariably comprise two subunits, each containing iron–sulphur centres: the large flavoprotein subunit (M_r 65000–72000) containing covalently-bound FAD, and the smaller iron–sulphur protein subunit (M_r 25000–30000). Additionally, these subunits may be associated with two small polypeptides (M_r 13500–17000 and 7000–14000) or a cytochrome *b* (see review by Hederstedt & Rutberg, 1981).

The succinate dehydrogenase of *Escherichia coli* has proved refractory to extensive purification, but structural information has been obtained indirectly using mutants and immunochemical methods. Thus, the cytoplasmic membranes of *sdh* nonsense mutants were shown to lack a polypeptide of M_r 67000 corresponding to the flavoprotein subunit (Spencer & Guest, 1974*a*), and immunoprecipitates have been found to contain polypeptides of M_r 73000 and 26000 (Jones *et al.*, 1982) or M_r 71000, 31000, 29000 and 13000, plus a polypeptide of M_r 19000 which could be cytochrome b_{556} (Owen & Condon, 1982).

Studies with mutants lacking succinate dehydrogenase have shown that the *sdh* genes are located between the citrate synthase gene (*gltA*) and 2-oxoglutarate dehydrogenase complex genes (*sucA,B*) at 16.7min in the *E. coli* linkage map

Abbreviations used: SDH, succinate dehydrogenase; FRD, fumarate reductase; CRP, cyclic AMP receptor protein; kb, 1000 base pairs.

* Present address: Department of Biochemistry, Imperial College of Science and Technology, South Kensington, London SW7 2AZ, U.K.

† To whom reprint requests should be addressed.

(Creaghan & Guest, 1972). This segment of the chromosome is also thought to contain several tRNA genes and the carboxylic acid transport gene: *gltA-cbt*-lysT-lysT-valT*-supG*-sdh-sucA-sucB* (Bachmann, 1983). *E. coli* possesses a similar but genetically-distinct enzyme, fumarate reductase (FRD; EC 1.3.99.1) encoded by the *frd* genes at 94.4min (Spencer & Guest, 1973, 1974*b*). Fumarate reductase is the terminal reductase of an anaerobic energy-generating electron-transport chain that transfers electrons from α-glycerophosphate, lactate, formate, NADH or H_2 to fumarate with the production of succinate. Although both enzymes catalyse the same reaction, they exhibit different affinities for succinate and fumarate and functional interchange is precluded by the mechanisms controlling their synthesis. Thus, succinate dehydrogenase is induced aerobically and strongly repressed during anaerobic growth or by glucose, whereas fumarate reductase synthesis is derepressed during anaerobic growth and repressed by oxygen or nitrate (Spencer & Guest, 1973; Ruiz-Herrera & Garcia, 1972; Takahashi, 1975; Gray *et al.*, 1966; Wimpenny & Cole, 1967). However, the possibility that the two enzymes are the products of common ancestral genes that have been duplicated and coupled to different regulators to enable their functional diversification has always been an attractive speculation.

More is known about the *E. coli* fumarate reductase than the corresponding succinate dehydrogenase because the *frd* genes were more readily cloned (Cole & Guest, 1978, 1980*a*; Guest, 1981*a*; Lohmeier *et al.*, 1981) and because the enzyme has proved easier to isolate and characterize (Dickie & Weiner, 1979; Lemire *et al.*, 1982). Studies with the cloned *frd* genes, including a complete nucleotide sequence analysis, have revealed an operon containing four genes *frdABCD* (Cole, 1982; Cole *et al.*, 1982; Grundström & Jaurin, 1982). The operon exhibits considerable sequence economy because the *frdA* gene partially overlaps the *frdB* gene, the distal gene (*frdD*) contains the promoter for the adjacent β-lactamase gene (*ampC*), and the *frd* transcriptional terminator functions as an attenuator for the growth-rate-dependent expression of the β-lactamase. The genes encode the flavoprotein subunit (M_r 66052, *frdA*), the iron–sulphur protein subunit (M_r 27092, *frdB*) and two very hydrophobic subunits (M_r 15000, *frdC*; M_r 13100, *frdD*) that are essential for anchoring the soluble two-subunit form of the enzyme into the membrane but could perform other functions (Cole, 1984). The synthesis of fumarate reductase has been amplified by increasing the gene-dosage with multi-copy plasmids, replicating *λfrd* phages, and strains containing multiple duplications of the chromosomal genes. These studies have shown that the enzyme can accumulate in the cytoplasm once the binding capacity of the membrane has been exceeded (Cole & Guest, 1979*a*), and that synthesis of the enzyme can be derepressed aerobically, presumably by titration of a specific repressor by multiple copies of the *frd* operator (Cole & Guest, 1979*b*, 1980*b*). This aerobic derepression leads to partial replacement of succinate dehydrogenase function in *sdh* mutants (Guest, 1981*a*).

The *gltA-sucB* region of the *E. coli* chromosome has now been cloned in phage and plasmid vectors and the approximate positions of the tricarboxylic acid cycle genes have been defined relative to the restriction map (Guest, 1981*b*; Spencer & Guest, 1982; Guest *et al.*, 1984). This has provided evidence for the existence of at least two *sdh* genes, encoding the flavoprotein (*sdhA*) and the iron–sulphur protein (*sdhB*), and the transcriptional polarities of the tricarboxylic acid cycle genes have been defined by post-infection labelling: $\overleftarrow{gltA}$-$\overrightarrow{sdh(AB)}$-$\overrightarrow{sucAB}$.

The present paper reports the complete nucleotide sequence of a 3614 base-pair segment of DNA containing the *sdhA* gene, two genes designated *sdhC* and *sdhD* that encode small hydrophobic proteins comparable to the *frdC* and *frdD* gene products, the intergenic region between the *gltA* and *sdh* genes, and part of the *gltA* gene. The sequence of the *sdhB* gene will be reported separately (M. G. Darlison & J. R. Guest, unpublished work). The results show that the succinate dehydrogenase of *E. coli* is expressed from an operon of four genes, *sdhCDAB*, and the striking homology that is apparent between the flavoprotein (and iron–sulphur protein) subunits of succinate dehydrogenase and fumarate reductase provide clear evidence of a close ancestral relationship.

Experimental

Bacteria, phages and plasmids

Strains used for genetic complementation and for enzymological studies were: PL2024 (*sdh⁺frd⁺gal trpA rpsL*) and its derivatives JRG780 (*sdh⁺ frd-11*), JRG1502 (*sdh-0 recA*), JRG1557 (*sdh-9 recA*), JRG1503 (*sdh-0 frd-11 recA*). Other strains used for specific purposes were: AB2480 (*uvrA6 recA8*), the maxicell strain; ED8641 (*hsdR supE recA56 met*) and GM242 (*dam-3 recA1*), hosts for plasmid propagation; C600 for lambda phage propagation, and JM101 (*Δlac-pro supE thi*/F′*traD36 proAB lacIᵠZΔM15*), the transfection and propagation host for M13 phages.

The lambda transducing phages containing the *sdh* genes, λG117 (*λgltA sdh sucAB*) and λG119 (*λgltA sdh*) have been described previously (Spen-

cer & Guest, 1982) as have the ColE1-*gltA*$^+$ hybrid plasmids, pLC26-17 and pLC31-28 (Guest, 1981*b*).

Derivatives of plasmid pBR322 containing segments of the *sdh* region (pGS65 and pGS91) have been described previously (Hull *et al.*, 1983) and several others (pGS125, 126 and 127) were constructed from pGS65 during the course of the present work, by *in vitro* deletion of the small *Hin*dIII and *Xho*I-*Sal*I fragments. Plasmid pGS133 is analogous to pGS65 in containing the same 4500 base-pair *Bam*HI fragment of bacterial DNA from λG117 but inserted into the low copy vector pLG339 (Stoker *et al.*, 1982).

Complementation tests

Complementation of *sdh* lesions by specific plasmids was assessed by growth tests with Amp^R or Kan^R transformants on plates of succinate minimal medium and peptone medium (Spencer & Guest, 1982).

Enzymology

Succinate dehydrogenase specific activities were assayed according to Creaghan & Guest (1972) using exponential cultures grown in glucose-free L broth with ampicillin (50 μg/ml) or kanamycin (25 μg/ml).

Sources of DNA

The DNA fragments used for sequence analysis were obtained from several of the original transducing phages and the plasmid derivatives (Spencer & Guest, 1982; Hull *et al.*, 1983). The fragments used for 'shot-gun' cloning in M13 are shown in Fig. 1 and they include: the 3900 base-pair *Nru*I fragment of pGS91, N_1–N_p (co-ordinate 972 in pBR322); the 2400 base-pair *Hin*dIII–*Xho*I fragment of pGS65 (H–X); the 2700 base-pair *Hin*dIII–*Nru*I fragment of pGS65 (H–N_2) and the 760 base-pair *Sst*I–*Hin*dIII fragment of λG119 (St_1–H). Other specific fragments were cloned directly into M13 from digests of these fragments, their sources, or from λG117. Phage and plasmid DNAs were prepared as described previously (Guest *et al.*, 1983) and restriction fragments were isolated by electrophoresis in agarose gel followed by electroelution (McDonell *et al.*, 1977).

