Identification, sequence determination, and expression of the flavodoxin gene from Desulfovibrio salexigens

ARTICLE *in* BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS · MAY 1990

Impact Factor: 2.3 · DOI: 10.1016/0006-291X(90)92393-E · Source: PubMed

Vol. 168, No. 2, 1990 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
April 30, 1990 Pages 809-817

IDENTIFICATION, SEQUENCE DETERMINATION, AND EXPRESSION OF THE FLAVODOXIN GENE FROM DESULFOVIBRIO SALEXIGENS

Larry R. Helms, Grigorios D. Krey, and Richard P. Swenson

Department of Biochemistry, The Ohio State University, Columbus, OH 43210

Received March 27, 1990

SUMMARY: Restriction fragments of genomic DNA from Desulfovibrio salexigens (ATCC 14822) containing the structural gene coding for the flavodoxin protein were identified using the entire coding region of the gene for the Desulfovibrio vulgaris (Hildenborough) flavodoxin as a probe (Krey, G.D., Vanin, E.F., and Swenson, R.P. (1988) J. Biol. Chem. 263, 15436-15443). A 1.4-kb PstI-HindIII fragment was ultimately identified which contains an open reading frame coding for a polypeptide of 146 amino acid residues that was highly homologous to the D. vulgaris flavodoxin, sharing a sequence identity of 55%. When compared to the X-ray crystal structure of the D. vulgaris protein, the homologous regions were largely confined to those portions of the protein which are in the immediate vicinity of the flavin mononucleotide cofactor binding site. Tryptophan-60 and tyrosine-98, which reside on either side of the isoalloxazine ring of the cofactor, are conserved, as are the sequences of the polypeptide loop that interacts with the phosphate moiety of the flavin. Acidic residues forming the interface of model electron-transfer complexes with certain cytochrome c proteins are retained. The flavodoxin holoprotein is over-expressed in E. coli from the cloned gene using its endogenous promoter.

Flavodoxins have been identified in several members of the Desulfovibrio species including D. gigas, D. salexigens, D. desulfuricans, and D. vulgaris. In each case, the protein has been isolated and characterized (1). However, within this group only the flavodoxin from D. vulgaris has had its primary sequence and X-ray crystal structure elucidated (2). The tertiary structures of this flavodoxin and those from a variety of other sources are remarkably homologous; however, the amino acid sequences of this family of proteins are quite different (3, 4). Sequence variation extends over all portions of the structure, including the FMN binding site, greatly increasing the uncertainties in the identification of functionally important amino acid residues by visual inspection of the flavodoxin crystal structures themselves. However, examination of a closely related family of flavodoxin protein sequences should more readily localize homologous regions which are likely to be functionally significant. For example, these regions might be important in the regulation of the redox properties of the bound cofactor and in electron transfer processes. Recently, we reported the nucleotide sequence of the structural gene for the

0006-291X/90 $1.50

flavodoxin from D. vulgaris (5). Using the coding region from this gene as a probe, we report here the cloning and sequence of a fragment of genomic DNA from Desulfovibrio salexigens which codes for the flavodoxin from this organism.

MATERIALS AND METHODS: Desulfovibrio salexigens cultures obtained from American Type Culture Collection (ATCC 14822) were revived and grown in modified Baar's medium under anaerobic conditions (6). E. coli strains XL1-Blue (Stratagene) and AG1 were grown in LB medium. Genomic DNA was purified as described previously (5).

A 450-bp HindIII fragment (designated DVF450H3) containing the entire coding region of the D. vulgaris flavodoxin (DVF) gene (5) was radiolabeled by a modification of the method of Feinberg and Vogelstein (7) and used as a probe for flavodoxin structural genes. DNA restriction fragments that putatively contain the D. salexigens flavodoxin (DSF) gene were identified by hybridization of Southern blots of ClaI-digested D. salexigens genomic DNA with the radiolabeled DVF450H3 probe as follows. The ClaI restriction fragments were separated by agarose gel electrophoresis and transferred to GeneScreen Plus membranes (Dupont) by the Southern blot procedure. Membrane filters were prehybridized by established methods and hybridized overnight at 55°C with the radiolabeled probe (6.2×10^6 dpm). Filters were washed twice, each for 20 min at 55°C in 3x SSC/0.5% SDS, dried, and subjected to autoradiography.

