SEQUENCE OF *ESCHERICHIA COLI* D-SERINE DEHYDRATASE

Location of the pyridoxal-phosphate binding site

Emile SCHILTZ and Wilfried SCHMITT
Department of Physiological Chemistry, The University of Würzburg, Koellikerstr. 2, 8700 Würzburg, FRG

Received 23 September 1981

1. Introduction

D-Serine dehydratase (EC 4.2.1.14) from *Escherichia coli* is a pyridoxal 5′-phosphate (PLP)-dependent enzyme which catalyzes the conversion of D-serine to pyruvate and ammonia. D-Serine is a competitive antagonist of β-alanine in the biosynthetic pathway to pantothenate and coenzyme A [1]. The dehydratase was reported to have M_r 45 500 [2] and in contrast to other PLP-dependent enzymes it is a monomer. We have chosen it as a model for structural studies on the environment of PLP in enzymes catalyzing α,β-elimination reactions [3–5] and also for elucidating the degree of structural similarity among PLP enzymes. Whereas numerous hydrolases and oxidoreductases have been examined in molecular detail by crystallographers and protein chemists, similar studies on the large group of PLP-dependent enzymes have only just begun. The impressive results on aspartate aminotransferase and phosphorylase do not yet allow us to draw general conclusions, e.g., on the structure of a PLP domain or on the catalytically active amino acid or pyridine side chains. This paper describes the nearly complete primary structure of D-serine dehydratase established by isolation and sequence analysis of peptides obtained by cleavage with trypsin, clostripain, *Staphylococcus aureus* protease, CNBr and 2-nitro-5-thiocyanobenzoate (NTCB). The fragments could be ordered in two large pieces:

(i) The N-terminal part with 185 residues which includes the PLP binding lysine at position 118; and
(ii) The C-terminal fragment of 264 residues which contains all 5 cysteines.

This adds up to 449 amino acids and a minimum M_r of 48 790/molecule. The characteristics of the sequence as well as results of a preliminary comparison with other PLP enzymes are discussed.

2. Materials and methods

D-Serine dehydratase was prepared as in [3] from *E. coli* K-12 mutant C 6, kindly provided by Professor E. E. Snell and grown in kg amounts by Merck (Darmstadt). Reduction and carboxymethylation with iodo[$^{14}C_2$]acetic acid were performed according to [6]. Enzymic digestions on 1–2 μmol dehydratase with trypsin, clostripain and *Staph. aureus* protease as well as secondary cleavage of peptides with chymotrypsin, thermolysin and carboxypeptidases A and B were carried out as in [3,7,8]. Cleavages by CNBr after Met [9] and with 2-nitro-5-thiocyanobenzoate (generous gift of Dr R. Gracy) before Cys [10] were performed under standard conditions. Peptides were prefractionated on Sephadex columns and purified by ion-exchange chromatography (Sephadex ion-exchangers for peptides larger than 20 residues and Aminex A5 for smaller peptides) and cellulose thin-layer electrophoresis and/or chromatography [3,7,8,12]. Acid hydrolysis, amino acid analysis and sequence determination were performed as in [7].

3. Results and discussion

The nearly complete sequence of D-serine dehydratase is proposed in table 1. It includes all peptides isolated in good yield and characterized so far from digests of the protein by trypsin, clostripain and *Staph. aureus* protease as well as the fragments from CNBr and NTCB cleavages. The N-terminal sequence

Published by Elsevier/North-Holland Biomedical Press
00145793/81/0000–0000/$02.75

