Volume 212, number 2, 193–198 FEB 04402 February 1987

Hypothesis

An integrated prediction of secondary, tertiary and quaternary structure of glucose dehydrogenase

J. Hönes, K.-D. Jany, G. Pfleiderer and A.F.V. Wagner

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

Received 8 December 1986

Based on homology of partial sequences, on physico-chemical evidence and on studies using chemical modification, we came to the tentative conclusion that tetrameric glucose dehydrogenases from *Bacillus megaterium* and *B. subtilis* should have a structure closely related to that of lactate dehydrogenase. The overall homology of primary structures was found to be very low, however, and independent predictions of secondary structure produced a clearly different pattern of β-strands and α-helices. We nevertheless tried a manual prediction based on the hydrophobic nature of internal β-sheet and on the amphiphilic character of external helices. This treatment led to the identification of analogues of all the β-strands present in lactate dehydrogenase with the exception of βC. Six amphiphilic helices were identified corresponding to αB, αC, αD, α1F, α2F and α3G in lactate dehydrogenase. Conserved functional residues were found at analogous positions. The Q and R intersubunit contacts could be identified and partial proteolysis was found to occur on the outer surface of the tetramer. The structure was found to explain the better binding of NADP as compared to NAD^+ and offered a rationalization of the role of the essential lysine at position 201.

Integrated structural prediction; Structural homology; Sequence homology; Physico-chemical characterization; Chemical modification

1. INTRODUCTION

Pure predictions of protein structure usually try to follow the hierarchic pathway [1]. From the known primary structure, the secondary structure is predicted first [2,3]. These methods yield correct structural predictions for about two-thirds of the amino acids [4]. This is not enough to proceed to the prediction of tertiary structure, however. The inclusion of 'super-secondary-structure' into prediction algorithms seems to be promising [5], but in general, the problem of predicting from the amino acid sequence only up to the level of the functional oligomeric protein remains unsolved.

Many proteins of unknown structure are characterized fairly well by means of chemical and physico-chemical methods. Subunit composition, functionally important residues and content of α-helices and β-sheet may thus be used to add information relevant for the prediction of structure. In contrast to the 'pure' predictions, this approach is focused on single proteins rather than on a general prediction algorithm. It is an individualistic approach needing modification for any new problem. The principle of our method of prediction is given in fig.1. Detailed results with glucose dehydrogenase are described in the following sections.

Correspondence address: J. Hönes, Institut für Organische Chemie, Biochemie und Isotopenforschung der Universität Stuttgart, Pfaffenwaldring 55, 7000 Stuttgart 80, FRG

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/87/$3.50

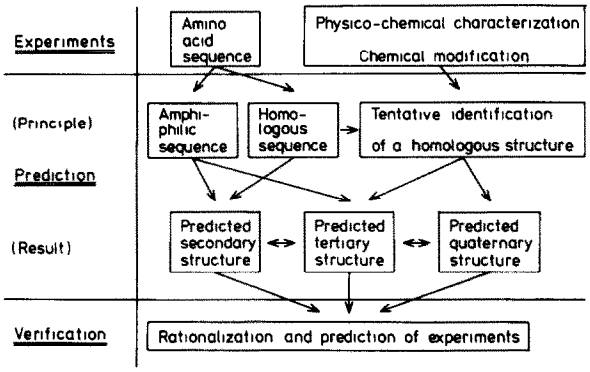


Fig.1. Scheme of structural prediction. The scheme proceeds via three levels from experiments to prediction and finally verification. The key step is the tentative identification of a structurally homologous protein with known structure. This might be tried from sequence data alone as described recently [14]. Our approach, in contrast, uses additional experimental evidence in order to narrow the range of possible structural analogues and to compensate for the limited precision [4] of the usual methods of secondary structure prediction [2,3].