Cloning in M13 and transfection

The overall sequencing strategy involved a combination of 'shot-gun' cloning and directed cloning of specific fragments using several sources of DNA and appropriate cloning sites of M13mp8-11 (Messing & Vieira, 1982). The clones that were sequenced are indicated in Fig. 1. The 'shot-gun' clones were generated in M13mp8 as follows: ultrasonic fragments of N_1–N_p were ligated into the vector *Sma*I site according to Darlison *et al.* (1984); *Sau*3A fragments of H–X were cloned into the *Bam*HI site; and *Msp*I fragments of H–N_2 and St_1–H were cloned into the *Acc*I site. The specific fragments, their sources and the corresponding vectors, were cloned as follows: B_1–H (λG117) in M13mp8 and 9; B_1–RI (λG117) in M13mp8; St_1–M and M–H (*Msp*I digest of St_1–H) in M13mp10 and M13mp11, respectively; H–M (*Msp*I digest of H–N_2) in M13mp8; Rs_1–Rs_2 (*Rsa*I digest of H–X) in M13mp8; and H–N_2 was cloned in M13mp9. Transfection of *E. coli* JM101 was performed according to published procedures (Sanger *et al.*, 1980).

Nucleotide sequence analysis

Single stranded M13 DNA templates were prepared and sequenced by the dideoxy chain-termination method using a 17-nucleotide synthetic primer (Sanger *et al.*, 1977). All the clones were screened initially by 'A-tracking', to avoid generating redundant data, and the nucleotide sequences were compiled and analysed with the Staden computer programs (Staden, 1979, 1980, 1982; Staden & McLachlan, 1982).

Radioactive labelling of the polypeptide products of cloned genes

Plasmid-encoded polypeptides were labelled with [^{35}S]methionine and identified by polyacrylamide-gel electrophoresis and autoradiography using (i) the 'maxicell' procedure of Sancar *et al.* (1979) according to Shaw & Guest (1982) and (ii) the *in vitro* transcription–translation system described by Pratt *et al.* (1981). The latter method was also used with isolated DNA fragments. Labelled polypeptides were analysed by gel electrophoresis and autoradiography by using the sodium dodecyl sulphate/urea polyacrylamide gel system of Kadenbach *et al.* (1983) with acrylamide 15% (w/v) and a range of standard proteins of M_r 6200–116000.

Materials

Restriction endonucleases, T4 DNA ligase and T4 DNA polymerase were purchased from Bethesda Research and New England Biolabs, and DNA polymerase (Klenow fragment) from Boehringer Corp. All enzymes were used as directed by the suppliers. L-[^{35}S]Methionine, [α-^{32}P]dATP and the prokaryotic DNA-directed translation kit (N.380) were supplied by Amersham International.

Results and discussion

Gene–protein relationships

The approximate position of the *sdh* region relative to the corresponding restriction map was deduced from the properties of the original ColE1-*gltA*$^+$ plasmids (Guest, 1981*b*) and the λ*gltA* and

λ*gltA sdh* transducing phages (Spencer & Guest, 1982). These showed that the citrate synthase gene is situated in the 3.1 kb *Eco*RI–*Hin*dIII fragment (RI–H in Fig. 1) and, in order to obtain good *sdh* complementation, this segment has to be extended by at least 4 kb in the direction of the second *Bam*HI site (B_2 in Fig. 1). Further studies have confirmed this location because good nutritional complementation of *sdh recA* and *sdh frd recA* mutants was obtained with pGS65, which contains the *Bam*HI fragment (B_1–B_2), but not with pGS91, which overlaps this fragment only as far as the *Xho*I site (X in Fig. 1). The results suggest that succinate dehydrogenase is expressed from the 4.5 kb *Bam*HI fragment, but this interpretation is complicated by the fact that plasmids containing smaller inserts exhibit partial complementation (pGS125) or weak but significant complementation (pGS126 and pGS127). Likewise, enzymological studies with plasmid-containing derivatives of JRG1503 (*sdh frd recA*) have shown that, relative to the corresponding parental strain (JRG780, *sdh*$^+$ *frd*), good succinate dehydrogenase specific activities are expressed with pGS133 (200%) and pGS65 (50%), but significant activities are also detected with pGS125 (20%), pGS126 and pGS127 (10%) compared with pGS91 and pBR322 (2%). It would appear that these nutritional and enzymological tests are complicated by the presence of other promoters, such as the vector *tet* promoter, and partial complementation by the flavoprotein subunit, which is presumably encoded in the common *Hin*dIII–*Xho*I fragment (H–X in Fig. 1). Thus in order not to exclude any part of the *sdh* region it was decided to sequence the entire 5.0 kb *Nru*I–*Bam*HI segment (N_1–B_2 in Fig. 1). Subsequent work has shown that the two sites, N_1 and B_2, are located in the segments of DNA encoding the *N*-terminal ends of citrate synthase and 2-oxoglutarate dehydrogenase respectively (Hull *et al.*, 1983; Darlison *et al.*, 1984).

The polypeptides expressed from the 4.5 kb *Bam*HI segment (B_1–B_2 in Fig. 1) were detected by the maxicell procedure after transforming the maxicell strain (AB2480) with the series of plasmids shown in Fig. 1 and the corresponding vectors (pBR322 and pLG339). They were also detected by *in vitro* transcription–translation studies using plasmid DNA and isolated restriction fragments (B_1–B_2 and H–X). These studies showed that pGS65, pGS133 and the isolated *Bam*HI fragment (B_1–B_2) direct the synthesis of two polypeptides of estimated M_r 70000 and 26000 corresponding to the flavoprotein (SDHA) and iron–sulphur protein (SDHB) subunits (the *sdhA* and *sdhB* gene products), and two smaller polypeptides of M_r 13500 (SDHC) and 11500 (SDHD), the products of two previously unidentified genes designated *sdhC* and *sdhD*. The approximate positions of the coding regions were deduced by observing that SDHB is not expressed by pGS91, pGS125, pGS127 or the H–X fragment, and that SDHC could not be detected with pGS126, pGS127 or the H–X fragment. These results are consistent with the ultimate gene locations shown in Fig. 1.

Nucleotide sequence and location of coding regions

The overall strategy for sequencing the *sdh* region involved a combination of 'shot-gun' and directed cloning of several different fragments of a 5.0 kb segment of bacterial DNA (see the Experimental section). The M13 clones used for sequencing the 3614 base-pair *Nru*I–*Xho*I fragment (N_1–X) containing the *gltA–sdhA* segment are summarized in Fig. 1 and the complete nucleotide sequence is presented in Fig. 2. The sequence is fully-overlapped, and except for part of the *gltA* region, all of it was derived from both DNA strands and each segment from at least two independent clones.

The coding regions were detected with the computer program FRAMESCAN (Staden & McLachlan, 1982) using the *frd* and other *E. coli* genes as standards. Three reading frames having the polarity that had been defined for the *sdh* genes (Spencer & Guest, 1982), were found at positions 1091–1477m 1474–1818 and 1821–3584 (Fig. 2). They all exhibited consistently high scores with respect to codon usage and the largest, encoding a polypeptide of M_r 64268, corresponds in size and position to the *sdhA* structural gene. The others, encoding hydrophobic polypeptides of M_r 14167 and 12792, were tentatively designated *sdhC* and *sdhD* because of their similarity to the *frdC* and *frdD* structural genes (Grundström & Jaurin, 1982). The larger, which is encoded by the DNA flanking the *Hin*dIII site, clearly corresponds to the M_r 13500 polypeptide detected in the transcription–translation studies, and the smaller corresponds to the M_r 11500 polypeptide expressed by all of the plasmids. Only two coding regions with the opposite polarity were detected in the complementary sequence. One corresponds to the proximal segment of the *gltA* structural gene, co-ordinates 383–1 (Fig. 2). The other forms part of a long open reading frame (co-ordinates 3497–1793) which is mainly in the *sdhA* complement. Only the 3488–2682 segment exhibits a typical *E. coli* codon usage and there is no evidence that this potential coding region is expressed.

