The Nucleotide Sequence of the *serA* Gene of *Escherichia coli* and the Amino Acid Sequence of the Encoded Protein, d-3-Phosphoglycerate Dehydrogenase*

(Karen L. Tobey and Gregory A. Grant)

From the Department of Biological Chemistry and the Division of Dermatology, Department of Medicine, Washington University School of Medicine, St. Louis, Missouri 63110

The nucleotide sequence of *serA*, the structural gene of *Escherichia coli* which codes for d-3-phosphoglycerate dehydrogenase, has been determined. The structural gene contains 1233 nucleotides which code for the 409 amino acids of the subunit of the tetrameric enzyme, as well as the initiator methionine, which is cleaved from the mature protein, and the termination codon. The majority of the primary structure of the enzyme has been confirmed by automated Edman degradation of peptide fragments produced by a variety of cleavage agents. Comparison of the amino acid sequence of phosphoglycerate dehydrogenase with other NAD-dependent oxidoreductases reveals less than 20% homology, although conservation of certain specific residues in the coenzyme binding domain appears to be evident.

*serA* gene, as well as the nearly complete amino acid sequence of the corresponding phosphoglycerate dehydrogenase.

**EXPERIMENTAL PROCEDURES**

DNA Cloning, Mapping, and Sequencing—The cloning vector pBR325, K12 genomic DNA, and the *E. coli* serine auxotroph χ536 (HfrOR11 glu V42 : *SerAl3 T3*) were kindly supplied by Dr. Roy Curtiss III of the Department of Biology. The Klenow fragment of DNA polymerase was either from Bethesda Research Laboratories or a gift from Dr. John Majors of the Department of Biological Chemistry. T4 polynucleotide kinase, all restriction endonucleases, and the sequencing vectors M13 mp18 and mp19, were from New England Biolabs. T4 DNA ligase was purchased from Sigma, and the M13 E. coli host strain JM 105 was from Amersham. All synthetic oligonucleotides used for probes and sequencing primers were made on an Applied Biosystems Model 380A DNA synthesizer and were all 17 bases (17-mer) in length.

Plasmid isolation was accomplished by a modification of the preparative methods of Birnboim (21). T3-CDTA* was used in place of acetate MOPS buffer, and only a single ethanol precipitation was performed. Step 5a was deleted from the procedure, and plasmid was simply incubated with ribonuclease A for 60 min at room temperature and then precipitated with ethanol. In those cases where it appeared that endonuclease contamination was a problem, the plasmid preparation was treated with 50 μg/ml Proteinase K at 65 °C for 60 min in 10 mM Tris, pH 7.5, 5 mM EDTA, and 0.5% SDS. The protein was then extracted with phenol, and the plasmid was ethanol-precipitated.

All endonuclease digestions were performed in the optimal salt buffers as described by Maniatis et al. (22) except for EcoRI digestion which was performed in 100 mM Tris, pH 7.5, 50 mM NaCl, and 5 mM MgCl2. All ligations were performed at 16 °C as described by New England Biolabs. All plasmid transformations were carried out as described by Dagert and Ehrlich (23), and all M13 transformations were accomplished as described by Amersham (24). Southern blot analysis of restriction fragments was performed as described by Maniatis et al. (22) with a 3P-labeled synthetic oligonucleotide based on amino acid sequence data. Sanger dyeoxy sequencing of restriction fragments subcloned into M13 was performed as described (24).

Protein Fragmentation, Isolation, and Sequencing—d-3-Phosphoglycerate dehydrogenase was isolated from cultures of *E. coli* strain B as previously described (8) and was alkylated with iodo[14C]acetate acid after reduction with dithiothreitol (25). The S-14Ccarboxymethyl-d-3-phosphoglycerate dehydrogenase was subjected to cyanogen bromide digestion (25) after succinylation (26), to trypsin digestion (27) at the arginyl residues after citraconylation (28), and to digestion by endoproteinase Lys-C, chymotrypsin, elastase, and *Staphyloccoccus aureus* V8 protease. The conditions used for all of the proteolytic digestions were essentially the same as those described for trypsin except that, in the case of elastase, incubations were performed for 4 h rather than overnight. Peptides produced by cyanogen bromide digestion of carboxymethylated and succinylated d-3-phosphoglycerate dehydrogenase and by trypsin, carboxymethylated and citraconylated d-3-phosphoglycerate dehydrogenase were initially prepared by gel filtration on Sephadex G-100 and G-
RESULTS AND DISCUSSION

DNA Sequence of the serA Gene—*E. coli* K12 genomic DNA was partially digested by serial dilution with endonuclease *Sau3A*I and fragments of 4–6 kb were selected for introduction into the unique *BamH*I site in the tetracycline resistance gene of pBR325. The resulting plasmid selected, pGT17, contains a functional *serA* gene as judged by its ability to transform a *serA* mutant to wild-type.