A genomic mini-library was constructed using the Bluescript vector (Stratagene) as follows. Genomic DNA was digested to completion with the restriction enzyme ClaI and fragments size-selected by centrifugation though a linear NaCl gradients as described previously (5). Appropriately sized fragments were ligated into the ClaI site in the Bluescript vector using conditions described by Grundström et al. (8). E. coli XL1-Blue cells were transformed by the procedure described by Hanahan (9) and transformation mixtures plated onto GeneScreen Plus membranes which had been placed on LB media supplemented with ampicillin (75 µg/ml) and tetracycline (15 µg/ml). Colony hybridizations were as follows. All plates were replica plated onto GeneScreen Plus membranes and allowed to grow to 1-2mm in diameter (ca. 8 hrs at 37°C). The filters were first laid for 7 min on Whatman 3MM filter paper saturated with 0.5M NaOH and then washed with 1M Tris-HCl, pH 7.4 for 5 min and finally for 5 min with 0.5M Tris-HCl, pH 7.4, containing 1.5M NaCl. The filters were then scrubbed with a gloved hand in 6x SSC/0.5% SDS at 25°C and allowed to air dry. Filters were prehybridized at 65°C with sonicated and denatured Bluescript vector (5 µg/ml) to minimize nonspecific interactions of the probe with this cloning vector. Hybridization with radiolabeled DVF450H3 proceeded overnight at 65°C. The filters were washed twice in 1x SSC/0.5% SDS at 65°C, each for 20 min. Positive colonies were visualized by autoradiography and rescreened by hybridization at 65°C of the DVF450H3 probe to Southern blots of ClaI-digested recombinant plasmid preparations as described above. Restriction fragments were subcloned into Bluescript for restriction mapping and sequence determination. Nucleotide sequences were determined using the dideoxy nucleotide chain termination method of Sanger et al. (10). All other recombinant DNA procedures were taken from Maniatis et al. (11) unless otherwise noted. DNA and protein sequences were analyzed using software from DNAstar, Inc. The amino acid similarity (PAM) matrix described by Lipman and Pearson (12) was used in the protein sequence comparisons. Purification of the DSF protein expressed in E. coli and SDS-polyacrylamide gel electrophoresis were carried out essentially as described by Krey et al. (5).

RESULTS AND DISCUSSION

Isolation of the gene - We have previously cloned and determined the nucleotide sequence for the flavodoxin from Desulfovibrio vulgaris (5). In this study, the entire coding region of the D. vulgaris flavodoxin (DVF) gene (as the

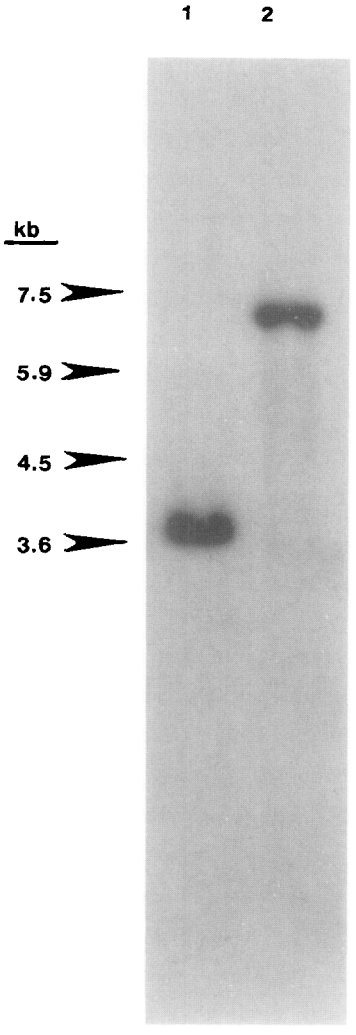

Figure 1. Autoradiogram of the Southern blot of Desulfovibrio salexigens genomic DNA digested with the restriction endonucleases ClaI (**Lane 1**) and NcoI (**Lane 2**). Hybridization was to a radiolabeled 450bp HindIII fragment containing the entire coding region of the flavodoxin gene from Desulfovibrio vulgaris.