Features of the nucleotide sequence and organization of the sdh genes

The main feature of the nucleotide sequence is the *sdhA* structural gene encoding a polypeptide of

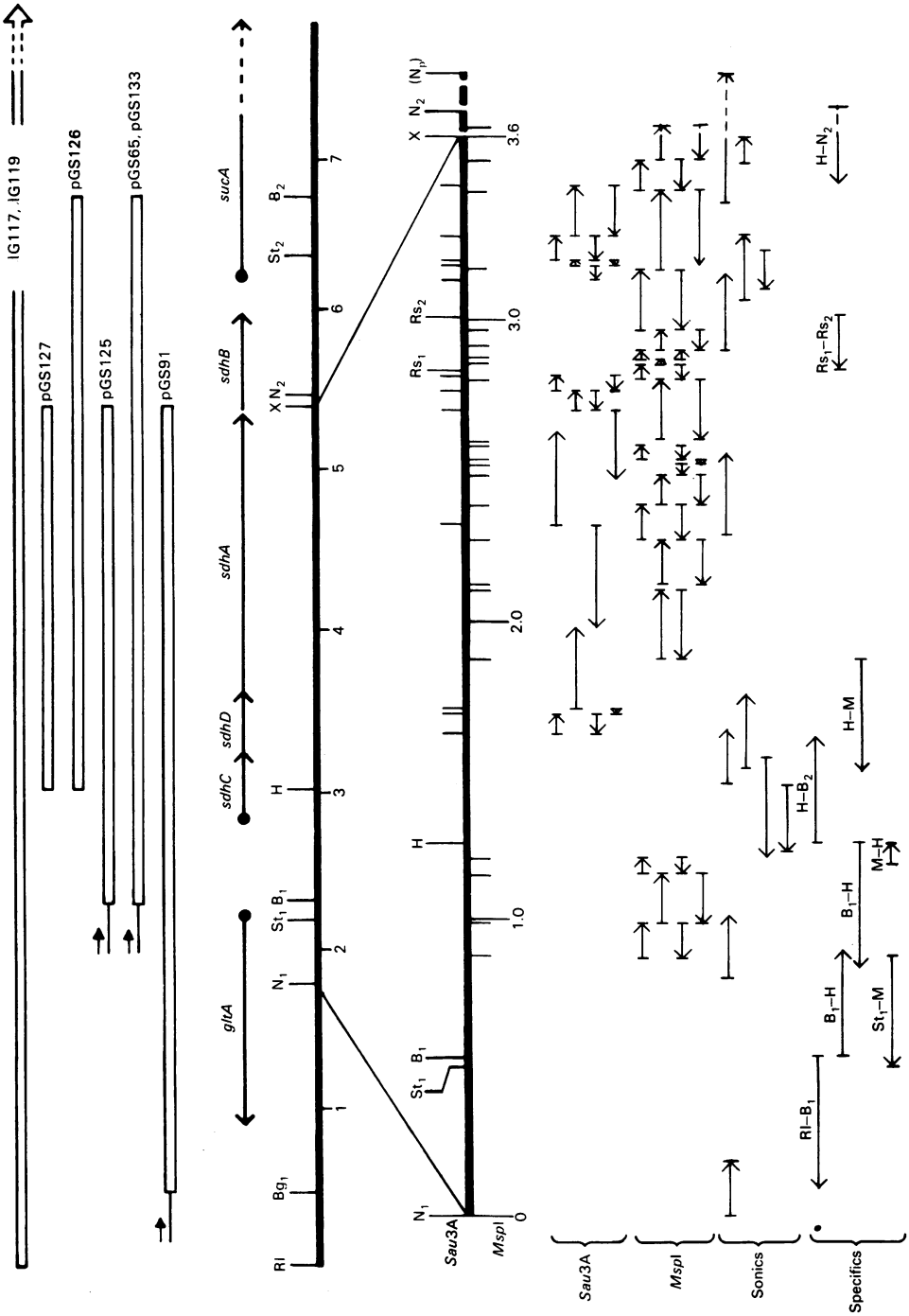


Fig. 1. *Organization of the succinate dehydrogenase genes of E. coli and summary of the DNA sequence data obtained from M13 clones*

A physical map of the region containing the genes encoding citrate synthase (*gltA*), succinate dehydrogenase (*sdhCDAB*), and 2-oxoglutarate dehydrogenase (*sucA*) is shown (scale kilobase-pairs, kb). The approximate positions and transcriptional polarities of the genes are indicated, and their sizes (drawn to scale) are based on the sizes of the corresponding products. The left to right orientation corresponds to clockwise in the *E. coli* linkage map. Segments of DNA cloned in λ vectors (λG117, λG119) and in the *Bam*HI and *Sal*I sites of the plasmid vectors pBR322 and pLG339 are indicated by the open bars: where present the promoter for the plasmid *tet* gene is denoted by a small arrow. The restriction targets for *Bam*HI (B), *Bgl*II (Bg), *Hin*dIII (H), *Msp*I (M), *Nru*I (N), *Eco*RI (RI), *Rsa*I (Rs), *Sst* (St) and *Xho*I (X) have been defined by restriction mapping (Spencer & Guest, 1982) and DNA sequence analysis (Hull *et al.*, 1983, and the present work). N_p refers to the vector *Nru*I site in pGS91. The *Msp*I and *Sau*3A targets are those identified in 'shot-gun' cloning. The lines below the expanded restriction map show the positions, directions and extents of sequences obtained from the two DNA strands.

M_r 64268 (587 amino acid residues excluding the initiating methionine) that corresponds closely to the size predicted for the flavoprotein subunit of succinate dehydrogenase, SDHA. The coding region is terminated by a UAA codon at co-ordinate 3585, and this is just 18 base pairs from the *sdhB* structural gene, which is initiated, without changing the reading frame, by a methionine codon at co-ordinate 3603 (Darlison & Guest, 1984). The translational initiation site proposed for the *sdhA* gene is the AUG codon at co-ordinate 1821 (Fig. 2), and compared with all the other potential initiators in the vicinity, it gives by far the highest score when the PERCEPTRON algorithm is applied (Stormo *et al.*, 1982). The initiation codon is overlapped by the termination codon (UGA, co-ordinate 1819) of an upstream open reading frame which in turn contains a purine-rich region that could serve as a ribosome-binding site for *sdhA* translation (Fig. 2).

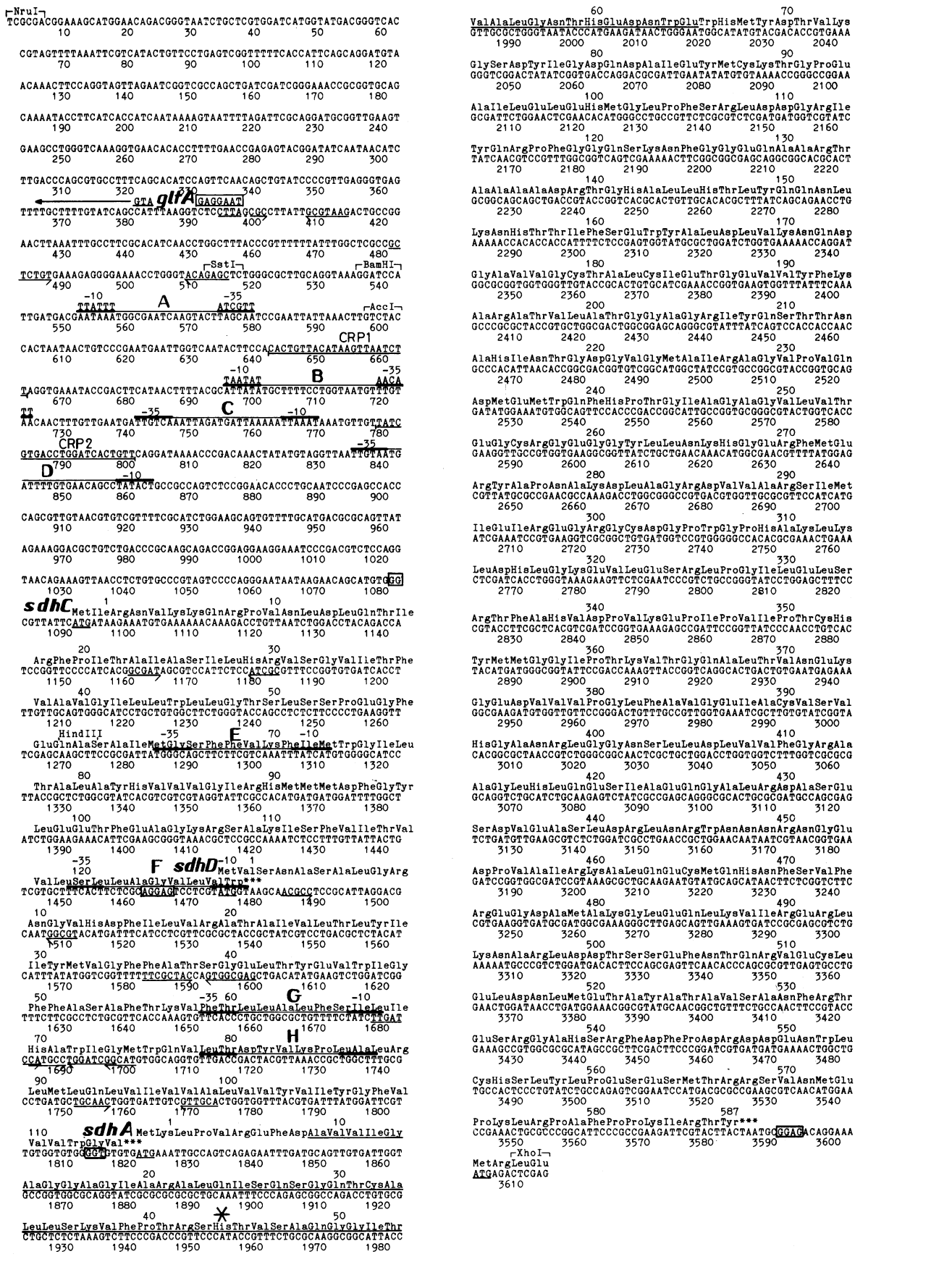
The other identifiable feature in the nucleotide sequence is part of the *gltA* coding region, co-ordinates 383–1 (Fig. 2). This encodes 126 amino acid residues at the *N*-terminal end of citrate synthase and the sequence agrees with that reported recently by Ner *et al.* (1983). Two potential *gltA* promoters (**A** and **B**) are indicated in Fig. 2.