Fig. 1 shows the restriction map of the genomic insert containing the *serA* gene and the sequencing strategy employed. Both strands of the *serA* gene were sequenced using the Sanger dideoxy method. The two *BamH*I/*HindIII* restriction fragments were overlapped by sequencing through the interior *BamH*I site with the *HindIII*/*PstI* fragment as indicated. Initially, Southern blot analysis of the restriction fragments indicated that the *serA* gene was located in the genomic insert proximal to the unique *HindIII* site in pBR325. Therefore, sequencing was initiated by priming the 1.3-kb *HindIII*/*BamH*I restriction fragment subcloned into M13 mp18 with a synthetic oligonucleotide complementary to the pBR325 sequence just prior to the insert site. The complete restriction fragment was sequenced by synthesizing new oligonucleotides based on the furthest sequence determined and using them for subsequent priming of the template in separate sequencing reactions. Sequencing of the opposite strand, cloned into M13 mp19, was started with M13 universal primer and sequenced in the same manner. The *HindIII*/*PstI* fragment cloned into M13 mp19 was primed with universal primer, and the opposite strand placed into M13 mp18 was primed initially with an oligonucleotide based on the sequence determined at the 3′ end of the 1.3-kb *HindIII*/*BamH*I fragment. The sequence was completed in a similar manner using the 3.2-kb *BamH*I/*HindIII* fragments placed in both M13 mp18 and mp19 and primed with oligonucleotide complementary to the previously determined sequence. The complete sequence is shown along with the protein translated sequence in Fig. 2.

Protein Sequence of the 3-Phosphoglycerate Dehydrogenase—

\[S-\text{[14C]}\text{Carboxymethyl-D-3-phosphoglycerate dehydrogenase}\]

was digested with cyanogen bromide after succinylation (CN), trypsin after citraconylation (CN), endoproteinase Lys-C (L), chymotrypsin (C), elastase (E), and *S. aureus* strain V8 protease (S) to produce peptides for amino acid sequence analysis. S-Carboxymethyl-D-3-phosphoglycerate dehydrogenase is insoluble in dilute aqueous buffers at both acidic and basic pH, but the protein becomes soluble after succinylation or citraconylation. However, since S-carboxymethyl-D-3-phosphoglycerate dehydrogenase is insoluble under the conditions used for digestion with chymotrypsin, V8 protease, elastase, and endoproteinase Lys-C, these digestions were performed on the suspended precipitate. In all cases, the digest cleared soon after addition of the protease. Initially, digestion of S-carboxymethyl-D-3-phosphoglycerate dehydrogenase with cyanogen bromide yielded mostly insoluble peptides. After succinylation, the solubility of the peptides improved, but it was still not possible to isolate more than a few relatively small peptides. The pools obtained from Sephadex G-75 chromatography of the trypsin-digested citraconylated D-3-phosphoglycerate dehydrogenase were decitraconylated by incubation in 4.5% formic acid at 37°C for 3 h. This procedure produced some acid-insoluble peptides, but they all readily dissolved in 0.1 M ammonium bicarbonate, pH 8.0.

Of the many peptides produced by the numerous digestions employed, 90 peptides were isolated that produced reliable sequence information. However, only 29 of these produced sequences that provided unique information. These peptides are shown in Fig. 3 along with their positions in the amino acid sequence. Inspection of Fig. 3 demonstrates that it was not possible to deduce the complete sequence of D-3-phosphoglycerate dehydrogenase based on the peptide data alone. Six gaps, ranging in size from 5–16 residues, as well as six 1–2-residue overlaps, remained after all of the peptides were placed. In some cases, the gaps, which account for 66 residues, represent the inability to sequence to the end of a particular peptide, while in other cases it is obvious that peptides representing that region of the sequence were simply not found. The amino acid sequence through these gap regions, which correspond to residues 93–97, 109–124, 167–177, 192–203, 335–342, and 351–364 had to be deduced from the DNA sequence. In addition, the overlap regions at residues 75–76, 216–219, 224–225, 286–257, 269–370, and 374–377 were also supplied by the DNA sequence. Overall, the unique sequence determined from the peptide data represents 342 residues of the 409 total residues.