450 bp HindIII fragment, designated DVF450H3) was used as a probe in the identification and isolation of the D. salexigens flavodoxin (DSF) gene. The DVF450H3 probe was found to hybridize to two fragments of molecular sizes of 3.8 and 7.2-kb in Southern blots of ClaI- or NcoI-digested genomic DNA, respectively, as shown in Figure 1. A mini-library was prepared in E. coli XL1-Blue cells by transformation with size-selected (3.0 to 5.0-kb) ClaI restriction fragments cloned into the Bluescript vector. Of approximately 1600 recombinant clones, eight colonies were found to hybridize to the DVF450H3 probe and contain a plasmid with a 3.8-kb insert also hybridizing to the probe.

The region responsible for hybridization was further confined to a 1.0-kb PstI-ClaI fragment located at the extreme 3'-end of the insert of one of these

clones. Analysis of the nucleotide sequence of this fragment suggested that this clone was missing at least 12% of the 3'-end of the coding region for the DSF protein (based on homology to the DVF sequence). Knowing that the 5' to 3' orientation of the gene was from the PstI site to the ClaI site, genomic DNA was digested with PstI and various other restriction endonucleases to generate larger fragments more likely to contain the entire structural gene. A 1.1-kb PstI-EcoRI fragment and a 1.4-kb PstI-HindIII fragment were identified on Southern blots of these double-digests using the PstI-ClaI fragment as a probe. The 1.4-kb PstI-HindIII fragment was chosen for cloning. A mini-library of size-selected PstI-HindIII fragments was prepared and screened as described above. Of approximately 1200 recombinant clones, six colonies were observed to strongly bind the probe and to contain a plasmid with the 1.4-kb insert which hybridized to the 1.0-kb PstI-ClaI probe. One of these clones was chosen for nucleotide sequence analysis.

Nucleotide sequence analysis - The partial restriction map and sequencing strategy involved in the determination of the nucleotide sequence of the portion of the 1.4-kb PstI-HindIII fragment contained the entire DSF sequence shown in Figure 2. The nucleotide sequence is reported in Figure 3. Examination of this sequence revealed an open-reading frame beginning with the initiation codon ATG and coding for a polypeptide of 146 amino acid residues. The amino acid sequence derived from this open-reading frame, also shown in Figure 3, is highly homologous to the flavodoxin protein isolated from D. vulgaris, a protein consisting of 148 amino acid residues. A detailed comparison will follow.

The DSF and DVF genes share a sequence alignment of 278 nucleotides for a DNA sequence homology of 63%. The G + C content of this open-reading frame is 47%, a value significantly lower than that observed for the corresponding gene from D. vulgaris (62%). Examination of the 5'-flanking sequence revealed regions highly homologous to the consensus promoter sequences of E. coli (13). The sequence -TTCACAT- beginning at nucleotide -109 is very similar to the "-35 site" while the sequence -TAAATT- beginning at nucleotide -87 is homologous to the "-10 site" of E. coli promoters. The region bounded by these two sites is very A/T-rich and is of the optimal length for the separation of these two

