Biochem. J. (1984) **223**, 507–517
Printed in Great Britain

Nucleotide sequence encoding the iron–sulphur protein subunit of the succinate dehydrogenase of *Escherichia coli*

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

(*Received 18 May 1984/Accepted 27 June 1984*)

The nucleotide sequence of a 961 base-pair segment of DNA containing the *sdhB* gene, which encodes the iron–sulphur protein subunit of the *E. coli* succinate dehydrogenase, has been determined. The *sdhB* structural gene comprises 711 base pairs (237 codons, excluding the translational initiator and terminator). It is separated by a 15 base-pair intergenic region from the preceding flavoprotein gene (*sdhA*) and is the distal gene of an operon that also includes genes (*sdhC* and *D*) encoding two hydrophobic subunits, $\overrightarrow{sdhCDAB}$. The distal end of the *sdh* operon is linked to the 2-oxoglutarate dehydrogenase gene (*sucA*) by a complex region of dyad symmetry that is homologous with several potential intercistronic regulatory elements or transcriptional attenuators. The *sdhB* structural gene encodes a polypeptide of M_r 26637 that is strikingly homologous with the iron–sulphur protein subunit of fumarate reductase (38% identity, increasing to 58% when conservative changes are included). Both subunits contain 11 cysteine residues, 10 being conserved in three clusters resembling those found in ferredoxins. This work completes the sequence of a 9897 base-pair segment of DNA containing seven tricarboxylic acid cycle genes encoding three enzymes or enzyme complexes, citrate synthase (*gltA*), succinate dehydrogenase (*sdh*), and the 2-oxoglutarate dehydrogenase complex (*suc*), that are organized thus: $\overleftarrow{gltA}$-$\overrightarrow{sdhCDAB}$-$\overrightarrow{sucAB}$.

In *Escherichia coli*, the interconversion of succinate and fumarate is catalysed by two genetically-distinct enzymes, succinate dehydrogenase and fumarate reductase (EC. 1.3.99.1). They are membrane-bound flavoprotein complexes which participate in electron transfer but are synthesized under different physiological conditions (Hirsch *et al.*, 1963; Spencer & Guest, 1973, 1974*a*,*b*). Succinate dehydrogenase is synthesized during aerobic growth and it functions in the tricarboxylic acid cycle by transferring reducing equivalents to the respiratory chain. In contrast, fumarate reductase is synthesized during anaerobic growth and it functions as the terminal reductase of an anaerobic electron transport chain using fumarate as the ultimate acceptor.

Abbreviations used: SDH, succinate dehydrogenase; FRD, fumarate reductase; 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.

Succinate dehydrogenases and fumarate reductases from different sources invariably contain a large flavoprotein subunit (M_r 65000–72000) with a covalently-bound FAD coenzyme, and a smaller iron–sulphur protein subunit (M_r 25000–30000), associated with a cytochrome *b* or two hydrophobic subunits (Hederstedt & Rutberg, 1981; Davis *et al.*, 1977; Cole & Guest, 1982; Cole, 1984; Unden *et al.*, 1980). The flavoprotein and iron–sulphur protein subunits contain iron–sulphur centres detectable by e.s.r. spectroscopy, but in most cases, their exact location is uncertain. The succinate dehydrogenase of *E. coli* has proved refractory to purification and until recently structural information was based on studies with mutants (Spencer & Guest, 1974*a*) or immunoprecipitated material (Jones *et al.*, 1982; Owen & Condon, 1982). The succinate dehydrogenase genes (*sdh*) are located at 16.7min in a cluster of tricarboxylic acid cycle genes that includes the citrate synthase (*gltA*) and 2-oxoglutarate dehydrogenase complex (*sucA* and *sucB*) genes. This cluster has been cloned in phage and plasmid vectors (Spencer & Guest, 1982; Hull

et al., 1983) and studies with the cloned genes have confirmed the existence of two genes (*sdhA* and *sdhB*) encoding the flavoprotein (SDHA) and iron–sulphur protein (SDHB) subunits. The location and transcriptional organization of the genes has been established, as shown in Fig. 1. Furthermore, the sequence of a 3600 base-pair segment of DNA containing the *sdhA* gene and two proximal genes, *sdhC* and *sdhD*, encoding small hydrophobic subunits (SDHC and SDHD), has been defined (Wood *et al.*, 1984). This work paralleled earlier studies in which the fumarate reductase operon (*frdABCD*) was cloned (Cole & Guest, 1980*a*,*b*) and the structures of the corresponding gene products (FRDA,B,C and D) were established (Cole, 1982; Cole *et al.*, 1982; Grundström & Jaurin, 1982). In addition, it revealed a remarkable structural homology between the two flavoprotein subunits, and also the presence of analogous but not homologous pairs of small hydrophobic subunits.

The present work reports the sequence of the *sdhB* gene, encoding the iron–sulphur protein subunit (SDHB) of succinate dehydrogenase, and the marked similarity between the primary structures of the iron–sulphur protein subunits of succinate dehydrogenase and fumarate reductase. This work not only completes the sequence of the *sdh* genes, but also the sequence of a 9897 base-pair segment containing seven tricarboxylic acid cycle genes: $\overleftarrow{gltA}$-$\overrightarrow{sdhCDAB}$-$\overrightarrow{sucAB}$ (Ner *et al.*, 1983; Hull *et al.*, 1983; Darlison *et al.*, 1984; Spencer *et al.*, 1984; Wood *et al.*, 1984).

Experimental

Sources of DNA

The DNA fragments used for sequencing the *sdhB* gene were obtained from λG117, a *gltA sdhCDAB sucAB* transducing phage (Spencer & Guest, 1982) and pGS65, a pBR322 derivative containing a 4500 base-pair *Bam*HI fragment that expresses succinate dehydrogenase (Hull *et al.*, 1983; Wood *et al.*, 1984). Phage and plasmid DNA were prepared as described previously (Guest *et al.*, 1983) and restriction fragments were isolated after electrophoresis either by phenol extraction from low-melting-point agarose (Kühn *et al.*, 1979), or by electroelution and subsequent purification by chromatography on DE52 DEAE-cellulose (McDonell *et al.*, 1977; Smith, 1980).