The presence of two overlapping coding regions (*sdhC* and *sdhD*) situated between the *gltA* and *sdhA* structural genes and having the same polarity as the *sdhA* gene suggests that they may form part of a longer *sdh* transcriptional unit containing a total of four genes. Thus, in order to gain further insight into the organization and expression of the genes in this segment, the sequence has been carefully analysed for the presence of transcriptional promoters (−35 and −10 sequences), terminators, and CRP-binding sites. There are several potential promoters (Hawley & McClure, 1983) but no obvious terminators (Rosenberg & Court, 1979). The promoter sequences that are most like the canonical sequence are located at co-ordinates 741–769 (**C**), 835–863 (**D**), 1284–1311 (**E**), 1450–1478 (**F**), 1651–1679 (**G**) and 1711–1738 (**H**), as shown in Fig. 2. All but the first two (**C** and **D**) occur in what appear to be coding regions, so the simplest explanation would be to propose that **C** or **D** is the promoter for a putative *sdh* operon containing four genes: *sdhCDAB*, and if active, the other sequences could represent secondary or internal promoters for expressing groups of distal genes, *sdhDAB* (promoter **E**) or *sdhAB* (promoter **F**, **G** or **H**). A consequence of this organization is that the citrate synthase and succinate dehydrogenase genes would exhibit divergent expression from a single region that contains two potential promoters, each separated by a putative CRP-binding site (see below).

The first of the extra genes (*sdhC*), located between AUG and UAA codons (co-ordinates 1091 and 1478), encodes a protein of M_r 14167 (128 residues excluding the initiating methionine) that has been tentatively identified as the larger of the small labelled polypeptides, SDHC. The gene lacks an obvious ribosome-binding site but analysis with the PERCEPTRON algorithm (Stormo *et al.*, 1982) indicates that of the four translational initiation sites proposed for the *sdh* genes, only that for *sdhC* scores well. Examination of the four *frd* genes likewise indicates that only the *frdA* gene scores well. This suggests that in operons containing very close or overlapping genes, only the first genes have translational initiation sites that score well in this analysis. The *sdhC* coding region overlaps the putative *sdhD* gene by four base pairs. The *sdhD* gene is located between AUG and UGA codons (co-ordinates 1474 and 1819), it is preceded by a good ribosome-binding site (co-ordinates 1462–1466) and it encodes a polypeptide of M_r 12792 (114 residues excluding the initiating methionine) that presumably corresponds to the smallest labelled polypeptide, SDHD.

Genes that are subject to catabolite repression are activated by the complex formed between cyclic AMP and its receptor protein (CRP), which interacts with a conserved sequence of nucleotides close to the corresponding RNA polymerase binding site (Valentin-Hansen, 1982; Chapon & Kolb, 1983). Screening the *gltA–sdhA* region for

Fig. 2. *Nucleotide sequence of the sdhC, D and A genes and primary structures of the corresponding products*
The nucleotide sequence of 3614 base pairs of the non-coding (sense strand) of the succinate dehydrogenase flavoprotein gene (*sdhA*) and two genes (*sdhC* and *sdhD*) encoding small subunits, is presented in the 5′–3′ direction. The intergenic region between the *gltA* (citrate synthase) and *sdh* genes, the coding strand for part of the *gltA* structural gene, and the beginning of the succinate dehydrogenase iron–sulphur protein gene (*sdhB*), are also included. The nucleotide co-ordinates are assigned relative to the first base of the *Nru*I site (N_1) in the *gltA* gene and key restriction sites are indicated. The primary structures of the *sdhC,D* and *A* gene products are shown above the nucleotide sequence, the FAD-binding region of the flavoprotein is underlined and the histidine residue at the FAD-attachment site is denoted by an asterisk. Potential ribosome-binding sites for *gltA* and *sdh* genes are boxed and putative −35 and −10 (Pribnow) promoter sequences are indicated by lines and letters (**A–H**) above the nucleotide sequence. Potential CRP-binding sites are denoted CRP and underlined. Translation initiation sites are also underlined and relevant stop sites are denoted by asterisks thus: ***. Regions of hyphenated dyad symmetry in the *gltA–sdhA* intergenic region are underlined by converging arrows.


NruI
glfA
GAGGAAT
SstI
BamHI
AccI
CRP1
CRP2
sdhC
sdhD
sdhA
XhoI

potential CRP-binding sites that resemble the 22 base-pair consensus sequence 5'-dAANTGT-GANNTANNTCAC/AATTT (Chapon & Kolb, 1983) revealed two sites of good quality: co-ordinates 642–663 for CRP1 and 779–800 for CRP2 (Fig. 2). These are situated between each of two pairs of putative *gltA* and *sdh* promoters. Their presence is consistent with the fact that citrate synthase and succinate dehydrogenase are subject to catabolite repression (Gray *et al.*, 1966; Ruiz-Herrera & Garcia, 1972; Takahashi, 1975). However, at this stage it is not possible to deduce from the organization of these sites whether the tricarboxylic acid cycle genes are expressed (*a*) in a simple and independent manner, involving, e.g., promoter **A** (*gltA*) regulated by CRP1 and promoter **D** (*sdh*) regulated by CRP2, or (*b*) in a more complex and interdependent fashion from an overlapping regulatory region. Alternative interpretations of the gene organization in the *gltA–sdhA* intergenic region have not been ruled out. For example, the *sdhC* and *sdhD* genes could be expressed independently of the other *sdh* genes, or of each other, in transcriptional units of the type: *sdhC* and *sdhD* with *sdhAB*; *sdhCD* and *sdhAB*; *sdhC* with *sdhDAB*, If so, there would be a need for active promoters in the *sdhC* or *sdhD* coding regions (possibly promoters **E–H**): the existence of such overlapping coding and regulatory sequences in *E. coli* has already been described (Grundström & Jaurin, 1982).

Other features of the nucleotide sequence

The nucleotide sequence contains a number of regions of hyphenated dyad symmetry, and those between the *gltA* and *sdhA* structural genes that could form stable stem-and-loop structures in the corresponding RNA transcripts are shown in Fig. 2. Most of them occur in presumptive coding regions and none are particularly striking, their free energy values being in the range -25.1 to $-46.9\,kJ\cdot mol^{-1}$ (-6.0 to $-11.2\,kcal\cdot mol^{-1}$) (calculated according to Tinoco *et al.*, 1973). It may be significant that the strongest (co-ordinates 1484–1511) is located just downstream of the *sdhC* coding region in the *sdhD* gene.

The sequence has been examined for the presence of the tandem *lysT*, *valT* and *supG* genes and the *cbt* gene that may be located between the *gltA* and *sdhA* genes, but no tRNA genes could be detected nor could any further coding regions be found. However, the *cbt* gene lacks identifiable features, and the possibility that *cbt* corresponds to *sdhC* or *sdhD* has not been ruled out. This would mean that *cbt* has a dual function in succinate dehydrogenase and carboxylate transport.

The sequence overlaps that of the *gltA* region published by Ner *et al.* (1983) by 1270 base pairs and it contains four differences. These involve insertions of T and G at co-ordinates 590 and 870, and deletions of T and G at co-ordinates 637 and 1260, respectively, but none affect the interpretation of the sequence except the last, which would change the reading frame of *sdhC* to one with a very atypical codon usage terminating 43 codons earlier at the UAG codon, co-ordinate 1350 in Fig. 2.

Codon usage. The codon usages for the *sdhA*, *C* and *D* genes are typically non-random (Table 1) and very similar to those of the *frdA*, *C* and *D* genes. By examining the choice of degenerate and modulating codons it is possible to assess the degree of expression of the genes, because highly expressed genes use few, if any, of the modulatory codons corresponding to minor tRNA species and favour codons with intermediate codon–anticodon interaction energies (Grosjean & Fiers, 1982). The proportions of modulatory codons in the *sdhA*, *C* and *D* genes (1.5, 5.9 and 1.8%) and optimal energy codons (57, 50 and 69%) correspond to those in the *frdA*, *C* and *D* genes and suggest that the genes are not highly expressed. When compared with other tricarboxylic acid cycle genes, the *sdh* genes resemble the 2-oxoglutarate dehydrogenase gene (*sucA*) in appearing less strongly expressed than the related *sucB*, *aceE*, *aceF* and *lpd* genes (Spencer *et al.*, 1984).

Primary structure and composition of the flavoprotein subunit, SDHA

The primary structure of the flavoprotein subunit of the *E. coli* succinate dehydrogenase (SDHA) is presented in Fig. 2. It contains 587 residues, excluding the initiating formylmethionine, and corresponds to a protein of M_r 64268 (64988 including the flavin cofactor). The amino acid composition and polarity derived from the nucleotide sequence are shown in Table 2 and they clearly resemble those of the flavoprotein subunits of the *Rhodospirillum rubrum* and bovine heart mitochondrial enzymes (Davis *et al.*, 1977) and the *E. coli* fumarate reductase (Cole, 1982).