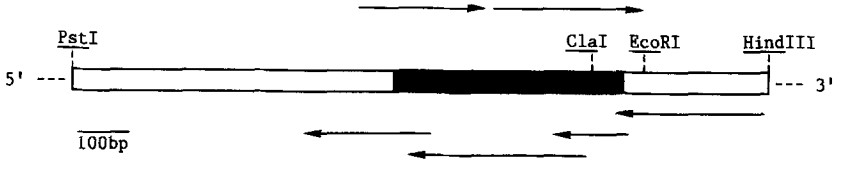

Figure 2. Partial restriction map and nucleotide sequencing scheme for the 1.4 kb PstI-HindIII restriction fragment. The darkened region represents the coding region for the Desulfovibrio salexigens flavodoxin. The arrows indicate the regions sequenced on each DNA strand.

```
AGACTAACAATAAAAATTTTGTTTTCTCACAAACAACATTTTGAACGCGCATATATCGGTACGTCTTAAGGAGTGCATA -160

ACACCACTCAGGCGCGGCGTAAACACTCAATTACCGATTGGCTGACATTTTCACATTTTCTTAATATTTTTATAAATTG  -80
                                                    -35                  -10

ATATTGACTTTGAATTTCACTCTCACTACAAACTAACACAACGTTATCAAACCGATTACAGATTTATAAGGAGGACACT   -1
                                                                          SD

ATG TCC AAA TCA CTG ATC GTT TAC GGC TCT ACT ACC GGA AAT ACT GAA ACA GCC GCC GAA   60
Met Ser Lys Ser Leu Ile Val Tyr Gly Ser Thr Thr Gly Asn Thr Glu Thr Ala Ala Glu
  1                                 10                                  20

TAC GTG GCT GAA GCG TTT GAA AAC AAA GAA ATT GAT GTG GAA CTT AAA AAT GTT ACT GAT  120
Tyr Val Ala Glu Ala Phe Glu Asn Lys Glu Ile Asp Val Glu Leu Lys Asn Val Thr Asp
                                  30                                  40

GTC AGT GTT GCC GAT CTC GGC AAC GGA TAC GAC ATC GTG CTA TTC GGC TGC TCT ACC TGG  180
Val Ser Val Ala Asp Leu Gly Asn Gly Tyr Asp Ile Val Leu Phe Gly Cys Ser Thr Trp
                                  50                                  60

GGC GAA GAA GAA ATT GAA TTG CAG GAC GAC TTC ATC CCC CTC TAC GAT TCC CTC GAA AAC  240
Gly Glu Glu Glu Ile Glu Leu Gln Asp Asp Phe Ile Pro Leu Tyr Asp Ser Leu Glu Asn
                                  70                                  80

GCA GAC CTG AAG GGC AAG AAA GTA TCT GTC TTC GGA TGC GGT GAC TCC GAT TAC ACT TAT  300
Ala Asp Leu Lys Gly Lys Lys Val Ser Val Phe Gly Cys Gly Asp Ser Asp Tyr Thr Tyr
                                  90                                 100

TTT TGC GGT GCA GTA GAT GCC ATC GAA GAA AAA CTC GAA AAA ATG GGG GCC GTC GTC ATA  360
Phe Cys Gly Ala Val Asp Ala Ile Glu Glu Lys Leu Glu Lys Met Gly Ala Val Val Ile
                                 110                                 120

GGT GAC AGC CTC AAG ATC GAT GGC GAC CCG GAA CGC GAT GAG ATT GTA AGC TGG GGT TCA  420
Gly Asp Ser Leu Lys Ile Asp Gly Asp Pro Glu Arg Asp Glu Ile Val Ser Trp Gly Ser
                                 130                                 140

GGA ATA GCG GAC AAA ATT TAA ATTAATACAGCTCAGCTATCCTTACGAAG
Gly Ile Ala Asp Lys Ile ***
```

Figure 3. Nucleotide sequence of the flavodoxin gene from Desulfovibrio salexigens. The sequence shown is the nontranscribed strand. The translated amino acid sequence for the flavodoxin protein is indicated below the open coding region of the gene. SD, potential Shine-Dalgarno sequence; "-35" and "-10", putative promoter sites.