2. EXPERIMENTAL RESULTS

The amino acid sequence of glucose dehydrogenase from *Bacillus megaterium* was determined some years ago [6]. Recently, the DNA-derived sequence of the closely homologous enzyme from *B. subtilis* was also published [7]. In a computer search for homologies to other known dehydrogenases, a small strand of homologous sequence at the amino-terminus was found [8]. This strand comprised the glycine-rich pyrophosphate-binding sequence. Predictions of secondary structure using the methods of Chou and Fasman [2] and Robson and Suzuki [3] led to patterns of α-helix and β-sheet without apparent homology to dehydrogenases of known structure.

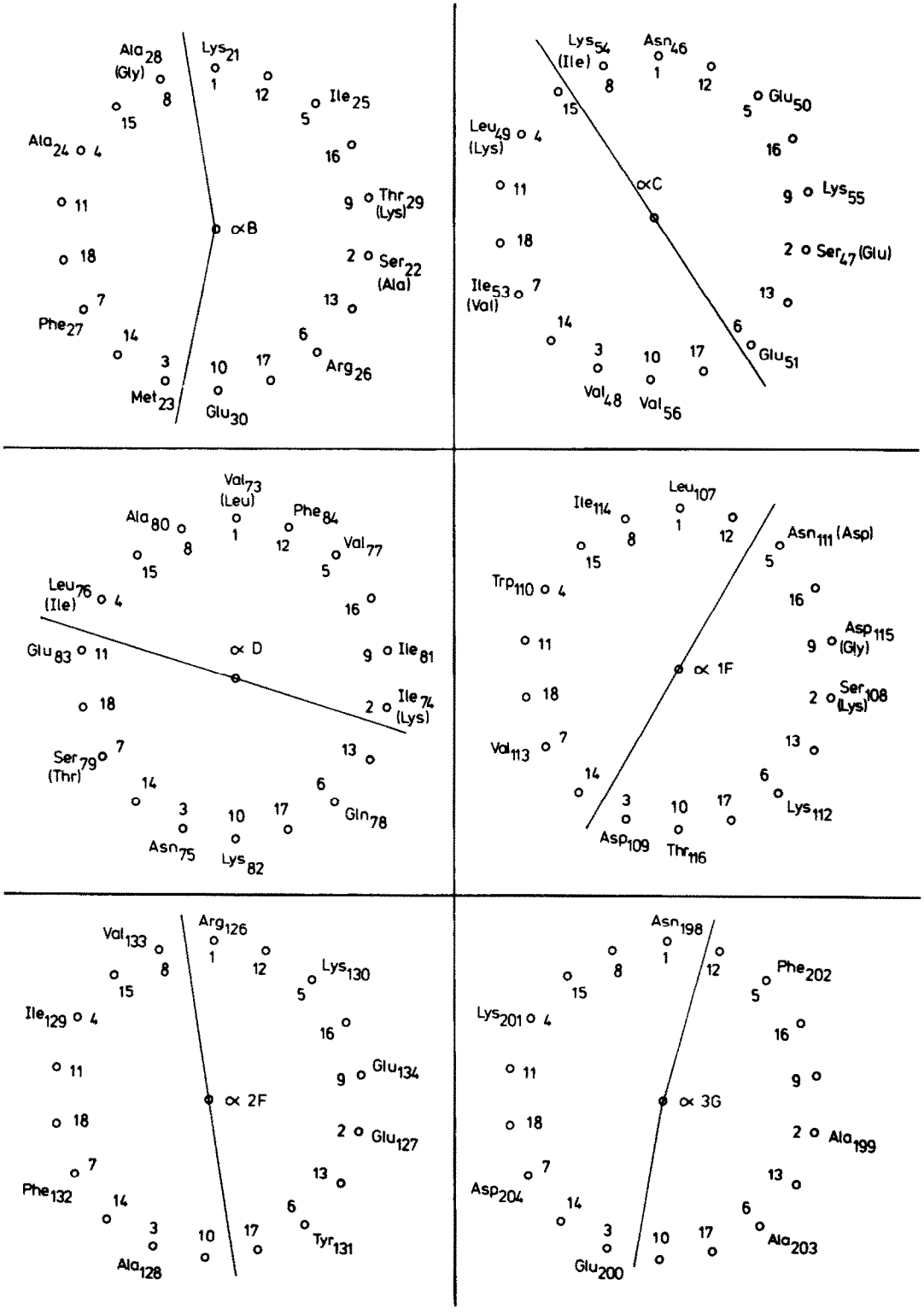
Chemical modification using bromoacetylpyridine was used to identify the essential histidine at position 148 [6]. The sequence around this essential residue was found to be closely homologous to a small stretch of sequence comprising the essential histidine in glyceraldehyde phosphate dehydrogenase. Further essential residues detected by chemical modification are Lys 201 and Tyr 254 [6]. Transamination of the amino-terminal residue by reaction with glyoxylic acid and Cu^{2+} [9] was used to introduce hydrazine labels. Circular dichroism and fluorescence of the labels provided strong evidence for a direct contact of two and two amino-termini. This was directly confirmed by crosslinking [10]. Spectrophotometric titration with 3,6,8-triiodo-7-hydroxy-4-methylcoumarin showed the existence of one hydrophobic binding site per subunit (Hönes, J., unpublished). In lactate and malate dehydrogenase, this substance has been shown to bind to the hydrophobic adenine-binding pocket [11]. Circular dichroism of the enzyme was found to be very similar to the spectra of lactate and malate dehydrogenase (not shown).