Cloning in M13

Most of the nucleotide sequence was obtained by 'shot-gun' cloning the products of *Msp*I digestion of the *Xho*I–*Pst*I fragment of pGS65 (X–P_1 in Fig. 1) into the *Acc*I site of M13mp8 (Messing & Vieira, 1982). Specific fragments, generated by double digestion of λG117 or pGS65 (*Xho*I plus *Bam*HI, X–B_2; *Xho*I plus *Sst*I, X–St_2; *Hin*dIII plus *Nru*I, H–N_2) were cloned into the relevant sites of M13mp8, M13mp11 and M13mp9, respectively (Messing & Vieira, 1982). The *Hin*dIII–*Nru*I fragment of pGS65 (H–N_2 in Fig. 1) provided the sequence that overlapped the *Xho*I site and it was also the source of two M13mp8 derivatives with complementary *Msp*I inserts containing the *Xho*I site. Transfection of *E. coli* strain JM101 and the identification of recombinant plaques were by published procedures (Sanger *et al.*, 1980).

Nucleotide sequence analysis

Single-stranded DNA templates were prepared and sequenced by the dideoxy chain-termination method using a 17-nucleotide synthetic primer and the methods of Sanger *et al.* (1977). Nucleotide sequences were compiled and analysed with the Staden computer programs (Staden, 1979, 1980, 1982; Staden & McLachlan, 1982).

Materials

The sources of materials have been described previously (Stephens *et al.*, 1983; Darlison *et al.*, 1984; Wood *et al.*, 1984) except that the 17-nucleotide primer was purchased from Celltech and the T4 DNA ligase was from Boehringer Corp.

Results and discussion

Sequencing strategy

The iron–sulphur protein (SDHB) of the *E. coli* succinate dehydrogenase has an estimated M_r of 26000–28000 (Spencer & Guest, 1982; Cole & Guest, 1982; Jones *et al.*, 1982) and this corresponds to an *sdhB* coding region of approx. 800 base pairs. This coding region has been located between the succinate dehydrogenase flavoprotein gene (*sdhA*) and the 2-oxoglutarate dehydrogenase gene (*sucA*) using the maxicell and *in vitro* transcription–translation techniques (Wood *et al.*, 1984). Thus the *sdhB* gene is expressed from pGS65 and the 4.5kb *Bam*HI fragment, B_1–B_2 (Fig. 1), but not from plasmids such as pGS91, nor from the *Hin*dIII–*Xho*I fragment, H–X (Fig. 1), which lack the 1.4kb *Xho*I–*Bam*HI segment, X–B_2. Post-infection labelling studies with λ*glt-suc* transducing phages have also indicated that the *sdhB* gene is expressed with the same polarity as the *sdhA* and *sucA* genes (Spencer & Guest, 1982; Cole & Guest, 1982). Recent studies have also shown that the termination codon for the *sdhA* gene is just 24 base pairs upstream of the *Xho*I site (Wood *et al.*, 1984) and that the translation initiation site for the *sucA* gene is 276 base pairs upstream of a convenient *Pst*I site, P_1 (Darlison *et al.*, 1984), which indicates that the *sdhB* coding

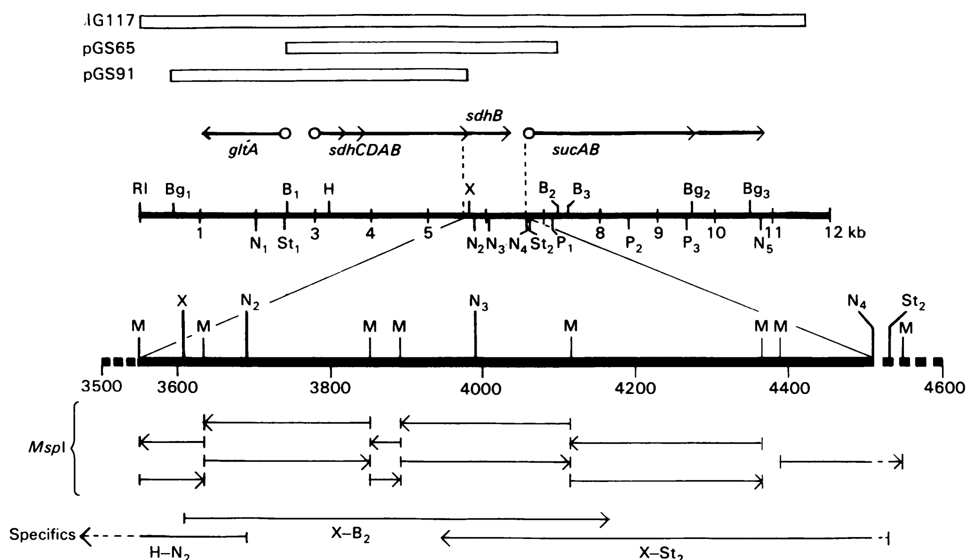

Fig. 1. *Organization of the succinate dehydrogenase genes of E. coli and summary of DNA sequence data obtained from M13 clones*

The segment of the *E. coli* linkage map at 16.7 min containing the genes encoding citrate synthase (*gltA*), succinate dehydrogenase (*sdhCDAB*) and the specific components of the 2-oxoglutarate dehydrogenase complex (*sucAB*), is aligned with the restriction map (drawn to scale; units kilobase pairs, kb). Left to right corresponds to clockwise in the *E. coli* linkage map and the positions and transcriptional polarities of the genes are indicated by the arrows above the restriction map. The segments of DNA cloned in a λ*gltA-sucB* transducing phage (λG117) and into the *Bam*HI and *Sal*I sites of pBR322 (pGS65 and pGS91) are denoted by the open bars. The restriction targets for *Bam*HI (B), *Bgl*II (Bg), *Eco*RI (RI), *Hin*dIII (H), *Nru*I (N), *Pst*I (P), *Sst*I (St) and *Xho*I (X) have been defined by restriction mapping and the *Msp*I (M) sites were identified in the 'shot-gun' cloning and sequencing. The arrows below the expanded section of the restriction map show the positions, directions and extents of the sequences obtained from M13 clones. The nucleotide co-ordinates in the *sdhB* region are numbered from the *Nru*I target (N_1) according to Wood *et al.* (1984).

region is located mainly in the 1.3 kb *Xho*I–*Pst*I fragment ($X–P_1$; Fig. 1).

