Complete amino acid sequence of glucose dehydrogenase from *Bacillus megaterium*

Klaus-Dieter Jany, Wolfgang Ulmer, Marion Fröschle and Gerhard Pfleiderer

Institut für Organische Chemie, Biochemie und Isotopenforschung der Universität Stuttgart, Pfaffenwaldring 55, D-7000 Stuttgart 80, FRG

Received 3 November 1983

The amino acid sequence of glucose dehydrogenase from *Bacillus megaterium* has been determined. The enzyme consists of 4 identical subunits, each containing 262 amino acid residues. Its structure was established using manual Edman degradation procedures after modification of the enzyme in the native form with reagents specific to the amino acids histidine, tyrosine, tryptophan and lysine in order to identify residues involved in catalysis or located in the subunit binding area.

Glucose dehydrogenase *Dehydrogenase* *Modification* *Amino acid sequence*
Bacillus megaterium

1. INTRODUCTION

Glucose dehydrogenase is a key enzyme during the early stage of sporogenesis of *Bacillus megaterium*. The enzyme catalyses the oxidation of D-glucose without prior phosphorylation to D-β-gluconolactone using NAD or NADP as a coenzyme [1]. The enzyme is a tetrameric protein (M_r 118000) displaying the unusual property of dissociating into its inactive monomers under very mild conditions and reassociating to give complete reactivation [2].

Chemical modifications of the enzyme in the tetrameric and/or monomeric state followed by enzymatic investigations and sequence analysis were carried out as an approach toward the identification of residues involved in catalysis or located in the subunit binding area.

Here we present the complete primary structure of the subunit of glucose dehydrogenase and the amino acid residues susceptible to the modification.

2. MATERIALS AND METHODS

Glucose dehydrogenase was a generous and much appreciated gift from E. Merck, Darmstadt. The enzyme was further purified to an activity of 440 U/mg and assayed as in [1].

The following modifications were performed:

(i) Tyrosine modification with tetranitromethane (Serva):
(a) Nitration of the enzyme in the tetrameric state was carried out in 0.2 M Tris–HCl (pH 8.7) containing 1 M NaCl with a 1.5 molar excess of reagent with respect to tyrosine for 15 min.
(b) For modification in the monomeric state, the enzyme was dissociated in 50 mM Tris–glycine (pH 8.7) and after 25 min the monomers were treated with the same excess of reagent as above for 15 min. In both cases the reaction was stopped by cysteine.

(ii) Histidine modification with 3-bromoacetylpyridine [3]:
The enzyme (tetrameric state) in 0.1 M Na-phosphate (pH 7.8) containing 1 M NaCl was inactivated with a 3.1 molar excess of 3-bromo[*carbonyl*-^{14}C]acetylpyridine (a gift from Professor Woenckhaus, Frankfurt) for 22 min. The carbonyl group was subsequently reduced by sodium borohydride at 0°C [4].

Published by Elsevier Science Publishers B.V.

(iii) Tryptophan modification with dimethyl-(2-hydroxy-5-nitrobenzyl)sulfonium bromide (Koshland-reagent I) [5]:
The tetramer was dissociated into monomers as described above. For modification of tryptophan residues, the pH was rapidly adjusted to 6.5 and a 100-fold molar excess of reagent with respect to the Trp residues was added. The reaction was stopped by cysteine after 10 min.

(iv) Lysine modification with methyl acetimidate and dansyl acetimidate:
The synthesis of the free base of dansyl acetimidate and the selective modification of the enzyme are described in [21].

Sequence analyses were performed on peptides resulting from the following cleavages (fig.2):

(i) The enzyme nitrated in the tetrameric state (1.2 NO_2–Tyr residues) was cleaved with cyanogen bromide in 70% formic acid [6]. CNBr:methionine, 60:1; 22 h, room temperature.

(ii) The enzyme nitrated in the monomeric state (2.3 NO_2–Tyr residues) was cleaved with BNPS-skatole in 60% acetic acid [7]. BNPS-skatole:tryptophan, 35:1; 48 h, room temperature.