It should also be noted that two amino acid substitutions between *E. coli* strain K12, which was used for DNA sequencing, and strain B, which was used for protein isolation, were also revealed. These are an Ile in K12 for a Phe at position 78 and a Glu in K12 for an Asp at position 147. Both substitutions are conservative in nature and there does not appear to be any appreciable difference in enzymatic activity from the two strains as a result of these differences.

Sequence Comparison—A search of the National Biomedical Research Foundation (NBRF) protein sequence data bank with the FASTP algorithm of Lipman and Pearson (29) failed to reveal any significant overall similarity between D-3-phosphoglycerate dehydrogenase and any other protein. Similarly, direct comparison of the D-3-phosphoglycerate dehydrogenase sequence with that of other dehydrogenases using the NBRF ALIGN program yielded little, if any, significant homology. For instance, comparison to lactate dehydrogenase and mitochondrial malate dehydrogenase showed less than 20% identity with D-3-phosphoglycerate dehydrogenase. This is not an unexpected observation since, historically, very little sequence homology among different dehydrogenases has been...
found (30). Nevertheless, among the nicotinamide adenine dinucleotide-dependent oxidoreductases for which crystal structures are available, some striking structure-function relationships have been observed (30). These dehydrogenases can be divided into roughly two domains, the nucleotide binding domain and the catalytic domain. The nucleotide binding domain, which is comprised of roughly the first half of the enzyme, displays a high degree of conservation of tertiary structure, and it appears that coenzyme binding takes place in nearly identical fashion in each protein. Generally, it consists of a four-stranded parallel β sheet and one α helix with virtually identical arrangement. When the coenzyme-enzyme interactions are considered, equivalent structural elements among the different oxidoreductases appear to make similar contacts with equivalent parts of the bound nucleotide.

Inspection of the amino acid sequences of the oxidoreductases in these areas revealed a consistent conservation of residue type in essentially two areas (30). With this in mind, a more detailed inspection of D-3-phosphoglycerate dehydrogenase, also a 2-hydroxy acid dehydrogenase, revealed similar elements of primary structure. The first of these is in the structure made up of the PA sheet and αB helix and consists of the generally conserved hexapeptide sequence G-X-(G or A)-X-X-G. Some representative sequences in this area, as well as that of D-3-phosphoglycerate dehydrogenase, are shown in Fig. 4, residues 17-22. These sequences show that,
while some variation can occur in the conserved positions, they are generally occupied by small side chain residues. Also, note the conservation of a basic residue at position 11 and a hydrophilic residue, usually valine, at position 15. The first glycine of the hexapeptide comes into close proximity with the adenine ribose of the bound coenzyme. It appears that any other residue, that is one with a side chain, would affect the position of the coenzyme. The glycine at position 6 (residue 22 in Fig. 4) of the hexapeptide is not in direct contact with the coenzyme, but any large side chain here would probably have some effect on the sharp turn between $\beta A$ and $\alpha B$. This residue is highly conserved, although it is an alanine in cytoplasmic malate dehydrogenase. The 3rd residue of the hexapeptide is also involved in the turn and, although it is generally conserved, somewhat more diversity appears possible here. The histidine present in this position in the D-3-phosphoglycerate dehydrogenase sequence is the bulkiest side chain yet observed in this position, but it is not yet known if the existence of this pattern of conservation in D-3-phosphoglycerate dehydrogenase also complicates homology studies. Again, it is not yet possible to distinguish what portions of the polypeptide are analogous to the backbone of the other 2-hydroxy acid dehydrogenases and which may represent areas of deviation from the smaller enzymes if the higher order structural homology within this family extends to D-3-phosphoglycerate dehydrogenase. Finally, the primary structure does not reveal any information pertaining to the mechanism of allosteric control, but it is interesting to speculate whether, as discussed above, the possible contribution of a non-homologous aspartyl residue to the putative His-Asp pair may be involved. X-ray crystallographic analysis and oligonucleotide directed mutagenesis studies designed to address questions of the structure-function relationship are currently underway.

Acknowledgments—We thank Maria Zapp, James Sacchettini, Mark Frazier, and Michael Burg for excellent technical assistance throughout the course of this investigation. We would also like to thank Dr. Roy Curtis III and Dr. William Jacobs for their invaluable technical assistance in the early stages of this work and Dr. Luis Glaser and Dr. Arthur Eisen for their continued support and encouragement. The oligonucleotide primers used in this study were synthesized in the Department of Biological Chemistry’s Protein Chemistry Laboratory.

REFERENCES

Translated Sequence of Phosphoglycerate Dehydrogenase