The overall strategy that has been adopted for sequencing the *sdhB* gene has involved combining a 'shot-gun' analysis of the $X–P_1$ fragment with the cloning and sequencing of three specific fragments that were needed to align the 'shot-gun' clones ($X–B_2$ and $X–St_2$) and to overlap the *Xho*I target ($H–N_2$).

Nucleotide sequence and location of coding region

A restriction map showing the locations, directions and extents of sequencing of the M13 clones used to analyse the *sdhB* region is shown in Fig. 1, and the complete and unambiguous sequence of a 961 base-pair segment containing the *sdhB* gene is shown in Fig. 2. The nucleotides are numbered from the first base of the *Nru*I site (N_1) in the *gltA* gene (Fig. 1) and the sequence extends the previously reported 3614 base-pair sequence by a further 898 base pairs to another *Nru*I site (N_4, final co-ordinate 4512; Fig. 2). All of the sequence was obtained from at least two independent clones and 97.4% (100% of the coding region) was obtained from both DNA strands.

The computer program FRAMESCAN (Staden & McLachlan, 1982), which predicts regions of coding DNA by a statistical analysis of codon usage in all reading frames, was used to search both strands and only one coding region was detected. It corresponds to a 714 base-pair open reading frame (co-ordinates 3603–4316) having the same transcriptional polarity as the other *sdh* genes (Wood *et al.*, 1984). The coding region starts with an AUG initiation codon, ends with an UAA (ochre) codon, and encodes a polypeptide of M_r 26637. It therefore has the size, position and polarity predicted for the *sdhB* structural gene.

Features of the nucleotide sequence

The putative *sdhB* coding region is preceded by a well-placed ribosome binding site having four consecutive bases, d(G-G-A-G) at co-ordinates 3591–3594, that are complementary to the 3′-terminal sequence of 16S ribosomal RNA (Fig. 2). This translation initiation region satisfies seven of

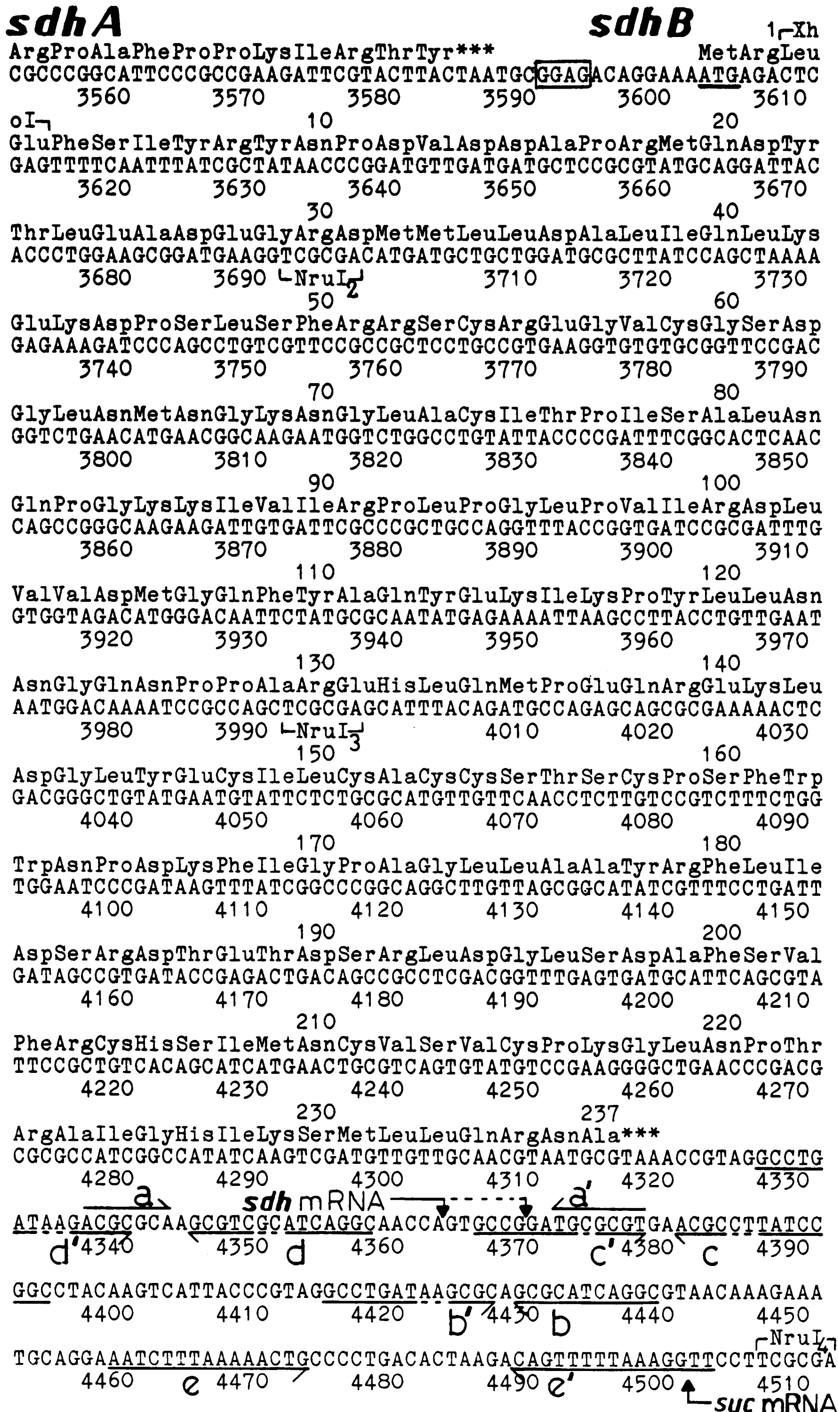

Fig. 2. *Nucleotide sequence of the sdhB gene and primary structure of the succinate dehydrogenase iron–sulphur protein subunit*