(iii) The enzyme modified with 3-bromoacetylpyridine (1.3 mol modifier incorporated/mol subunit) was hydrolysed at the Asp–Pro bond in 70% formic acid for 90 h at 37°C.

(iv) The 3-bromoacetylpyridine inactivated enzyme was also digested with chymotrypsin (Worthington) in 0.1 M NH_4HCO_3 (pH 7.8) at a substrate:protease ratio of 100:1 (w/w) for 40 min at 25°C.

(v) The enzyme modified at the lysine residues (1.3 dansyl groups incorporated) was digested with special purified trypsin [8] in 50 mM NH_4HCO_3 (pH 8.4) at a substrate:protease ratio of 50:1 (w/w) for 5 h at 25°C after blocking the remaining lysines with citraconic anhydride [9].

(vi) The enzyme with modified tryptophan residues was digested with *Staphylococcus aureus* V8 protease (Miles) in 50 mM NH_4HCO_3 (pH 7.8). Protein:protease, 30:1 (w/w); 20 h, 25°C.

(vii) The unmodified, but citraconylated enzyme was digested with *Astacus fluviatilis* protease [10] kindly donated by Professor Zwilling, Heidelberg in 0.1 M Tris–HCl (pH 8.2). Substrate:protease, 50:1 (w/w); 4 h, 25°C.

Limited proteolysis of the enzyme was carried out with proteinase K in 0.2 M Tris–HCl (pH 7.8) containing 1 M NaCl. Substrate:proteinase K, 60:1 (w/w); 90 min, 25°C [10,11].

Large sized peptides (>25 residues) were subfragmented. In addition to the mentioned cleavage procedures, proteinase K, thermolysin and Lys-C proteinase (Boehringer) were used. The fragments were isolated by a combination of gel filtration, ion exchange chromatography and thin layer chromatography [12].

Amino acid analyses were performed on a Biotronik LC 6000 analyser after hydrolysis of the peptides in 6 N HCl at 108°C. The presence of tryptophan was checked by the Ehrlich reaction [13] and its content was determined after hydrolysis in methanesulfonic acid [14].

All sequential degradations were performed manually using the dansyl-Edman procedure [12,15] and the 4-(dimethylamino)azobenzene-4-isothiocyanate/phenylisothiocyanate double coupling method [16]. Most of the peptides were sequenced by both methods. C-terminal sequences were determined by kinetic analysis using carboxypeptidases A, B, or Y.

3. RESULTS AND DISCUSSION

Our work on the determination of the amino acid sequence of glucose dehydrogenase from *B. megaterium* was initiated as the result of physicochemical investigations on the reversible dissociation-association process of its subunits [2]. In contrast to the most common strategy, we have modified amino acid side chains believed to be involved in catalysis or in binding of the subunits and characterized the properties of the modified enzyme species prior to fragmentation. To each modification another primary cleavage was performed. This procedure permitted not only the elucidation of the structure but also the identification of essential residues in one process.

The amino acid sequence of the subunit of glucose dehydrogenase is shown in fig.1. Its structure was derived from the analysis of 8 independent sets of peptide fragments as presented in fig.2. Each peptide is overlapped by at least 3 residues. The amino acid composition is in good agreement with the results derived from the sequence, except for valine and tryptophan. One more tryptophan residue was found by the sequence studies. The lower recovery of valine in the amino acid analysis may be explained by the two Val–Val–Val sequences [9–11, 34–36].

A detailed description of the data used to establish this structure together with the description of the effects upon the different modifications on the enzyme will be reported subsequently.