Even if some of these experiments were not strict proof, the following conclusions seemed reasonable. The enzyme was supposed to contain the pyrophosphate-binding structural moiety, the hydrophobic adenine-binding pocket and the essential histidine known from many other dehydrogenases. The content of α- and β-structure evidently was very similar to that in lactate dehydrogenase. Furthermore, an intersubunit contact similar to the R contact in lactate dehydrogenase [12] was probable. We concluded that glucose dehydrogenase most likely had a structure closely homologous to that of lactate dehydrogenase.

3. PREDICTION OF STRUCTURE

The structure of lactate dehydrogenase is characterized by an approximate alternation between hydrophobic β-strands and amphiphilic α-helices. Therefore β-strands in glucose dehydrogenase are predicted at the positions of hydrophobic peptides around amino acid nos 10, 37, 62,

Fig.2. Predicted amphiphilic helices in glucose dehydrogenase. The helical-wheel method is used [15]. The helix is viewed along its axis with the amino-terminal residue at the top. The start and end of helices are ascribed to the first and last residue fitting into the amphiphilic periodicity. The borderline between the hydrophilic and hydrophobic regions is indicated. Exchanges in the *B. subtilis* enzyme are given in parentheses. Glu 52 is deleted from the *B. megaterium* sequence since it is absent in the *B. subtilis* protein and would break the amphiphilic periodicity of αC.


Ala28 (Gly)
Lys21
Ile25
Ala24
Thr29 (Lys)
∝B
Ser22 (Ala)
Phe27
Arg26
Met23
Glu30
Lys54 (Ile)
Asn46
Glu50
Leu49 (Lys)
∝C
Lys55
Ser47 (Glu)
Ile53 (Val)
Glu51
Val48
Val56
Val73 (Leu)
Phe84
Ala80
Val77
Leu76 (Ile)
∝D
Ile81
Glu83
Ile74 (Lys)
Ser79 (Thr)
Gln78
Asn75
Lys82
Leu107
Ile114
Asn111 (Asp)
Trp110
∝1F
Asp115 (Gly)
Ser108 (Lys)
Val113
Lys112
Asp109
Thr116
Arg126
Val133
Lys130
Ile129
∝2F
Glu134
Glu127
Phe132
Tyr131
Ala128
Asn198
Phe202
Lys201
∝3G
Ala199
Asp204
Ala203
Glu200