sites. The sequence -AGGAGG- located just upstream to the initiation codon is nearly identical to the Shine-Dalgarno site observed in E. coli. This region is separated from the initiation codon by five nucleotides, a separation similar to that observed in other bacteria (14). Because of these strong similarities,

```
                10        20        30        40        50        60        70
                .         .         .         .         .         .         .
D.s. MSKSLIVYGSTTGNTETAAEYVAEAFENKEIDVELKNVTDVSVADLGNGYDIVLFGCSTWGEEEIELQDDFIPLY
     ¦:¦:¦¦¦¦¦¦¦¦¦¦¦¦¦ :¦¦ :¦ .:.: . :¦: ::.:.¦..::¦ :¦:¦:¦¦:¦¦¦¦¦¦::.¦¦¦¦¦¦¦¦¦:
D.v. MPKALIVYGSTTGNTEYTAETIARELADAGYEVDSRDAASVEAGGLFEGFDLVLLGCSTWGDDSIELQDDFIPLF
                                                                  # *   * *

          80        90        100       110       120       130       140
          .         .         .         .         .         .         .
D.s. DSLENADLKGKKVSVFGCGDSDYTYFCGAVDAIEEKLEKMGAVVIGDSLKIDGDP--ERDEIVSWGSGIADKI
     ¦¦¦¦::: :¦:¦¦: ¦¦¦¦¦¦.¦.¦¦¦¦¦¦¦¦¦¦¦¦¦.::¦¦ :: ¦:¦:¦¦¦¦  .¦¦:¦¦:¦: :: : ¦
D.v. DSLEETGAQGRKVACFGCGDSSYEYFCGAVDAIEEKLKNLGAEIVQDGLRIDGDPRAARDDIVGWAHDVRGAI
                  * #                                 *
```

Figure 4. Comparison of the flavodoxin protein sequences from D. vularis (D.v.) and D. salexigens (D.s.) using the PAM matrix of Lipmann and Pearson (ref 12). Identical (¦), conservative (:), and semiconservative (.) homologies are shown. Underscored regions form portions of the FMN binding site in the D. vulgaris flavodoxin. #, aromatic residues flanking the FMN isoalloxazine ring; *, acidic residues possibly involved in complex formation with cytochrome c_3 (ref 16).

expression of this protein in E. coli using the endogenous D. salexigens promoter might be anticipated.

Sequence comparisons - The protein sequence for DSF was compared to that of DVF. As shown in Figure 4, the two sequences can be directly aligned to reveal several regions of sequence identity. A somewhat better alignment was observed near the carboxyl terminus if a double deletion was introduced after residue 130. By analogy to the DVF crystal structure, this "deletion" would eliminate a short bend on the surface of the protein which immediately precedes the final α-helix. With this adjustment, the two proteins share a sequence identity of 55%. Of the remaining sequence, 26% and 8% were found to be conservatively and semiconservatively homologous, respectively. The molecular weight for the protein derived from this open reading frame is 15,814 daltons, consistent with the reported value for the purified protein. Also, the amino acid composition derived from the gene sequence is consistent with published data (1).

When compared to the X-ray crystal structure of the DVF, the strongly homologous regions were largely confined to those portions of the protein which are in the immediate vicinity of the flavin mononucleotide cofactor binding site. All known flavodoxin structures share a highly conserved sequence of predominantly hydroxyamino acids near their amino terminus which forms a loop which surrounds and hydrogen bonds to the terminal phosphate group of the cofactor (residues 10-15 in D. vulgaris) (3). The DSF sequence is identical to DVF in this region. It is likely, then, that similar types of hydrogen bonding interactions between these residues and oxygen atoms of the phosphate group as well as portions of the ribityl side chain are conserved in the DSF (2).

The most striking feature of the binding site for the isoalloxazine ring of the FMN is the presence of the aromatic residues typtophan-60 and tyrosine-98 which lie on the inner and outer face of the flavin ring system, respectively (2). These two residues are conserved in DSF. This general theme is repeated in other flavodoxins (e.g., Anacystis nidulans (4)), but is by no means universal (e.g., Clostridum MP (3)). Their conservation in many flavodoxins implies an important functional role. Indeed, our studies involving the substitution of these residues in DVF by in vitro mutagenesis suggests that the redox properties of the bound FMN are very dependent on the chemical nature of the amino acid side chains at these positions (15).