Comparisons between the FAD and NAD(P) domains of several enzymes has revealed the existence of a common secondary structural arrangement, the FAD-binding fold (Thieme *et al.*, 1981; Wierenga *et al.*, 1983). Furthermore, there is a high degree of sequence conservation in the β–α–β segment that contains the AMP-binding site of this fold (Arscott *et al.*, 1982; Stephens *et al.*, 1983; Rice *et al.*, 1984). The flavin AMP-binding sites are generally situated close to the *N*-terminus, and such a site is found between residues 8 and 36 in the SDHA protein (Fig. 2). The AMP-binding sites of several *E. coli* flavoproteins are compared with that of human erythrocyte glutathione reductase in

Table 1. *Codon usage in the sdhA, C and D genes*

The AUG initiation codons are not included with the methionine codons. The codon pairs enclosed in solid boxes are those whose use varies particularly between strongly and weakly expressed genes, and the codons marked by arrows are potential modulatory codons (Grosjean & Fiers, 1982).

Gene...	sdh	A	C	D		sdh	A	C	D		sdh	A	C	D		sdh	A	C	D	
	UUU	Phe	5	4	2	UCU	Ser	6	1	2	UAU	Tyr	11	2	3	UGU	Cys	7	0	0	
	UUC	Phe	14	5	7	UCC	Ser	10	6	1	UAC	Tyr	3	0	3	UGC	Cys	4	0	0	
	UUA	Leu	0	0	1	UCA	Ser	0	1	0	UAA	End	1	1	0	UGA	End	0	0	1	
	UUG	Leu	3	0	3	UCG	Ser	7	0	0	UAG	End	0	0	0	UGG	Trp	7	3	4	
	CUU	Leu	3	4	0	CCU	Pro	0	2	0	CAU	His	7	1	2	CGU	Arg	25	0	0	
	CUC	Leu	5	4	2	CCC	Pro	0	1	0	CAC	His	12	2	0	CGC	Arg	13	3	3	
→	CUA	Leu	0	1	0	CCA	Pro	4	0	0	CAA	Gln	5	2	1	CGA	Arg	1	0	0	←
	CUG	Leu	40	6	10	CCG	Pro	20	0	1	CAG	Gln	16	1	1	CGG	Arg	0	1	0	←
	AUU	Ile	12	4	3	ACU	Thr	5	1	0	AAU	Asn	5	2	1	AGU	Ser	0	0	1	
	AUC	Ile	14	7	7	ACC	Thr	24	4	5	AAC	Asn	20	0	1	AGC	Ser	6	2	1	
→	AUA	Ile	0	2	0	ACA	Thr	0	1	1	AAA	Lys	21	5	2	AGA	Arg	1	2	0	←
	AUG	Met	15	5	3	ACG	Thr	4	1	1	AAG	Lys	1	0	0	AGG	Arg	0	0	0	←
	GUU	Val	14	4	5	GCU	Ala	7	2	4	GAU	Asp	21	1	1	GGU	Gly	26	5	3	
	GUC	Val	8	5	4	GCC	Ala	11	1	3	GAC	Asp	11	1	1	GGC	Gly	26	4	3	
	GUA	Val	4	2	2	GCA	Ala	13	2	2	GAA	Glu	29	4	1	GGA	Gly	2	1	2	←
	GUG	Val	15	4	8	GCG	Ala	25	5	2	GAG	Glu	14	1	1	GGG	Gly	5	0	0	←

Table 2. *Amino acid compositions of the flavoprotein subunits of three succinate dehydrogenases and a fumarate reductase*

The DNA-derived compositions of the *E. coli* enzymes from the present work and Cole (1982) are compared with the analytical values for the *Rhodospirillum rubrum* and beef heart mitochondrial enzymes (Davis *et al.*, 1977). The initiating formylmethionine residues are not included.

	Residues in the flavoprotein subunits of:			
	Succinate dehydrogenase from:			Fumarate reductase from
Amino Acid	*E. coli*	*R. rubrum*	Beef heart	*E. coli*
Asp	32	52	57	35
Asn	25			23
Thr	33	40	38	36
Ser	29	21	37	20
Glu	43	53	67	44
Gln	21			22
Pro	24	25	30	22
Gly	59	53	65	66
Ala	56	66	59	61
Val	41	38	45	41
Met	15	13	14	20
Ile	26	25	31	23
Leu	51	44	55	48
Tyr	14	16	21	17
Phe	19	19	22	20
Lys	22	20	26	27
His	19	18	17	21
Arg	40	34	41	38
Cys	11	9	14	10
Trp	7	6	6	7
Total	587	552	645	601
M_r	64268	59864	70054	65835
Polarity (%)	45	43	44	44

Table 3. They all contain a characteristic stretch of four hydrophobic residues (β_A) preceding an invariant glycine residue, a conserved stretch of residues with small side chains (α_A), and a second relatively hydrophobic segment (β_B). The glutamate residue that recognizes the adenine ribose hydroxy groups in the FAD- and NAD-binding sites of glutathione reductase and lipoamide

Table 3. *Amino acid sequences around the FAD-binding sites of several flavoproteins*

The alignments for several *E. coli* flavoproteins are based on the structure of human erythrocyte glutathione reductase (GR; Thieme *et al.*, 1981; Rice *et al.*, 1984) to highlight the apparent conservation of the β–α–β structure (broken boxes) in the AMP-binding segments (upper portion), and the structural diversity of the adjacent segments (lower portion). The strong homologies around the flavin-attachment sites of fumarate reductase (FRDA; Cole, 1982), succinate dehydrogenase (SDHA) and the 23-residue flavopeptide of the bovine heart mitochondrial enzyme (Kenney *et al.*, 1972) contrast with the sequences of NADH dehydrogenase (NADHDH; Young *et al.*, 1981), which does not have a covalently-bound FAD coenzyme, and of the disulphide oxidoreductases (GR and lipoamide dehydrogenase; LPDH, Stephens *et al.*, 1983), which interact with the isoalloxazine ring via their cystine peptides. Identical residues in the regions of FRDA and SDHA corresponding to the flavopeptide are enclosed in solid boxes and the critical histidine residues are denoted by an asterisk, as are the active-site cysteine residues of GR and LPDH.

```
                    βA                  αA                                  βB
GR(human)  (22) D Y L V I G G S G G L A S A R R A A E L - - - - - G A R A A V V E
LPDH        (7) Q V V V L G A G P A G Y S A A F R C A D L - - - - - G L E T V I V E
NADHDH      (6) K I V I V G G A G G L E M A T Q L G H K L G R K K K A K I T L V D
FRDA        (6) D L A I V G A G G A G L R A A I A A A Q A N P - - - N A K I A L I S
SDHA        (8) D A V V I G A G G A G I A R L A Q I S Q S - - - - - G Q T C A L L S

                        *                   *
S H - K L G G T C V N V G C V P K K V M W N T A V H S E F M H D H A (83)  GR(human)
R Y N T L G G V C L N V G C I P S K A L L H V A K V I E E A K A L A (69)  LPDH
R N H S H L W K P L L H E V A T G S L D E G V D A L S Y L A H A R N (74)  NADHDH
K V Y P M R S H T V A A E G G - S A A V A Q D H - D S F E Y H F H D (68)  FRDA
K V F P T R S H T V S A Q G G I T V A L G N T H E D N W E W H M Y D (70)  SDHA
          S H T V A A Z G G I B L A A G B M D E B Z W R                  SDH(bovine)
                    *
```

dehydrogenase is replaced by aspartate in NADH dehydrogenase, but the acid residue is not retained in succinate dehydrogenase or fumarate reductase.

The FAD cofactor of the bovine heart mitochondrial succinate dehydrogenase is covalently bound via the 8α-methyl group of the isoalloxazine ring and the *tele*-N atom of a histidyl residue in the protein (Salach *et al.*, 1972; Singer & Edmondson, 1974). Furthermore, the amino acid sequence of a 23-residue flavopeptide from this enzyme has been deduced (Kenney *et al.*, 1972) and comparison with the *E. coli* flavoprotein reveals a segment that is identical at 15 of the 23 positions and this identifies the histidine residue at position 44 as the flavin-attachment site (Table 3). The binding site for the covalently-bound FAD cofactor of the *E. coli* fumarate reductase has also been identified by its homology with the bovine heart flavopeptide (Cole, 1982). In fact the sequence containing the critical histidine residue, Ser-**His**-Thr-Val-Xaa-Ala-Xaa-Gly-Gly, is common to all three enzymes (Table 3). It is interesting to note that, relative to the AMP-binding site, the flavin-attachment site is located at approximately the same position in the primary structure as the active-site disulphide bridge that interacts with the isoalloxazine ring in glutathione reductase and lipoamide dehydrogenase (Table 3). In the case of NADH dehydrogenase, which contains neither covalently-bound FAD nor an active-site disulphide bridge, there is no obvious homology with the other flavoenzymes in the region immediately beyond the AMP-binding site (Table 3). These examples clearly illustrate three different arrangements of the polypeptide chain around the FAD cofactor.