90, 122, 155 and 230. Further β-strands are predicted from sequence homology with glyceraldehyde-phosphate dehydrogenase at around residue 145 and from sequence homology with lactate dehydrogenase at residues 187, 212 and 244. The two latter homologies are certainly weak, but 9 of the 12 β-strands necessary to form the complete β-pleated sheet can be clearly identified. The sequence of glucose dehydrogenase certainly is compatible with the existence of a β-pleated sheet homologous to the structural element found in lactate dehydrogenase. The only element lacking is βC at the amino-terminal border of the sheet. This deletion at the outer border of the sheet certainly is not crucial for the interconnection of strands. Amphiphilic helices are expected at sequence positions between the β-strands. αB is identified easily by sequence homology. It is not a perfect amphiphilic helix since Ile 25 is located within the hydrophilic region (fig.2). Five other helices with perfect amphiphilicity are predicted at positions corresponding to αC, αD, α1F, α2F and α3G in lactate dehydrogenase (fig.2). αE, α1G, α2G and αH are lacking, however. The remaining parts of the sequence are predicted to be neither α-helix nor β-sheet. The whole prediction is summarized in fig.3. Finally, some details adding credibility to the prediction should be noted. Residues with functional importance for binding of NAD^+ in lactate dehydrogenase have identical or closely similar counterparts in glucose dehydrogenase (boxed residues in fig.3). Two exchanges between the enzymes from *B. megaterium* and *B. subtilis* are simple displacements by one helix turn most probably

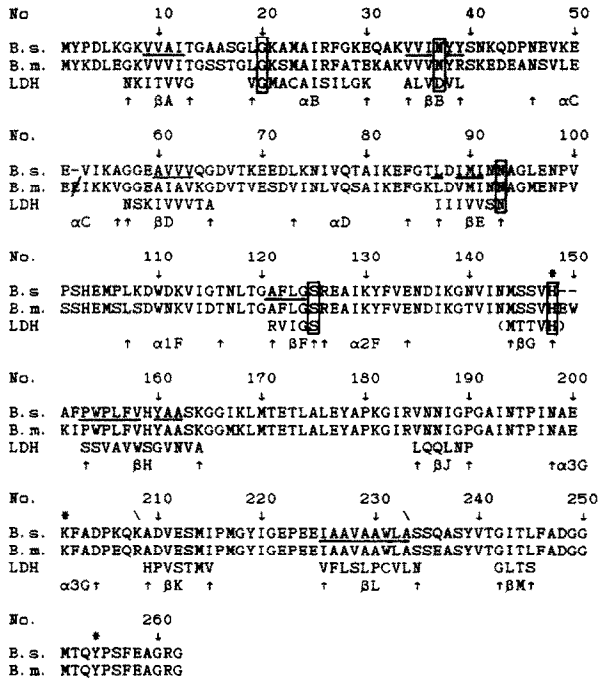


Fig.3. Predicted secondary structure of glucose dehydrogenases. The DNA-derived sequence of glucose dehydrogenase from *B. subtilis* (B.s., top) [7] is aligned to the protein sequence from *B. megaterium* (B.m., second line) [6]. The partial alignment to lactate dehydrogenase (LDH) from pig heart and to a short piece of sequence from glyceraldehyde-phosphate dehydrogenase is given in the third line. Symbols in the fourth line refer to the structure of lactate dehydrogenase [12]. Hydrophobic stretches used for the identification of β-strands are underlined. Helices are identified on the base of amphiphilicity (fig.2). Conserved functional residues are boxed. Essential residues in glucose dehydrogenase are marked by an asterisk. Proteolytic attack by trypsin at position 208 and by proteinase K at 231 is indicated by \. Ala 230 is inserted to correct for an error of the original sequence [6]. The original numbering is conserved, however.

not changing the position of the charged groups on the surface of the enzyme (Lys 55 ⟶ Lys 49 and Asp 115 ⟶ Asp 111; fig.2, helices αC and α1F). Finally the three helices αB, αC and α3G needed to form the intersubunit Q contact are present in glucose dehydrogenase. The imperfect amphiphilicity of αB may even be explained by its involvement in the contact. The results of the prediction seem to justify the initial tentative identification of the structure.