Glucose dehydrogenase consists of 4 identical subunits. The subunit chain has 261 amino acid residues and M_r 28352. Although the acid amino acids clearly dominate in the protein, there is no distinct acidic region in the molecule. The acid stretch 42–52 is flanked by a cluster of basic amino acids (Arg-39, Lys-41, Lys-55, -56). The different amino acids are rather evenly distributed throughout the molecule. No stretches of internal sequence homology are observed, but some repetitive or similar amino acid combinations are obvious (table 1). Microheterogeneities were found in the positions 81 (Ile⟶Glu) and 200 (Glu⟶Pro) as proved by the recovery of minor amounts of homogeneous peptide fractions during the different cleavages. Other heterogeneities may also exist, but we have not found any peptides, which could not unambiguously be aligned with the sequence. Also, within the isolated peptides there are no indications of heterogeneity.

A selective modification of the enzyme with imidoesters was achieved by reacting lysines not essential for activity with methyl acetimidate in the presence of the coenzyme NAD. After removal of the coenzyme, an essential lysine can be modified with dansyl acetimidate (Bozler, unpublished). The residue predominantly labelled by the dansyl group was identified as Lys-201.

```
                                10                  20                  30
Met-Tyr-Lys-Asp-Leu-Glu-Gly-Lys-Val-Val-Val-Ile-Thr-Gly-Ser-Ser-Thr-Gly-Leu-Gly-Lys-Ser-Met-Ala-Ile-Arg-Phe-Ala-Thr-Glu-

                                40                  50                  60
Lys-Ala-Lys-Val-Val-Val-Asn-Tyr-Arg-Ser-Lys-Glu-Asp-Glu-Ala-Asn-Ser-Val-Leu-Glu-Glu-Glu-Ile-Lys-Lys-Val-Gly-Gly-Glu-Ala-

                                70                  80                  90
Ile-Ala-Val-Lys-Gly-Asp-Val-Thr-Val-Glu-Ser-Asp-Val-Ile-Asn-Leu-Val-Gln-Ser-Ala-Ile-Lys-Glu-Phe-Gly-Lys-Leu-Asp-Val-Met-

                               100                 110                 120
Ile-Asn-Asn-Ala-Gly-Met-Glu-Asn-Pro-Val-Ser-Ser-His-Glu-Met-Ser-Leu-Ser-Asp-Trp-Asn-Lys-Val-Ile-Asp-Thr-Asn-Leu-Thr-Gly-

                               130                 140                 150
Ala-Phe-Leu-Gly-Ser-Arg-Glu-Ala-Ile-Lys-Tyr-Phe-Val-Glu-Asn-Asp-Ile-Lys-Gly-Thr-Val-Ile-Asn-Met-Ser-Ser-Val-(His)-Glu-Trp-

                               160                 170                 180
Lys-Ile-Pro-[Trp]-Pro-Leu-Phe-Val-His-[Tyr]-Ala-Ala-Ser-Lys-Gly-Gly-Met-Lys-Leu-Met-Thr-Glu-Thr-Leu-Ala-Leu-Glu-Tyr-Ala-Pro-

                               190                 200                 210
Lys-Gly-Ile-Arg-Val-Asn-Asn-Ile-Gly-Pro-Gly-Ala-Ile-Asn-Thr-Pro-Ile-Asn-Ala-Glu-(Lys)-Phe-Ala-Asp-Pro-Glu-Glu-Arg-Ala-Asp-

                               220                 230                 240
Val-Glu-Ser-Met-Ile-Pro-Met-Gly-Tyr-Ile-Gly-Glu-Pro-Glu-Glu-Ile-Ala-Ala-Val-Ala-Trp-Leu-Ala-Ser-Ser-Glu-Ala-Ser-Tyr-Val-

                               250                 260
Thr-Gly-Ile-Thr-Leu-Phe-Ala-Asp-Gly-Gly-Met-Thr-Gln-(Tyr)-Pro-Ser-Phe-Glu-Ala-Gly-Arg-Gly
```

Fig.1. Amino acid sequence of glucose dehydrogenase from *B. megaterium*. (○) Residues that are part of the active site; (□) residues located in the subunit binding area.