The primary structures of the four *E. coli* flavoenzymes (succinate dehydrogenase, fumarate reductase, lipoamide dehydrogenase and NADH dehydrogenase) have been compared in pairs by using the proportional matching option of the graphics program DIAGON (Staden, 1982). This incorporates a scoring system based on MDM_{78} (a mutation data matrix that is very good for detecting distant relationships between amino acid sequences; Schwartz & Dayhoff, 1978), and diagonal lines in the matrices correspond to highly significant regions of homology. This analysis showed that the homologies are mainly centred on the AMP-binding sites of the FAD and NAD domains for all of the pairs except for SDHA and FRDA, which exhibit a remarkably high degree of mutual homology extending throughout their entire lengths (Fig. 3). An alignment based on the DIAGON analysis and refined by visual inspection is shown in Fig. 4. Some 246 of the 563 equivalenced residues (44%) are identical and a further 112 residues (20%) are conservatively changed at the 0.10 scoring limit of MDM_{78}. A similar high degree of homology is seen at the DNA level. The departure from colinearity between SDHA and FRDA is due mainly to three insertion-deletions at positions 122 (+7 residues), 277 (−11 residues) and 361 (+10 residues) in SDHA, and the larger size of FRDA is due to the

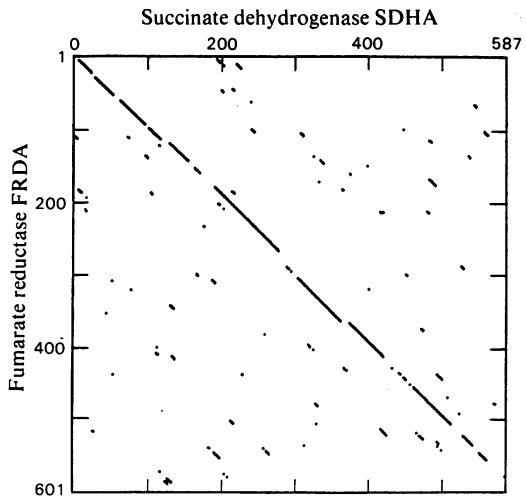


Fig. 3. *Amino acid sequence comparison of the flavoprotein subunits of succinate dehydrogenase and fumarate reductase*

A comparison matrix is shown for the flavoprotein subunits of the *E. coli* succinate dehydrogenase (SDHA) and fumarate reductase (FRDA; Cole, 1982). The computer program DIAGON (Staden, 1982) was used. The points correspond to the midpoints of each span of 21 residues giving a score equivalent to a double matching probability of <0.001. Numbers on the axes are residue numbers.

extra 21-residue segment at the *C*-terminus (Fig. 4). This remarkable sequence homology indicates that the two enzymes are closely related in evolution, rather than of entirely independent origin. It would appear that they have emerged by the duplication of a common ancestral gene and that their specific roles in aerobic and anaerobic metabolism have been imposed and refined by different transcriptional regulation and diverging evolution of the individual gene products.

The distribution of cysteine residues does not indicate which residues may be contributing to the iron–sulphur centres of the flavoprotein subunits. However, it is particularly striking that of the 21 cysteine residues in the two enzymes (11 in SDHA and 10 in FRDA), only one pair occupy identical positions in the alignment (Fig. 4). These residues are located in a very highly conserved region (positions 198–270 in SDHA and residues 189–261 in FRDA) and it would seem very likely that they are functionally important. Studies with bovine heart mitochondrial succinate dehydrogenase have shown that the flavoprotein contains a reactive thiol group at the active site (Kenney *et al.*, 1976). The flavoprotein subunit of the *Vibrio succinogenes* fumarate reductase has likewise been shown to contain an essential thiol group at the substrate site (Unden & Kröger, 1980). It is therefore likely that the corresponding active-site thiol residues in the *E. coli* enzymes are contributed by Cys-256 in SDHA and Cys-247 in FRDA. Histidine residues are more abundant in both flavoproteins (Table 2) and in the chosen alignment there are no fewer than eight conserved positions, in addition to the histidine residue at the FAD-attachment site (Fig. 4). It has been suggested that a histidine residue may have a proton donor/acceptor function in succinate dehydrogenase (Vik & Hatefi, 1981). This parallels the situation in glutathione reductase and lipoamide dehydrogenase and it may be significant that one of the conserved histidine residues is located close to the putative active-site cysteine residue, in a sequence His-Pro-Thr (residues 241–243 in SDHA) that is identical to the active-site histidine sequences of the disulphide oxidoreductases (Rice *et al.*, 1984).

The hydrophobic subunits

The two open reading frames that are found upstream of the *sdhA* gene encode two very hydrophobic products. These have been tentatively designated as the *sdhC* and *sdhD* genes because of their similarity to the *frdC* and *frdD* genes. The latter genes encode two hydrophobic subunits, FRDC (M_r 14883) and FRDD (M_r 12975) that are required to anchor the flavoprotein and iron–sulphur protein subunits of fumarate reductase to the cytoplasmic membrane (Lemire *et al.*, 1982, 1983) in a manner similar to that proposed for the F_0–F_1 components of ATP synthase (Fillingame, 1980). Bovine heart succinate dehydrogenase is also associated with two hydrophobic polypeptides that may likewise serve as anchor proteins (Girdlestone *et al.*, 1981; Merli *et al.*, 1979). The possibility that these subunits have additional functions in proton translocation or other aspects of electron transport, like the F_0 portion of ATP synthase (Ovchinnikov *et al.*, 1982), has not been investigated. The products of the *sdhC* and *sdhD* genes closely resemble their *frd* counterparts in size, polarity and the abundance of hydrophobic residues (Table 4). Furthermore, each protein contains three very hydrophobic and colinear segments of approx. 25 residues as shown by HYDROPLOT analysis (Fig. 5). These segments could represent transmembrane α-helices analogous to those suggested for bacteriorhodopsin, the F_0 portion of ATP synthase, and the *lac* carrier protein (Engelman *et al.*, 1980; Foster *et al.*, 1983). However, the similarity between the two pairs of hydrophobic proteins does not extend to their amino acid sequences, because no significant homologies were detected by DIAGON analysis. By the same analysis, both pairs of hydrophobic proteins shared short stretches of homology with

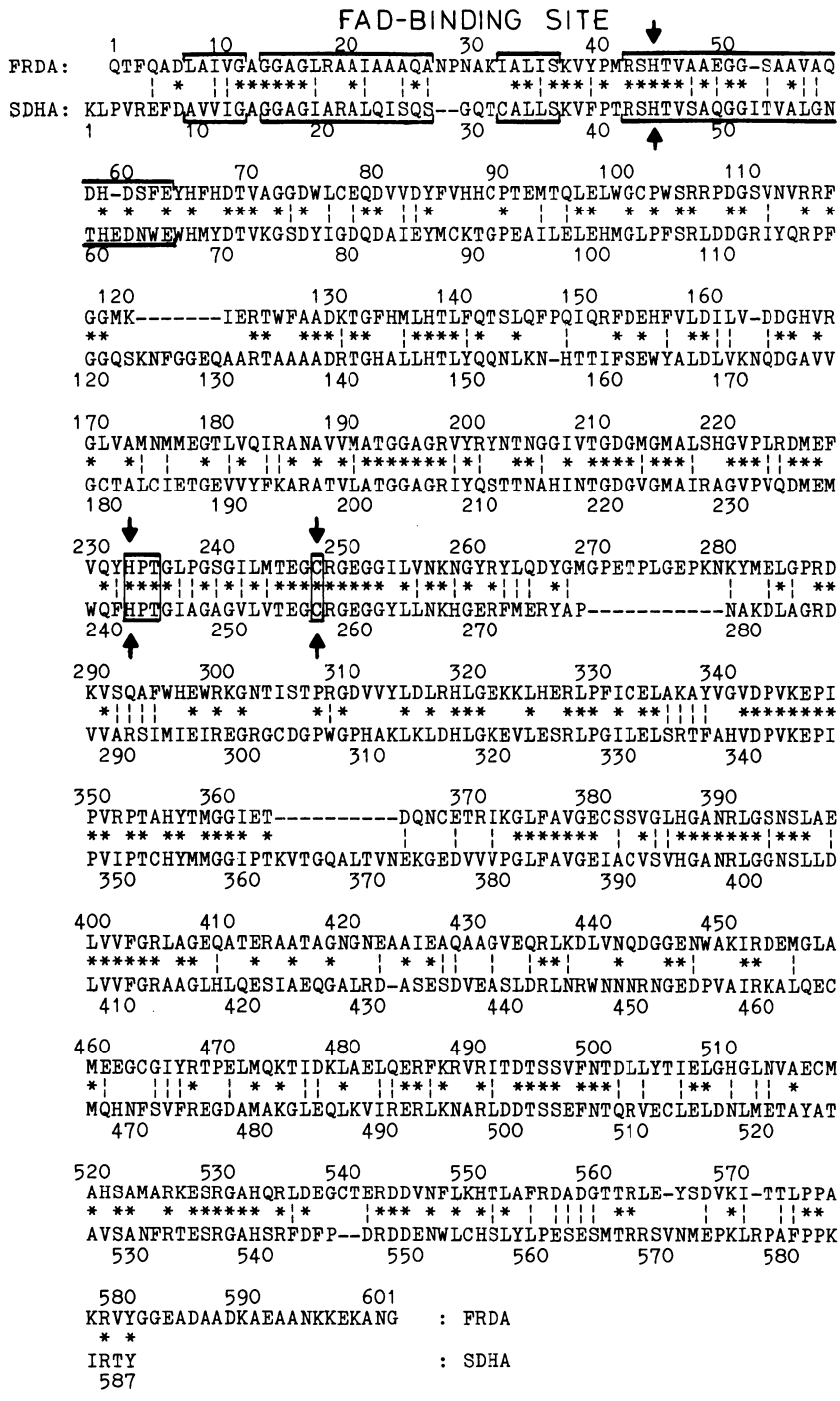


Fig. 4. *Alignment of the amino acid sequences of the flavoprotein subunits of succinate dehydrogenase and fumarate reductase*
The sequences, shown in single-letter amino acid code, have been aligned for maximum homology based on the DIAGON comparison (Fig. 3). Identical residues are identified by asterisks and conservatively-substituted residues with scores of $\geqslant 0.1$ in the MDM_{78} matrix (Schwartz & Dayhoff, 1978) are identified by vertical bars. The FAD-binding regions are indicated and the histidine residues at the flavin-attachment sites are denoted by arrows. The conserved cysteine residues and His-Pro-Thr sequences that may be in the active sites are boxed and arrowed.