4. VERIFICATION

The prediction presented above allows for a rationalization of the role of the essential Lys 201. Modification of this residue leads to loss of activity [6]. Furthermore, complete acetamidination of all other lysines under protection by NADP does not significantly alter the dissociation of the enzyme whereas additional modification of this lysine in the absence of protective coenzyme hampers the dissociation (Thumm, M., unpublished). According to the prediction, Lys 201 is located at the outside of helix α3G in contact with αC of the opposing subunit. Thus, the influence on dissociation is easily explained. The nicotinamide moiety of NAD^+ or NADP is bound on the opposite hydrophobic side of the same helix. The protective effect of coenzyme binding against dissociation and modification of Lys 201 may be rationalized by a small conformational change of helix α3G. Therefore, the binding of coenzyme is disturbed and activity is destroyed by modification. Obviously, the conformation of the holoenzyme is destabilised with Lys 201 modified. A comparable rationalization of the role of Tyr 254 is not possible, however. Partial proteolysis of the native tetrameric enzyme produces single nicks at the carboxy-termini of Arg 208 [10] and Ala 231 [13], respectively. According to the structural prediction, these sites are located at the beginning of βK and at the end of βL on the surface of the enzyme in close proximity to one another.

The aspartate in βB of lactate dehydrogenase binding 2′-OH of the adenosine ribose is replaced by Asn 37 in glucose dehydrogenase. This finding led to the prediction that NADP should be bound better than NAD^+ by our enzyme, since the unfavorable interaction of phosphate with carboxylate in lactate dehydrogenase should be replaced by a favorable one. Exactly this is the result of kinetic experiments and titrations (table 1). The crucial role of this residue for discrimination between NAD^+ and NADP has been suggested earlier (cf. [1], p.221).

Table 1

Binding of coenzymes to glucose dehydrogenase from *Bacillus megaterium*

Coenzyme	K_m (M) (pH 8.0)	K_d (M) (pH 6.5)
NAD^+	7×10^{-3}	$>1 \times 10^{-3}$
NADP	7.5×10^{-5}	1.7×10^{-5}

K_m was measured in 0.5 M Tris-HCl (pH 8.0) with 0.2 M glucose at 25°C. K_d was measured in 0.1 M phosphate buffer (pH 6.5) at 25°C by titration with NADP. The signal observed was the induced ellipticity from adenine at 255 nm. Curve fitting was done as described in [11]

According to these three pieces of evidence, the prediction seems reasonable. For a final judgement, we have to await the X-ray structural analysis which is in progress now.

ACKNOWLEDGEMENT

We gratefully acknowledge the predictions of secondary structure made by H. Ponstingl, DKFZ, Heidelberg.

REFERENCES

[1] Schulz, G.E. and Schirmer, R.H. (1979) Principles of Protein Structure, p.67, Springer, New York.
[2] Chou, P.Y. and Fasman, G.D. (1974) Biochemistry 13, 222–244.
[3] Robson, B. and Suzuki, E. (1976) J. Mol. Biol. 107, 327–356.
[4] Kabsch, W. and Sander, C. (1983) FEBS Lett. 155, 179–182.
[5] Taylor, W.R. and Thornton, J.M. (1983) Nature 301, 540–542.
[6] Jany, K.-D., Ulmer, W., Fröschle, M. and Pfleiderer, G. (1984) FEBS Lett. 165, 6–10.
[7] Lampel, K.A., Uratani, B., Chaudhry, G.R., Ramaley, R. and Rudikoff, S. (1986) J. Bacteriol. 166, 238–243.
[8] Jörnvall, H., Bahr-Lindström, H., Jany, K.-D., Ulmer, W. and Fröschle, M. (1984) FEBS Lett. 165, 190–196.

[9] Dixon, H.B.F. and Fields, R.J. (1972) Methods Enzymol. 25, 409–419.
[10] Wagner, V. (1986) Dissertation, Stuttgart.
[11] Hönes, J. (1985) Biol. Chem. Hoppe-Seyler 366, 561–566.
[12] Holbrook, J.J., Liljas, A., Steindel, S.J. and Rossmann, M.G. (1975) The Enzymes 11, 191–292.
[13] Jany, K.-D. and Nitsche, E. (1983) Hoppe-Seyler's Z. Physiol. Chem. 364, 839–844.
[14] Argos, P. and MohanaRao, J.K. (1986) Methods Enzymol. 130, 185–207.
[15] Schiffer, M. and Edmundson, A.B. (1967) Biophys. J. 7, 121–135.