The nucleotide sequence of 961 base pairs containing the non-coding (sense strand) of the succinate dehydrogenase iron–sulphur protein gene (*sdhB*) is shown. The intergenic region between the succinate dehydrogenase flavoprotein gene (*sdhA*) and *sdhB*, and part of the intergenic region between the *sdhB* gene and the 2-oxoglutarate dehydrogenase gene (*sucA*) are also included. The latter contains three highly-conserved elements of hyphenated dyad symmetry denoted by the converging arrows (designated d′d, c′c and b′b), and two unrelated sequences (aa′ and ee′), as well as a termination and initiation site for the transcripts of the *sdh* and *suc* operons, respectively. The nucleotide

the eight criteria defined by Stormo *et al.* (1982) and Atkins' (1979) rule, the nearest upstream stop codon being the UAA (ochre) termination site for *sdhA* (co-ordinates 3585–3587). The intergenic region between *sdhA* and *sdhB* comprises 15 base pairs (excluding the translational stop and start codons), so translation would be re-initiated without changing the reading frame. It is very rich in purine bases, which is typical of many bacterial intercistronic regions (Stephens *et al.*, 1983), and because there are no obvious promoter and terminator sequences (Hawley & McClure, 1983; Rosenberg & Court, 1979) in the regions flanking the intergenic region, it is concluded that the *sdhB* gene is transcribed as part of the *sdh* operon. The sequence downstream of the *sdhB* termination codon appears not to contain a typical *rho*-dependent or *rho*-independent transcriptional terminator. However, the *sdhB–sucA* intergenic region is particularly complex with respect to regions of hyphenated dyad symmetry, and this is indicated by the converging arrows in Fig. 2. The most striking feature is the presence of an IS-like intercistronic regulatory element analogous to those described by Higgins *et al.* (1982). It comprises three highly-conserved elements of hyphenated dyad symmetry, d′d(4327–4358), c′c (4366–4394) and b′b(4415–4439), plus an unrelated segment, aa (4337–4378), which could all form stable stem-and-loop structures in the corresponding transcripts. The significance of this region and an analogous segment at the distal end of the *sucB* gene has been discussed previously with respect to the transcriptional organization of the *suc* genes (Darlison *et al.*, 1984; Spencer *et al.*, 1984). However, recent attempts to identify transcripts of this region have revealed the existence of an *sdh* transcript which terminates just distal to the first element d′d at co-ordinates 4363–4369 (Fig. 2; M. E. Spencer, personal communication). Thus it is concluded that the *sdhB* gene is the distal gene of the succinate dehydrogenase operon and that the potential stem-and-loop structure, d′d, forms part of the termination apparatus. Evidence has also been obtained for the initiation of a *suc* mRNA at co-ordinate 4502, but no transcript spanning the *sdhB* and *sucA* genes has been detected (M. E. Spencer, personal communication). This suggests that the *suc* genes are expressed independently of the *sdh* genes and that the large *sdhB–sucA* intergenic element is probably not functioning as an intercistronic regulatory element to decrease the expression of distally-related genes, as has been suggested for other comparable structures (Higgins *et al.*, 1982; Valentin-Hansen *et al.*, 1984).

Codon usage. The codon usage of the *sdhB* gene is compared with that of the analogous *frdB* gene in Table 1, and the patterns are non-random and typical of many *E. coli* genes. By examining the use of modulatory codons corresponding to minor tRNA species, and the preferences for codons with intermediate interaction energies in a diagnostic set of eight pairs of codons, it is possible to assess the degree of expression of the genes (Grosjean & Fiers, 1982). The proportions of modulatory codons (2.5%) and optimal energy codons (46%) resemble those in the *sdhA,C* and *D* genes and the *frdA,C* and *D* genes, and suggest that *sdhB* is not strongly expressed (Wood *et al.*, 1984). This contrasts with the *frdB* gene, which resembles a highly expressed gene, because it uses a high proportion of optimal energy codons (65%) and no modulatory codons. The reason for this difference is not known.

Primary structure and composition of the iron–sulphur protein subunit (SDHB)

The primary structure of the *sdhB* gene product (SDHB), deduced from the nucleotide sequence, is shown in Fig. 2. It comprises a polypeptide of 237 residues, assuming that the initiating formylmethionine residue is removed post-translationally. The M_r predicted for SDHB, 26637, is in excellent agreement with previous estimates for the iron–sulphur protein subunit, M_r 26000–28000. A comparison of size, composition and polarity for the iron–sulphur protein subunits of three succinate dehydrogenases and a fumarate reductase is presented in Table 2. The four proteins resemble each other with respect to size and polarity, but there appear to be significant differences as well as similarities in composition. The cysteine contents of the iron–sulphur protein subunits of the succinate dehydrogenase (SDHB) and the fumarate reductase (FRDB) of *E. coli* are identical (11 residues/mol). Moreover, they are distributed in three clusters at approximately the same sites in the two polypeptide chains, and their organization resembles that found in several ferredoxins (Fig. 3). This contrasts with the corresponding flavoprotein subunits where the cysteine residues (11 in SDHA and 10 in FRDA) appear to be more randomly dispersed and only one occupies the same position in both polypep-

co-ordinates are numbered relative to the first base of a *Nru*I site (N_1) in the *gltA* structural gene (Wood *et al.*, 1984). A potential ribosome-binding site is boxed, the proposed translation initiation site is underlined and relevant stop sites are denoted by asterisks thus: ***. The primary structure of the *sdhB* gene product is shown directly above the nucleotide sequence.

Table 1. *Codon usage in the sdhB and frdB genes*

The AUG initiation codons are not included with the methionine codons in this table. Data for the *frdB* gene derives from Cole *et al.* (1982). The codon pairs enclosed in boxes are those whose use varies between strongly and weakly expressed genes, and the arrows indicate potential modulatory codons (Grosjean & Fiers, 1982).

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

Table 2. *Amino acid compositions of the iron–sulphur protein subunits of three succinate dehydrogenases and a fumarate reductase*

The amino acid composition predicted for the iron–sulphur protein subunit of the *E. coli* succinate dehydrogenase is compared with the analytical values for the corresponding subunits of the *Rhodospirillum rubrum* and beef heart mitochondrial enzymes (Davis *et al.*, 1977) and with the DNA-derived composition of the *E. coli* fumarate reductase iron–sulphur protein subunit (Cole *et al.*, 1982). The initiating formylmethionine residues are not included.