Table 4. *Amino acid compositions of the hydrophobic subunits of succinate dehydrogenase and fumarate reductase of E. coli*
The DNA-derived compositions of the *E. coli* proteins from the present work and from Grundström & Jaurin (1982) do not include the initiating formylmethionine residues.

	Residues in the hydrophobic subunits			
Amino acid	SDHC	FRDC	SDHD	FRDD
Asp	2	2	2	2
Asn	2	4	2	2
Thr	7	10	7	3
Ser	10	3	5	4
Glu	5	6	2	2
Gln	3	1	2	1
Pro	3	9	1	8
Gly	10	5	7	12
Ala	10	14	11	10
Val	15	12	19	12
Met	5	3	3	5
Ile	13	12	10	11
Leu	15	14	17	18
Tyr	2	4	6	2
Phe	9	9	9	8
Lys	5	10	2	3
His	3	1	2	5
Arg	6	4	3	4
Cys	0	0	0	1
Trp	3	7	4	4
Total	128	130	114	117
M_r	14167	14883	12792	12975
Polarity (%)	34	32	24	23

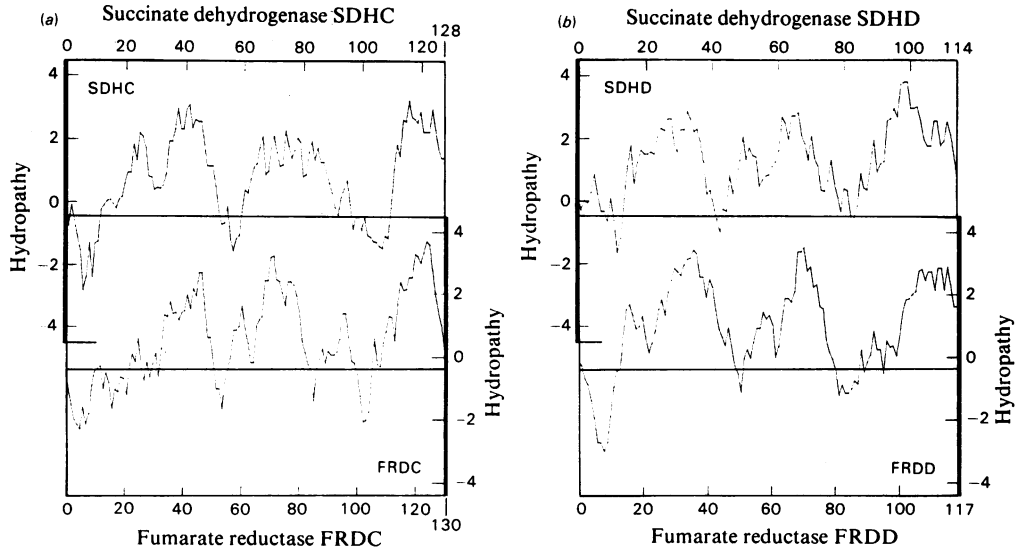


Fig. 5. *Comparisons of the hydropathy profiles of the hydrophobic subunits of succinate dehydrogenase and fumarate reductase*
Hydropathy profiles of the hydrophobic subunits of succinate dehydrogenase (upper panels) and fumarate reductase (lower panels): (*a*) SDHC and FRDC, (*b*) SDHD and FRDD. Consecutive hydropathy averages are plotted at the mid-point of a seven-residue segment as it advances from *N*- to *C*-terminus. Relative hydrophobicity and hydrophilicity (vertical axes) are recorded in the range +4.5 to −4.5 and a horizontal line representing the average for most sequenced proteins is included (Kyte & Doolittle, 1982). Numbers on the horizontal axes represent residue numbers.

the F_0 a and c subunits of the *E. coli* ATP synthase (Gay & Walker, 1981; Nielsen *et al.*, 1981); no homology with cytochrome b_{562} (Bethge *et al.*, 1981) could be detected. The lack of sequence homology between the hydrophobic subunits of succinate dehydrogenase and fumarate reductase contrasts sharply with the high degree of conservation observed for the flavoprotein subunits and for the iron–sulphur protein subunits (M. G. Darlison & J. R. Guest, unpublished work). Nevertheless, it would appear that the two pairs of hydrophobic gene products do represent analogous subunits of the enzyme complexes, but further biochemical and genetic evidence will be needed to confirm this preliminary characterization.

Organization of the succinate dehydrogenase genes

The nucleotide sequence of the *gltA–sucB* region of the *E. coli* chromosome confirms and extends earlier work in which the approximate positions and the transcriptional polarities of the *gltA*, *sdhA* and *B*, and *sucA* and *B* genes were defined (Spencer & Guest, 1982; Spencer *et al.*, 1984). The sequence indicates that the *sdh* operon contains no less than four genes situated between the *gltA* and *sucA* genes: $\overleftarrow{gltA}$-$\overrightarrow{sdhCDAB}$-$\overrightarrow{sucAB}$, in what is the largest cluster of tricarboxylic acid cycle genes. The transcription of the citrate synthase and succinate dehydrogenase genes appears to diverge from a single intergenic region containing several potential promoter and regulatory sites. This may be significant with respect to the overall control of expression of the tricarboxylic acid cycle. The very close proximity of the *gltA* and *sdh* genes is consistent with very high genetic linkage between these genes (97% P1 cotransduction) compared with that between that for the *sdh* and *suc* genes (90% P1 cotransduction). It may also be significant that at the distal end of the *sdh* operon, between the *sdhB* and *sucA* genes, there is a set of three palindromic sequences similar to the intercistronic regulatory elements found in several operons (Darlison *et al.*, 1984; Spencer *et al.*, 1984). This could mean that the transcriptional organization of the *sdh* and *suc* genes could be more complex than for two independent operons.

The overall composition of the *sdh* operon closely resembles that encoding the analogous *frd* operon, not only in possessing four comparable genes but also in the economy of sequence witnessed by the overlapping genes. The major difference concerns gene order, *sdhCDAB* compared to *frdABCD*, but it is not known whether this has any functional significance. The adjacent flavoprotein and iron–sulphur protein genes are in the same order (*AB*) and their similarity is indicative of close ancestral relationship and possible evolution by gene duplication. The genes encoding the hydrophobic proteins are also in the same order relative to each other (*CD*) but occupy different positions in the two operons. This could mean that the operons have evolved by fusing two pairs of ancestral genes in different orders, or that the genes of one or both operons have been reorganized since the duplication of an ancestral four-gene operon. The *C* and *D* genes appear to have diverged further than the *A* and *B* genes and this could reflect a greater selection pressure for maintaining catalytic rather than membrane-insertion functions. There is evidence that the specific functions of the *sdh* and *frd* genes in aerobic and anaerobic metabolism are not entirely imposed by regulatory compartmentation. This is because the functional interchangeability of the two operons, demonstrated by the partial replacement of *sdh* function by multiple copies of the *frd* operon (Guest, 1981*a*), appears not to be a reciprocal relationship. It has been observed that in multicopy situations, plasmid- and phage-encoded *sdh* operons do not significantly complement the nutritional lesions of *frd* mutants (R. J. Wilde & J. R. Guest, unpublished work). This suggests either that the succinate dehydrogenase system has a narrower functional specificity, or that the *sdh* operon is more stringently regulated than *frd*.

Future work should be aimed at (*a*) identifying the mRNA transcripts for the *sdh* and *gltA* genes, (*b*) defining the molecular mechanisms controlling *gltA* and *sdh* expression, and (*c*) amplifying and isolating the *sdh* gene products for further investigation of their gene–protein and structure–function relationships.

We thank Elizabeth P. Hull and Ruth Roberts for isolating some of the M13 clones, Roger Staden and Ian K. Duckenfield for generously providing and implementing computer programs, and the Science and Engineering Research Council for support by project grant (GR/B 91396) to J. R. G.