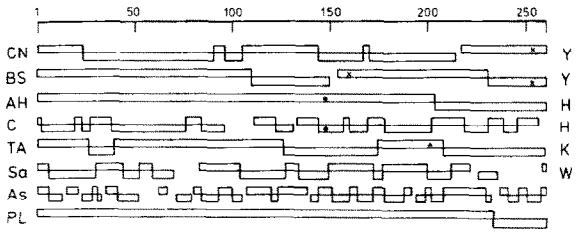

Fig.2. Diagram showing the main fragments generated by cyanogen bromide (CN), BNPS-skatole (BS), partial acid hydrolysis (AH), chymotrypsin (C), trypsin after blocking lysine residues (TA), *S. aureus* protease (Sa), *A. fluviatilis* protease (As) and limited proteolysis by proteinase K (PL). Fragments obtained from incomplete cleavages are not shown. The length of each bar is proportional to the size of that peptide. Residues (Y, H, K, W) modified prior to the fragmentation.

Table 1

Repetitive sequences in glucose dehydrogenase

Residues	Positions
-Ser-Lys-Gly-Gly-Met-Lys-Leu-	165-167
-Ala-Asp-Gly-Gly-Met-Thr-Gln-	249-251
-Gln-Ser-Ala-Ile-Lys-Glu-Phe-	80-83
-Arg-Glu-Ala-Ile-Lys-Tyr-Phe-	128-130
-Glu-Asn-Asp-Ile-Lys-Gly-Thr-	137,138
-Glu-Glu-Glu-Ile-Lys-Lys-Val-	53,54; 51-53
-Pro-Glu-Glu-Ile-Ala-Ala-Val-	224-226
-Ser-Asp-Val-Ile-Asn-Leu-Val-	73-75
-Gly-Thr-Val-Ile-Asn-Met-Ser-	141-143
-Ala-Ile-Arg-Phe-Ala-Thr-Glu-	26-28
-Ala-Glu-Lys-Phe-Ala-Asp-Pro-	201-203
-Ile-Thr-Leu-Phe-Ala-Asp-Gly-	246-248
-Gly-Lys-Val-Val-Val-Ile-Thr-	8-11
-Ala-Lys-Val-Val-Val-Asn-Tyr-	33-36
-Val-Ser-Ser-His-Glu-Met-Ser-	101-104
-Met-Ser-Ser-Val-His-Glu-Trp-	145-149

Reaction of the enzyme with the pseudo coenzyme analogon 3-bromoacetylpyridine results in a rapid loss of activity. The inactivation can be prevented by NAD, NMN, but not by ATP, AMP or glucose. The inactivation is correlated with the modification of one histidine residue which was subsequently identified as His-148.

Modification of the enzyme in the tetrameric state with tetranitromethane causes complete inactivation. NAD, AMP, ATP, but not ADP or substrate protect the enzyme from inactivation.

The inactivation is due to the modification of a single tyrosine which was later identified as Tyr-254. The enzyme remains in the tetrameric state upon the modifications and these modifications have no influence on the overall structure as indicated by a comparison of the circular dichroism (CD) patterns with those of the native enzyme. Further the enzyme has not lost its dissociation–association ability. Therefore we conclude that the modifications are occurring at the active site and His-148, Lys-201 and Tyr-254 are either involved in catalysis or binding of ligands.

Upon modification of the monomers with tetranitromethane two tyrosine residues react with the reagent and the modified subunits are unable to reassociate to the tetramer. As those modified residues Tyr-254 and Tyr-160 were identified. As indicated above, Tyr-254 is part of the active site, whereas Tyr-160 is not at all susceptible to a modification in the tetramer. Obviously Tyr-160 is protected by subunit contacts from the modifying reagent. Thus Tyr-160 must be located in the subunit binding area.

During the dissociation of the enzyme one tryptophan residue is exposed to the solvent as demonstrated by CD techniques ([2] and Maurer, personal communication) and we tried to locate this residue by modification of the enzyme in the monomeric state with Koshland-reagent I. However, by this procedure 2.6 residues react with the reagent and we were unable to decide whether Trp-150 or Trp-154 are parts of the subunit binding area. However, with respect to His-148 and Tyr-160, it would be reasonable to suspect that Trp-154 is the residue which is susceptible to the solvent upon dissociation of the enzyme.