	Composition of iron–sulphur protein subunits of:			
	Succinate dehydrogenase			Fumarate reductase
Amino acid	*E. coli*	*R. rubrum*	Beef heart	*E. coli*
Asp	18	24 (Asp + Asn)	28 (Asp + Asn)	14
Asn	12			13
Thr	6	11	14	12
Ser	16	9	14	14
Glu	11	21 (Glu + Gln)	24 (Glu + Gln)	13
Gln	9			9
Pro	17	17	14	14
Gly	16	16	14	14
Ala	14	21	20	24
Val	9	14	7	14
Met	8	4	6	8
Ile	14	12	17	14
Leu	27	21	20	17
Tyr	8	5	11	11
Phe	8	7	6	9
Lys	11	13	20	14
His	3	3	3	6
Arg	14	14	12	10
Cys	11	9	7	11
Trp	2	6	4	2
Total	237	227	241	243
M_r	26637	25223	27188	26990
Polarity (%)	43	42	48	43

	Organism	Protein	Sequence
I	*E. coli*	SDHB	53 S C R E G V C G S D G L N M N G K N G L A C 74
	E. coli	FRDB	56 S C R M A I C G S C G M M V N N V P K L A C 77
	S. platensis	Fd [2Fe-2S]	40 S C R A G A C S T C A (26 residues) L T C 79
II	*E. coli*	SDHB	147 E C I L C A C C S T S C P S F W 162
	E. coli	FRDB	147 G C I N C G L C Y A A C P Q F G 162
	P. aerogenes	Fd [4Fe-4S]	7 S C I A C G A C K P E C P V N I 22
	A. vinelandii	Fd [3Fe-3S]	7 N C I K C (4) C V E V C P V D C 24
	T. acidophilum	Fd [4Fe-4S]	122 D C I F C M A C E S V C P V R A 137
III	*E. coli*	SDHB	204 R C H S I M N C V S V C P 216
	E. coli	FRDB	203 S C T F V G Y C S E V C P 215
	P. aerogenes	Fd [4Fe-4S]	33 S C I D C G S C A S V C P 45
	A. vinelandii	Fd [4Fe-4S]	38 E C I D C A L C E P E C P 50
	T. acidophilum	Fd [4Fe-4S]	67 C C I A D G A C M D V C P (35) C 115

Fig. 3. *Comparison of the cysteine clusters of the iron–sulphur protein subunit of succinate dehydrogenase (SDHB) with the corresponding subunit of fumarate reductase (FRDB) and some bacterial ferredoxins*

The three cysteine-containing regions (I, II and III) of the *E. coli* SDHB subunit are compared with analogous regions of the *E. coli* FRDB subunit (Cole *et al.*, 1982) and the ferredoxins of *Spirulina platensis* (Tanaka *et al.*, 1976), *Peptococcus aerogenes* (Tsunoda *et al.*, 1968), *Azotobacter vinelandii* (Ghosh *et al.*, 1982), and *Thermoplasma acidophilum* (Wakabayashi *et al.*, 1983). The sequences are in one-letter code and arranged to align the cysteine residues of potential iron–sulphur binding sites. The numbers in parentheses refer to residues not shown, and those at the ends denote the amino acid positions of the adjacent residues.

tides (Wood *et al.*, 1984). The possible functions of the clusters of cysteine residues in the iron–sulphur protein subunits will be discussed more fully below.

Structural comparison between the iron–sulphur protein subunits of the succinate dehydrogenase and fumarate reductase of E. coli

The primary structures of SDHB and FRDB were analysed for homologies by using the computer program DIAGON (Staden, 1982). This program incorporates the MDM_{78} mutation data matrix of Schwartz & Dayhoff (1978) that is very good for detecting distant relationships between amino acid sequences. Regions of homology are indicated by diagonal lines, as shown by the line of identity in the self-comparison for SDHB and the two small regions of potential internal duplication (*aa'* and *bb'*, Fig. 4*a*). The former (*aa'*) corresponds to limited homology between the second and third cysteine clusters and the latter (*bb'*) identifies a region close to the third cysteine cluster that is related to the *N*-terminal segment. The self-comparison of FRDB also reveals homology between the second and third cysteine clusters (*cc'*) as well as some other small regions of unknown significance (Fig. 4*b*). The results suggest that the homologies between the cysteine clusters are slightly more extensive in FRDB than in SDHB. The third comparison matrix (Fig. 4*c*) shows that there is a remarkably high degree of well-aligned homology between the two iron–sulphur proteins. The homologies between the second and third cysteine clusters can again be seen (*dd'*) and it is apparent that the second clusters of both proteins and the third cluster of FRDB are more extensively related to each other than to the third cluster of SDHB.

An alignment, based on the DIAGON comparison but refined by visual inspection, is shown in Fig. 5. Some 87 of the 230 equivalenced residues (38%) are identical and a further 47 residues (20%) are conservatively replaced at the $\geqslant 0.10$ scoring limit of MDM_{78}. The divergence from complete colinearity for SDHB and FRDB is due to just six insertion-deletions of one or two residues, at positions 25, 30, 84, 167 (+1 residue), 90 (+2 residues) and 44(−1 residue) in SDHB, and the larger size of FRDB is due to extra residues at the *N*- and *C*-terminal ends (four and seven residues,