The portions of the polypeptide that flank these aromatic aromatic amino acid residues in DSF and DVF are highly homologous but nonidentical, especially in their immediate vicinity. Residues 61-64 form a loop which is positioned adjacent to the C(4)-N(5)-C(6) edge of the flavin. Glycine-61 is conserved in DSF. This is noteworthy in that the amide carbonyl of the homologous glycine residue in Clostridium MP, which points away from the flavin in the oxidized form of the protein, apparently rotates upon reduction of the flavodoxin to the

semiquinone to establish a new hydrogen bond with the hydrogen atom on N(5) of the flavin semiquinone. This interaction is thought to be at least partially responsible for the stabilization of the blue neutral form of the flavin radical (3). This glycine is followed by three glutamate residues in DSF, creating a more negative surface potential at this location than that provided by the two aspartate and the serine residues observed in DVF. With the exception of the transposition of conservatively homologous acidic and hydroxyamino acid residues flanking tyrosine-98, the so-called 90's loop is identically conserved. It is the peptide backbone heteroatoms of residues 95 and 100-102 in DVF that are involved in several hydrogen bonding interactions with the pyrimidine subnucleus of the isoalloxazine ring (2). The high degree of sequence identity in the FMN binding site in DSF implies similar types of interactions with the cofactor.

Flavodoxins, as a group, are quite acidic proteins. The calculated net charges for DSF and DVF at pH 7 are similar, with DSF being slightly more acidic. Also, the general asymmetric distribution of charged residues noted in

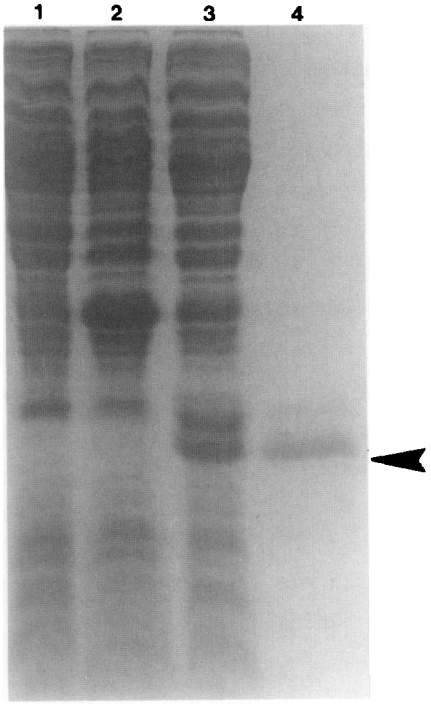

Figure 5. Sodium dodecyl sulfate-polyacrylamide gel electrophoretic analysis of the expression of recombinant *D. salexigens* flavodoxin in transformed *E. coli*. Electrophoresis was performed on 16% (w/v) polyacrylamide gels as described previously (ref 5). **Lane 1**, extract of untransformed cells; **Lane 2**, cells transformed with the Bluescript plasmid; **Lane 3**, cells transformed with the plasmid with the 1.4-kb *Pst*I-*Hind*III insert containing the entire coding region and 5'-flanking sequence for the *D. salexigens* flavodoxin; **Lane 4**, purified *D. salexigens* flavodoxin (Note: The faint trailing band is an apparent artifact as it also appears after re-electrophoresis of the excised lower band).

DVF is also conserved. Molecular modeling studies have suggested that several surface acidic residues may be important in the formation and stabilization of the postulated complex that forms between the flavodoxin and certain cytochrome proteins during electron transfer (16). In these models, these residues form ion pairs with corresponding basic residues on the surface of the cytochrome. The sequence results presented here are consistent with that model. Four of the acidic residues involved are identically conserved (Asp-69, -95, and -129; and Glu-66). The fifth residue, Asp-62, appears as a glutamate in DSF. These residues are located within homologous regions suggesting that the entire surface that may be involved in complex formation with donor/acceptor proteins is maintained in these flavodoxins.