By analogy with other dehydrogenases [17] it may be expected that glucose dehydrogenase will

contain a catalytic domain and a coenzyme-binding domain. The residues (His-148, Lys-201, Tyr-254) essential for the enzymatic activity probably correspond to His-176, Lys-183, Tyr-311 in the glyceraldehyde-3-phosphate dehydrogenases [18,19] and His-195, Tyr-239 in lactate dehydrogenase [20]. The exact assignment of the active site residues and the amino acids in the subunit binding area, as well as the elucidation of evolutionary relationships to other dehydrogenases, can be better accomplished by determination of the three-dimensional structure. These studies are now in progress.

ACKNOWLEDGEMENTS

This work was supported by the Deutsche Forschungsgemeinschaft (Ja 308/5). We thank Mrs E. Nitsche and Mrs E. Galaboff for their skillful technical assistance.

REFERENCES

[1] Pauly, H.E. and Pfleiderer, G. (1975) Hoppe-Seyler's Z. Physiol. Chem. 356, 1613–1623.
[2] Pauly, H.E. and Pfleiderer, G. (1977) Biochemistry 16, 4599–4604.
[3] Ulmer, W., Fröschle, M. and Jany, K.D. (1983) Eur. J. Biochem., in press.
[4] Woenckhaus, C., Berghäuser, J. and Pfleiderer, G. (1969) Hoppe-Seyler's Z. Physiol. Chem. 350, 473–483.
[5] Horton, H.R. and Tucker, W.P. (1970) J. Biol. Chem. 245, 3397–3401.
[6] Gross, E. (1967) Methods Enzymol. 11, 238–255.
[7] Fontana, A. (1972) Methods Enzymol. 25, 419–423.
[8] Jany, K.D., Keil, W., Mayer, H. and Kiltz, H.H. (1976) Biochim. Biophys. Acta 453, 62–66.
[9] Atassi, M.Z. and Habeeb, A.F.S.A. (1972) Methods Enzymol. 45, 546–553.
[10] Ponstingl, H., Krauhs, E., Little, M. and Kumpf, T. (1981) Proc. Natl. Acad. Sci. USA 78, 2757–2761.
[11] Jany, K.D. and Nitsche, E. (1983) Hoppe-Seyler's Z. Physiol. Chem. 364, 839–844.
[12] Jany, K.D. and Nitsche, E. (1983) Arch. Biochem. Biophys., in press.
[13] Reddi, K.K. and Kodicek, E. (1953) Biochem. J. 53, 286.
[14] Penke, B., Ferezi, R. and Kavacs, K. (1974) Anal. Biochem. 60, 45–50.
[15] Chen, R. (1976) Hoppe-Seyler's Z. Physiol. Chem. 357, 873–886.
[16] Wittmann-Liebold, B. and Lehmann, A. (1980) in: Methods in Peptide and Protein Sequence Analysis (Birr, C. ed) pp.49–79, Elsevier, Amsterdam, New York.
[17] Rossmann, M.G., Liljas, A., Bränden, C.-I. and Banaszak, L.J. (1975) The Enzymes, 3rd edn, vol.11, pp.61–102.
[18] Harris, J.I. and Waters, M. (1976) The Enzymes, 3rd edn, vol.13, pp.1–49.
[19] Nowak, K.K., Wolny, M. and Banaś, T. (1981) FEBS Lett. 134, 143–146.
[20] Kiltz, H.H., Keil, W., Griesbach, M., Petry, K. and Meyer, H. (1977) Hoppe-Seyler's Z. Physiol. Chem. 358, 123–127.
[21] Bozler, H., Jany, K.D. and Pfleiderer, G. (1983) Biochim. Biophys. Acta, in press.