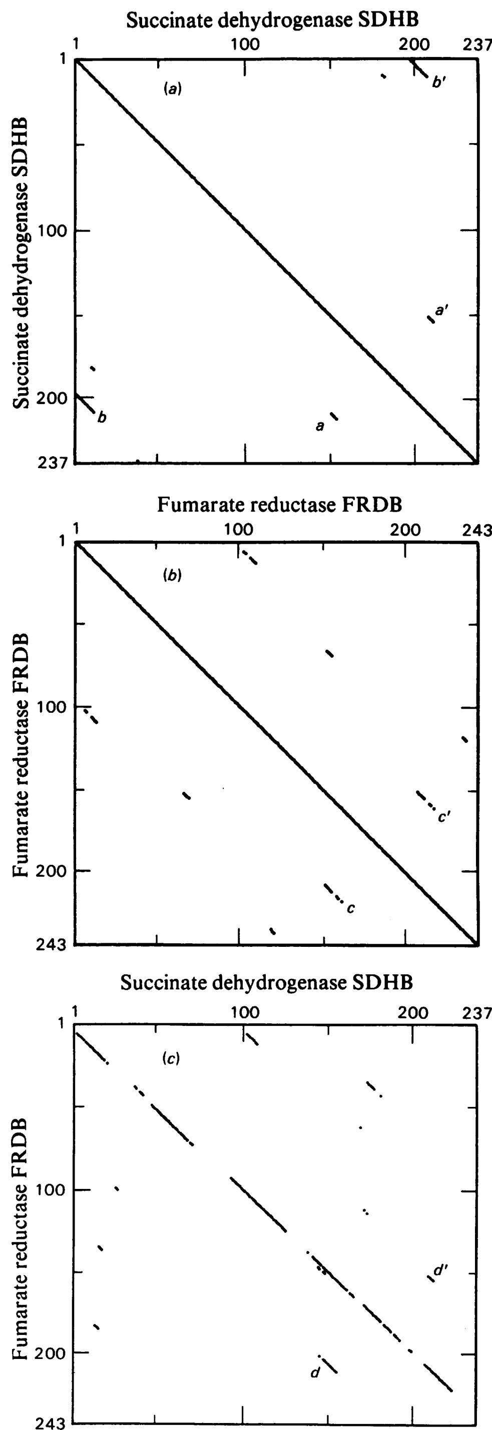

Fig. 4. *Amino acid sequence comparison matrices for the iron–sulphur protein subunits of succinate dehydrogenase and fumarate reductase*

The matrices depict self-comparisons for SDHB (*a*) and FRDB (*b*), and a cross-comparison for SDHB and FRDB (*c*). The DIAGON program of Staden (1982) was used and the points correspond to mid-points of 15-residue spans giving a double matching probability of <0.001. The regions marked *aa'*, *cc'* and *dd'* correspond to homologies between the second and third cysteine clusters (II and III) of the two subunits (see the text).

respectively). The distribution of hydrophobic and hydrophilic sequences, as revealed by HYDROPLOT analysis (Fig. 6), are fairly similar for the two proteins, as would be expected from the amino acid sequence homology. It is also interesting to note that the clusters of cysteine residues are located mainly in the most hydrophobic regions of both iron–sulphur proteins. The high degree of sequence homology observed for the proteins is also found at the DNA level when the nucleotide sequences of the *sdhB* and *frdB* genes are compared. The similarity between the iron–sulphur proteins and their genes closely parallels those already described for the flavoproteins, SDHA and FRDA, and their genes (Wood *et al.*, 1984), and it suggests that they have evolved by the duplication of common ancestral genes or of a common ancestral operon.

The iron–sulphur centres. The nature and location of the iron–sulphur centres in the succinate dehydrogenases and fumarate reductases from several sources is uncertain. The traditional view for the beef heart mitochondrial succinate dehydrogenase is that the flavoprotein contains two [2Fe–2S] clusters and the iron–sulphur protein contains a single [4Fe–4S] cluster, but there is an alternative view that the [4Fe–4S] cluster is in the flavoprotein and that there is only one [2Fe–2S] cluster, which is located in the iron–sulphur protein (Beinert & Albracht, 1982). The dehydrogenases from photosynthetic bacteria and *Bacillus subtilis* appear to fit the traditional pattern (Hederstedt & Rutberg, 1981), whereas the fumarate reductase of *Vibrio succinogenes* has the alternative configuration (Albracht *et al.*, 1981). E.s.r. spectroscopy suggests that the *E. coli* fumarate reductase contains a [2Fe–2S] centre in the iron–sulphur protein plus a high-potential cluster (possibly [4Fe–4S]) of uncertain location (Cole, 1984). The elucidation of the primary structures of the fumarate reductase and succinate dehydrogenase of *E. coli* offers further insights into the nature of the iron–sulphur centres, by comparing the distribution of cysteine residues. This is because studies with ferredoxins have shown that the cysteine residues at the iron–sulphur centres of plant and cyanobacterial [2Fe–2S] clusters and bacterial [3Fe–3S] and [4Fe–4S] clusters are strongly conserved (Tsukihara *et al.*, 1982; Cammack, 1983).

The cysteine residues of the flavoprotein subunits, 11 in SDHA and 10 in FRDA, are scattered

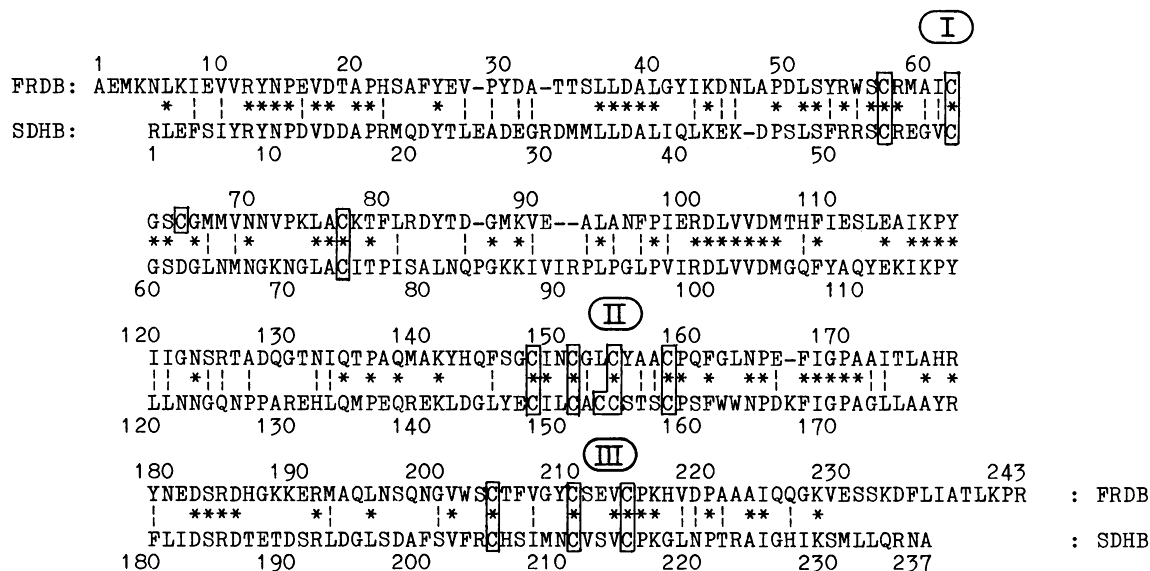

Fig. 5. *Alignment of the amino acid sequences of the iron–sulphur protein subunits of succinate dehydrogenase and fumarate reductase*