References

Arscott, L. D., Williams, C. H., Jr. & Schulz, G. E. (1982) in *Flavins and Flavoproteins* (Massey, V. & Williams, C. H., Jr., eds.), vol. 21, pp. 44–48, Elsevier/North-Holland Biomedical Press, Amsterdam

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

Bethge, P. H., Bellamy, H. D. & Mathews, F. S. (1981) *J. Mol. Biol.* **148**, 427–448

Chapon, C. & Kolb, A. (1983) *J. Bacteriol.* **156**, 1135–1143

Cole, S. T. (1982) *Eur. J. Biochem.* **122**, 479–484

Cole, S. T. (1984) *Biochem. Soc. Trans.* **12**, 237–238

Cole, S. T. & Guest, J. R. (1978) *Soc. Gen. Microbiol. Q.* **6**, 42–43

Cole, S. T. & Guest, J. R. (1979*a*) *Eur. J. Biochem.* **102**, 65–71

Cole, S. T. & Guest, J. R. (1979*b*) *FEMS Microbiol. Lett.* **5**, 65–67
Cole, S. T. & Guest, J. R. (1980*a*) *Mol. Gen. Genet.* **178**, 404–418
Cole, S. T. & Guest, J. R. (1980*b*) *Mol. Gen. Genet.* **179**, 377–385
Cole, S. T., Grundström, T., Jaurin, B., Robinson, J. J. & Weiner, J. H. (1982) *Eur. J. Biochem.* **126**, 211–216
Creaghan, I. & Guest, J. R. (1972) *J. Gen. Microbiol.* **71**, 207–220
Darlison, M. G., Spencer, M. E. & Guest, J. R. (1984) *Eur. J. Biochem.* **141**, 351–359
Davis, K. A., Hatefi, Y., Crawford, I. P. & Baltscheffsky, H. (1977) *Arch. Biochem. Biophys.* **180**, 459–464
Dickie, P. & Weiner, J. H. (1979) *Can. J. Biochem.* **57**, 813–821
Engelman, D. M., Henderson, R., McLachlan, A. D. & Wallace, B. A. (1980) *Proc. Natl. Acad. Sci. U.S.A.* **77**, 2023–2027
Fillingame, R. H. (1980) *Annu. Rev. Biochem.* **49**, 1079–1113
Foster, D. L., Boublik, M. & Kaback, H. R. (1983) *J. Biol. Chem.* **258**, 31–34
Gay, N. J. & Walker, J. E. (1981) *Nucleic Acids Res.* **9**, 3919–3926
Girdlestone, J., Blisson, R. & Capaldi, R. A. (1981) *Biochemistry* **20**, 152–156
Gray, C. T., Wimpenny, J. W. T. & Mossman, M. R. (1966) *Biochim. Biophys. Acta* **117**, 33–41
Grosjean, H. & Fiers, W. (1982) *Gene* **18**, 199–209
Grundström, T. & Jaurin, B. (1982) *Proc. Natl. Acad. Sci. U.S.A.* **79**, 1111–1115
Guest, J. R. (1981*a*) *J. Gen. Microbiol.* **122**, 171–179
Guest, J. R. (1981*b*) *J. Gen. Microbiol.* **124**, 17–23
Guest, J. R., Roberts, R. E. & Stephens, P. E. (1983) *J. Gen. Microbiol.* **127**, 65–79
Guest, J. R., Darlison, M. G., Spencer, M. E. & Stephens, P. E. (1984) *Biochem. Soc. Trans.* **12**, 220–223
Hawley, D. K. & McClure, W. R. (1983) *Nucleic Acid. Res.* **11**, 2237–2255
Hederstedt, L. & Rutberg, L. (1981) *Microbiol. Rev.* **45**, 542–555
Hull, E. P., Spencer, M. E., Wood, D. & Guest, J. R. (1983) *FEBS Lett.* **156**, 366–370
Jones, R. W., Kranz, R. G. & Gennis, R. B. (1982) *FEBS Lett.* **142**, 81–87
Kadenbach, B., Jarausch, J., Hartmann, R. & Merle, P. (1983) *Anal. Biochem.* **129**, 517–521
Kenney, W. C., Walker, W. H. & Singer, T. P. (1972) *J. Biol. Chem.* **247**, 4510–4513
Kenney, W. C., Mowery, P. C., Seng, R. L. & Singer, T. P. (1976) *J. Biol. Chem.* **251**, 2369–2373
Kyte, J. & Doolittle, R. F. (1982) *J. Mol. Biol.* **157**, 105–132
Lemire, B. D., Robinson, J. J. & Weiner, J. H. (1982) *J. Bacteriol.* **152**, 1126–1131
Lemire, B. D., Robinson, J. J., Bradley, R. D., Scraba, D. G. & Weiner, J. H. (1983) *J. Bacteriol.* **155**, 391–397
Lohmeier, E., Hagen, D. S., Dickie, P. & Weiner, J. H. (1981) *Can. J. Biochem.* **59**, 158–164
McDonell, M. W., Simon, M. N. & Studier, F. W. (1977) *J. Mol. Biol.* **110**, 119–146
Messing, J. & Vieira, J. (1982) *Gene* **19**, 269–276
Merli, A., Capaldi, R. A., Ackrell, B. A. C. & Kearney, E. B. (1979) *Biochemistry* **18**, 1393–1400
Ner, S. S., Bhayana, V., Bell, A. W., Giles, I. G., Duckworth, H. W. & Bloxham, D. P. (1983) *Biochemistry* **22**, 5243–5249
Nielsen, J., Hansen, F. G., Hoppe, J., Friedl, P. & von Meyenburg, K. (1981) *Mol. Gen. Genet.* **184**, 33–39
Ovchinnikov, Yu. A., Abdulaev, N. G. & Modyanov, N. N. (1982) *Annu. Rev. Biophys. Bioeng.* **11**, 445–463
Owen, P. & Condon, C. (1982) *FEMS Microbiol. Lett.* **14**, 223–227
Pratt, J. M., Boulnois, G. J., Darby, V., Orr, E., Wahle, E. & Holland, I. B. (1981) *Nucleic Acids Res.* **9**, 4459–4474
Rice, D. W., Schulz, G. E. & Guest, J. R. (1984) *J. Mol. Biol.* **174**, 483–496
Rosenberg, M. & Court, D. (1979) *Annu. Rev. Genet.* **13**, 319–353
Ruiz-Herrera, J. & Garcia, L. G. (1972) *J. Gen. Microbiol.* **72**, 29–35
Salach, J., Walker, W. H., Singer, T. P., Ehrenberg, A., Hemmerich, P., Ghista, S. & Hartmann, V. (1972) *Eur. J. Biochem.* **26**, 267–278
Sancar, A., Hack, A. M. & Rupp, W. D. (1979) *J. Bacteriol.* **137**, 692–693
Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463–5467
Sanger, F., Coulson, A. R., Barrell, B. G., Smith, A. J. H. & Roe, B. A. (1980) *J. Mol. Biol.* **143**, 161–178
Schwartz, R. M. & Dayhoff, M. O. (1978) *Atlas of Protein Sequence and Structure*, vol. 5, pp. 353–358, National Biomedical Research Foundation, Washington DC
Shaw, D. J. & Guest, J. R. (1982) *J. Gen. Microbiol.* **128**, 2221–2228
Singer, T. P. & Edmondson, D. E. (1974) *FEBS Lett.* **42**, 1–14
Spencer, M. E. & Guest, J. R. (1973) *J. Bacteriol.* **114**, 563–570
Spencer, M. E. & Guest, J. R. (1974*a*) *J. Bacteriol.* **117**, 947–953
Spencer, M. E. & Guest, J. R. (1974*b*) *J. Bacteriol.* **117**, 954–959
Spencer, M. E. & Guest, J. R. (1982) *J. Bacteriol.* **151**, 542–552
Spencer, M. E., Darlison, M. G., Stephens, P. E., Duckenfield, I. K. & Guest, J. R. (1984) *Eur. J. Biochem.* **141**, 361–374
Staden, R. (1979) *Nucleic Acids Res.* **6**, 2601–2610
Staden, R. (1980) *Nucleic Acids Res.* **8**, 3673–3694
Staden, R. (1982) *Nucleic Acids Res.* **10**, 2951–2961
Staden, R. & McLachlan, A. D. (1982) *Nucleic Acids Res.* **10**, 141–156
Stephens, P. E., Lewis, H. M., Darlison, M. G. & Guest, J. R. (1983) *Eur. J. Biochem.* **135**, 519–527
Stoker, N. G., Fairweather, N. F. & Spratt, B. G. (1982) *Gene* **18**, 335–341
Stormo, G. D., Schneider, T. D., Gold, L. & Andrzej, E. (1982) *Nucleic Acids Res.* **10**, 2997–3011
Takahashi, Y. (1975) *J. Biochem.* (*Tokyo*) **78**, 1097–1100
Tinoco, I., Borer, P. N., Dengler, B., Levine, M. D., Uhlenbeck, O. C., Crothers, D. M. & Gralla, J. (1973) *Nature* (*London*) *New Biol.* **246**, 40–41

Thieme, R., Pai, E. F., Schirmer, R. H. & Schulz, G. E. (1981) *J. Mol. Biol.* **152**, 763–782
Unden, G. & Kröger, A. (1980) *FEBS Lett.* **117**, 323–326
Valentin-Hansen, P. (1982) *EMBO J.* **1**, 1049–1054
Vik, S. B. & Hatefi, Y. (1981) *Proc. Natl. Acad Sci. U.S.A.* **78**, 6749–6753
Wierenga, R. K., Dreuth, J. & Schulz, G. E. (1983) *J. Mol. Biol.* **167**, 725–739
Wimpenny, J. W. & Cole, J. A. (1967) *Biochim. Biophys. Acta* **148**, 233–242
Young, I. G., Rogers, B. L., Campbell, H. D., Jaworowski, A. & Shaw, D. C. (1981) *Eur. J. Biochem.* **116**, 165–170