Gene expression - As described above, a strong homology between regions of the 5'-flanking sequence of the D. salexigens gene sequence and the consensus sequences for promoters in E. coli was noted. These similarities raised the possibility that expression of the cloned flavodoxin in E. coli might occur using its endogenous promoter. Examination of E. coli extracts by SDS-polyacrylamide gel electrophoresis did reveal the presence of an intense protein band in cells transformed with plasmids containing the gene and 5'-flanking regions (Figure 5, Lane 3) that is absent in extracts of untransformed cells or cells transformed with the vector lacking the insert (Figure 5, Lanes 1 and 2). This additional protein band comigrates with the purified flavodoxin (Lane 4). Expression levels were estimated to be at approximately 4-6% of total extractable proteins, a remarkably high level from an exogenous promoter but, perhaps, consistent with the promoter homology and the high copy number for Bluescript plasmids in E. coli. This protein was isolated from E. coli extracts and found to have molecular properties consistent with that reported for the wildtype DSF. These results contrast sharply with those obtained for the wildtype gene from D. vulgaris where no detectable expression was observed in E. coli using the endogenous promoter region (5). It should be noted, however, that even within strong expression vectors the D. vulgaris gene was efficiently expressed only after modification of the 5'-end of the coding region (5).

ACKNOWLEDGMENT: This research was supported in part by a grant from the National Institutes of Health (GM36490).

REFERENCES

1. Moura, I., Moura, J.J.G., Bruschi, M., and LeGall, J. (1980) Biochim. Biophys. Acta, 591, 1-8.
2. Watenpaugh, K.D., Sieker, L.C., and Jensen, L.H. (1973) Proc. Natl. Acad Sci. USA 70, 3857-3860.
3. Mayhew, S.G. and Ludwig, M.L. (1975) in The Enzymes (Boyer, P.D., ed.), 3rd edn., Vol 12, pp 57-118.
4. Smith, W.W., Pattridge, K.A., Ludwig, K.L., Petsko, G.A., Tsernoglou, D., Tanaka, N., Yasunobu, K.T. (1983) J. Mol. Biol. 165, 737-755.

5. Krey, G.D., Vanin, E.F., and Swenson, R.P (1988) J. Biol. Chem 263, 15436-15443.
6. Postgate, J.R. (1984) The Sulfate-Reducing Bacteria, 2nd ed., p.32, Cambridge University Press, Cambridge, England.
7. Feinberg, A.P., and Vogelstein, B. (1984) Anal. Biochem. 137, 266-267.
8. Grundström, T., Zenke, W.M., Wintzerith, M., Matthes, H.W.D., Staub, A., and Chambon, P. (1985) Nucleic Acids Res. 13, 3305-3316.
9. Hanahan, D. (1983) J. Mol. Biol. 166, 557.
10. Sanger, F., Nicklen, S., and Coulson, A.R. (1977) Prod. Natl. Acad. Sci. U.S.A. 74, 5463-5467.
11. Maniatis, T., Fritsch, E.F., and Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
12. Lipman, D.J., and Pearson, W.R. (1985) Science 227, 1435-1441.
13. Hawley, D.K., and McClure, W.R. (1983) Nucleic Acids Res. 11, 2237-2255.
14. Stormo G.D., Schneider, T.D., and Gold, L. (1982) Nucleic Acids Res. 10, 2971-2975.
15. Swenson, R.P., Eren, M., and Krey, G.D. (1988) J. Cell Biol 107, 621a.
16. Stewart, D.E., LeGall, J., Moura, I., Moura, J.J.G., Peck, Jr., H.D., Xavier, A.V., Weiner, P.K., and Wampler, J.E. (1988) Biochemistry 27, 2444-2450.