The sequences have been aligned for maximum homology based on the DIAGON comparison (Fig. 4*c*). Identical residues are identified by asterisks and conservatively-substituted residues with scores ⩾0.1 in the MDM_{78} matrix (Schwartz & Dayhoff, 1978) are denoted by vertical bars. The cysteine residues are boxed and the three clusters are labelled I, II and III. The FRDB sequence is from Cole *et al.* (1982) and the polypeptide chains are numbered from their *N*-termini, excluding the initiating formylmethionine residues. The dashes within the sequences identify sites of insertion and deletion.

with only one residue conserved, and they do not adopt distinctive ferredoxin-like clusters (Wood *et al.*, 1984; Cole, 1982). This does not exclude the formation of iron–sulphur centres at the tertiary structural level, as in the nitrogenase iron–sulphur proteins, nor for centres to be formed between the flavoprotein and iron–sulphur protein subunits. In this connection it may be relevant that the cysteine residues of the ferredoxin-like clusters occur in hydrophobic segments of the polypeptide chain and that several flavoprotein cysteines are located in hydrophobic regions. These include: the conserved cysteine residue, 256 in SDHA (247 in FRDA); residue 390 in SDHA, which is very close to residue 380 in FRDA in a highly-conserved segment; and residues 179, 183, 554 (in SDHA) and 332, 517 (in FRDA).

In the iron–sulphur proteins 10 of the 11 cysteine residues occupy identical positions in the alignment shown in Fig. 5, and they form three clusters that resemble those found in ferredoxins (Fig. 3). Cammack (1983) has pointed out the similarity between the first cluster in FRDB and the arrangement in the [2Fe–2S] centres of plant and cyanobacterial ferredoxins, e.g. *Spirulina platensis* (Fig. 3). The corresponding cluster in SDHB retains much of the homology but lacks one of the cysteine residues. Unless the aspartate residue (62) is functional, another cysteine residue would have to be contributed by the SDHA subunit or by one of the other clusters in the SDHB subunit. The arrangement of residues in the second and third clusters of SDHB and FRDB is similar to that found in the bacterial ferredoxins, e.g. *Peptococcus aerogenes*, 2 [4Fe–4S], and *Azotobacter vinelandii*, [3Fe–3S][4Fe–4S] (Fig. 3). The third clusters each lack the second cysteine residue so they may depend on residues contributed by the first clusters or the flavoprotein subunits. However, the arrangement in the third clusters is the same as that found in the unique cluster of the ferredoxin of the thermophilic and acidophilic bacterium, *Thermoplasma acidophilum* (Wakabayashi *et al.*, 1983). Here the unique cluster (residues 67–79), which precedes rather than follows a typical cluster (residues 122–137), may be completed by the cysteine residue at position 115 to produce what might be a two-[4Fe–4S] ferredoxin (Fig. 3).

The primary structures of the iron–sulphur protein subunits are consistent with them being the principal sites for several combinations of two iron–sulphur centres (single [2Fe–2S] and [4Fe–4S] centres, or two [4Fe–4S] centres), or even three centres (one [2Fe–2S] and two [4Fe–4S] centres) if

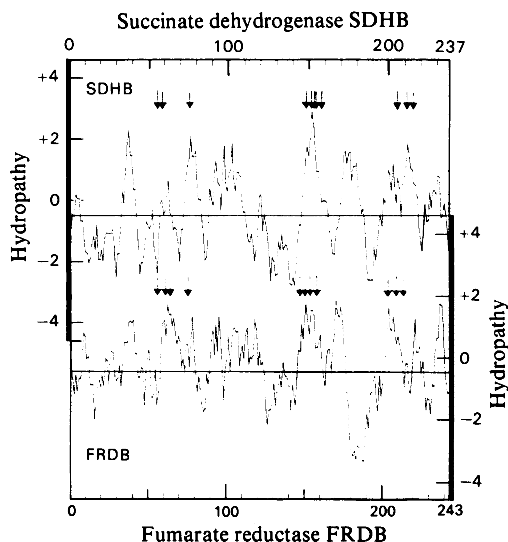

Fig. 6. *Comparison of the hydropathy profiles of the iron–sulphur protein subunits of succinate dehydrogenase and fumarate reductase*

The hydropathy profiles of SDHB (upper panel) and FRDB (lower panel) were derived using the range +4.5 to −4.5 according to Kyte & Doolittle (1982) and horizontal lines representing the average for most sequenced proteins are included. Consecutive hydropathy averages are plotted for a seven-residue segment as it advances from *N*- to *C*-terminus. The positions of the cysteine residues are indicated by arrows.

cysteine residues are contributed by the flavoprotein subunits. However, as pointed out by Cammack (1983), extrapolation from sequence data alone can be misleading, and it is possible that the second and third clusters could represent single centres of the [3Fe–xS] type that need five or six cysteine residues and are now thought to occur in *Azotobacter vinelandii* and *Desulphovibrio gigas* ferredoxins. Because the iron–sulphur protein subunits of succinate dehydrogenase and fumarate reductase appear to be extremely similar, it is presumed that differences in their redox potentials and other specific features of the iron–sulphur clusters are determined by differences in their environments.

Conclusion

The sequence of the 961 base-pair *sdhB* segment completes the sequence of the region encoding the *E. coli* succinate dehydrogenase. It also links two larger segments of DNA: (i) a 4599 base-pair segment encoding citrate synthase and the succinate dehydrogenase C, D and A subunits (Ner *et al.*, 1983; Wood *et al.*, 1984); and (ii) a 4620 base-pair segment encoding the specific E1o and E2o components of the 2-oxoglutarate dehydrogenase complex, *sucAB* (Darlison *et al.*, 1984; Spencer *et al.*, 1984). Together these form a continuous sequence of 9897 base pairs encoding what is the largest cluster of tricarboxylic acid cycle genes in the *E. coli* chromosome. The genes appear to be organized in three transcriptional units:

but the significance of the clustering and the potential transcription attenuator structures at each end of the *suc* operon, has still to be defined.

When compared with the analogous fumarate reductase transcriptional unit, *frdABCD*, the most notable features of the succinate dehydrogenase region are the striking sequence similarities between the corresponding flavoprotein and iron–sulphur protein subunits, the presence of equivalent but not homologous pairs of hydrophobic subunits, and the different gene order. It would appear that the flavoprotein and iron–sulphur protein subunits have evolved from common ancestors and that their functional compartmentation in aerobic and anaerobic respiration is primarily due to the regulatory mechanisms controlling the expression of the two operons. This work illustrates the value of gene-cloning and sequence analysis to characterize a membrane-bound enzyme that had hitherto resisted purification. It should now be possible to subclone segments of DNA encoding individual subunits or pairs of subunits in appropriate expression vectors for advancing studies on the proteins. In particular, amplification of the SDHA and SDHB proteins in the absence of the hydrophobic subunits may facilitate purification and further structural studies with the enzyme. The ability to create hybrids containing different segments of the dehydrogenase and reductase should also help to identify the regions concerned with the functional specificities of the two enzymes.

We thank Roger Staden and Ian Duckenfield for generously supplying and implementing the computer programs and R. Cammack for advice about interpreting the structures of potential iron–sulphur centres. We also acknowledge support from the Science and Engineering Research Council by project grant (GR/B 91396) to J. R.G.

References

Albracht, S. P. J., Unden, G. & Kröger, A. (1981) *Biochim. Biophys. Acta* **661**, 295–302

Atkins, J. F. (1979) *Nucleic Acids Res.* **7**, 1035–1041

Beinert, H. & Albracht, S. J. P. (1982) *Biochim. Biophys. Acta* **683**, 245–277

Cammack, R. (1983) *Chem. Scripta*, **21**, 87–95
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. (1980*a*) *Mol. Gen. Genet.* **178**, 409–418
Cole, S. T. & Guest, J. R. (1980*b*) *Mol. Gen. Genet.* **179**, 377–385
Cole, S. T. & Guest, J. R. (1982) *Biochem. Soc. Trans.* **10**, 473–475
Cole, S. T., Grundström, T., Jaurin, B., Robinson, J. J. & Weiner, J. H. (1982) *Eur. J. Biochem.* **126**, 211–216
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
Ghosh, D. O., O'Donnell, S., Furey, W., Jr., Robbins, A. H. & Stout, C. D. (1982) *J. Mol. Biol.* **158**, 73–109
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., Roberts, R. E. & Stephens, P. E. (1983) *J. Gen. Microbiol.* **129**, 671–680
Hawley, D. K. & McClure, W. R. (1983) *Nucleic Acids Res.* **11**, 2237–2255
Hederstedt, L. & Rutberg, L. (1981) *Microbiol. Rev.* **45**, 542–555
Higgins, C. F., Ferro-Luzzi Ames, G., Barnes, W. M., Clement, J. M. & Hofnung, M. (1982) *Nature (London)* **298**, 760–762
Hirsch, C. A., Raminsky, M., Davis, B. B. & Lin, E. C. C. (1963) *J. Biol. Chem.* **238**, 3370–3374
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
Kyte, J. & Doolittle, R. F. (1982) *J. Mol. Biol.* **157**, 105–132
Kühn, S., Fritz, H.-J. & Starlinger, P. (1979) *Mol. Gen. Genet.* **167**, 235–241
McDonell, M. W., Simon, M. N. & Studier, F. W. (1977) *J. Mol. Biol.* **110**, 119–146
Messing, J. R. & Vieira, J. (1982) *Gene* **19**, 269–276
Ner, S. S., Bhayana, V., Bell, A. W., Giles, I. G., Duckworth, H. W. & Bloxham, D. P. (1983) *Biochemistry* **22**, 5243–5249
Owen, P. & Condon, C. (1982) *FEMS Microbiol. Lett.* **14**, 223–227
Rosenberg, M. & Court, D. (1979) *Annu. Rev. Genet.* **13**, 319–353
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) in *Atlas of Protein Sequence and Structure*, vol. 5, suppl. 3, pp. 353–358, National Biomedical Research Foundation, Washington DC
Smith, H. O. (1980) *Methods Enzymol.* **65**, 371–380
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., Darlison, M. G., Lewis, H. M. & Guest, J. R. (1983) *Eur. J. Biochem.* **133**, 481–489
Stormo, G. D., Schneider, T. D. & Gold, L. M. (1982) *Nucleic Acids Res.* **10**, 2971–2996
Tanaka, M., Haniu, M., Yasunobu, K. T., Rao, K. K. & Hall, D. D. (1976) *Biochem. Biophys. Res. Commun.* **69**, 759–765
Tsunoda, I. N., Yasunobu, K. T. & Whiteley, H. R. (1968) *J. Biol. Chem.* **243**, 6262–6272
Tsukihara, T., Kobayashi, M., Nakamura, M., Katsube, Y., Fukuyama, K., Hase, T., Wada, K. & Matsubara, H. (1982) *Biosystems* **15**, 243–257
Unden, G., Hackenberg, H. & Kröger, A. (1980) *Biochim. Biophys. Acta* **591**, 275–288
Valentin-Hansen, P., Hammer-Jespersen, K., Boetius, F. & Svendsen, I. (1984) *EMBO J.* **3**, 179–183
Wakabayashi, S., Fujimoto, N., Wada, K., Matsubara, H., Kerscher, L. & Oesterhelt, D. (1983) *FEBS Lett.* **162**, 21–24
Wood, D., Darlison, M. G., Wilde, R. J. & Guest, J. R. (1984) *Biochem. J.* **222**, 519–534