Table 1
Partial amino acid sequence of *Escherichia coli* D-serine dehydratase

10 20
Met-Glu-Asn-Ala-Lys-Met-Asn-Ser-Leu-Ile-Ala-Gln-Tyr-Pro-Leu-Val-Lys-Asp-Leu-Val-
30 40
Ala-Leu-Lys-Glu-Thr-Thr-Trp-Phe-Asn-Pro-Gly-Thr-Thr-Thr-Leu-Ala-Glu-Gly-Leu-Pro-
50 60
Tyr-Val-Gly-Leu-Thr-Glu-Gln-Asp-Val-Gln-Asp-Ala-His-Ala-Arg-Leu-Ser-Arg-Phe-Ala-
70 80
Pro-Tyr-Leu-Ala-Lys-Ala-Phe-Pro-Glu-Thr-Ala-Ala-Thr-Gly-Gly-Ile-Ile-Glu-Ser-Glu-
90 100
Leu-Val-Ala-Ile-Pro-Ala-Met-Gln-Lys-Arg-Leu-Glu-Lys-Glu-Tyr-Gln-Gln-Pro-Ile-Ser-
110 120
Gly-Gln-Leu-Leu-Leu-Lys-Lys-Asp-Ser-His-Leu-Pro-Ile-Ser-Gly-Ser-Ile-Lys-Ala-Arg-
130 140
Gly-Gly-Ile-Tyr-Glu-Val-Leu-Ala-His-Ala-Glu-Lys-Leu-Ala-Leu-Glu-Ala-Gly-Leu-Leu-
150 160
Thr-Leu-Asp-Asp-Asp-Tyr-Ser-Lys-Leu-Leu-Ser-Pro-Glu-Phe-Lys-Gln-Phe-Phe-Ser-Gln-
170 180
Tyr-Tyr- X - X - X - X - X -Tyr-Ser-Ile-Ala-Val-Gly-Ser-Thr-Gly-Asn-Leu-Gly-Leu-
190' 200'
Ser-Ile-Gly-Ile-Met Ser-Ala-Arg-Ile-Gly-Phe-Lys-Val-Thr-Val-His-Met-Ser-Ala-Asp-
210' 220'
Ala-Arg-Ala-Trp-Lys-Lys-Ala-Lys-Leu-Arg-Ser-His-Gly-Val-Thr-Val-Val-Glu-Tyr-Glu-
230' 240'
Gln-Asp-Tyr-Gly-Val-Ala-Val-Glu-Glu-Gly-Arg-Lys-Ala-Ala-Gln-Ser-Asp-Pro-Asn-Cys-
250' 260'
Phe-Phe-Ile-Asp-Asp-Glu-Asn-Ser-Arg-Thr-Leu-Phe-Leu-Gly-Tyr-Ser-Val-Ala-Gly-Gln-
270' 280'
Arg-Leu-Lys-Ala-Gln-Phe-Ala-Gln-Gln-Gly-Arg-Ile-Val-Asp-Ala-Asp-Asn-Pro-Leu-Phe-
290' 300'
Val-Tyr-Leu-Pro-Cys-Gly-Val-Gly-Gly-Gly-Pro-Gly-Gly-Val-Ala-Phe-Gly-Leu-Lys-Leu-
310' 320'
Ala-Phe-Gly-Asp-His-Val-His-Cys-Phe-Phe-Ala-Glu-Pro-Thr-His-Ser-Pro-Cys-Met-Leu-
330' 340'
Leu-Gly-Val-His-Thr-Gly-Leu-His-Asp-Gln-Ile-Ser-Val-Gln-Asp-Ile-Gly-Ile-Asp-Asn-
350' 360'
Leu-Thr-Ala-Ala-Asp-Gly-Leu-Ala-Val-Gly-Arg-Ala-Ser-Gly-Phe-Val-Gly-Arg-Ala-Met-
370' 380'
Glu-Arg-Leu-Leu-Asp-Gly-Phe-Tyr-Thr-Leu-Ser-Asp-Gln-Thr-Met-Tyr-Asp-Met-Leu-Gly-
390' 400'
Trp-Leu-Ala-Gln-Glu-Glu-Gly-Ile-Arg-Leu-Glu-Pro-Ser-Ala-Leu-Ala-Gly-Met-Ala-Gly-
410' 420'
Pro-Gln-Arg-Val-Cys-Ala-Ser-Val-Ser-Tyr-Gln-Gln-Met-His-Gly-Phe-Ser-Ala-Glu-Gln-
430' 440'
Leu-Arg-Asn-Thr-Thr-His-Leu-Val-Trp-Ala-Thr-Gly-Gly-Gly-Met-Val-Pro-Glu-Glu-Glu-

Met-Asn-Gln-Tyr-Leu-Ala-Lys-Gly-Arg-OH

of 23 residues published earlier together with the tryptic coenzyme binding peptide [3] could be confirmed and extended up to position 185. Pyridoxal-5′-phosphate is bound to the ε-amino group of Lys 118. The last tryptic peptide of the N-terminal part starting with Gln 156 could not be isolated in good yield and 5 of its residues could not be determined unambiguously. The other positions were obtained by sequencing the corresponding chymotryptic and clostripain peptides. The central part of D-serine dehydratase contains the longest CNBr peptide (positions 198′–319′) and 4 of the 5 cysteines. In contrast, the C-terminal third of the protein is particularly rich in Met and contains only 1 Lys. In table 1 the positions of the second fragment are numbered provisionally 186′–449′. The gap between the 2 fragments could not be filled so far. It is possible that the preliminary length of the polypeptide chain is final since no other CNBr peptide was found [12]. M_r-Value determination of the longest NTCB-peptide by SDS–PAGE indicates that this fragment, which was found to be the N-terminal half of the protein, is probably not much larger than that sequenced so far: If the 5 undetermined positions 163–167 are given an average M_r of 110, the preliminary sequence of this peptide adds up to 26 070 M_r/molecule. From the mobility on slab gels M_r of 25 700, 27 000 and 27 500 were determined in 10%, 13% and 16% polyacrylamide, respectively. Also our M_r determinations of the total enzyme on SDS–PAGE (53 000, 50 900 and 48 000 in 10%, 13% and 16% polyacrylamide, respectively) corresponds reasonably well with the minimum value of 48 790 M_r calculated from the sequence data. The final proof for the link between the 2 fragments depends however on the isolation of an overlapping peptide. The purification will be difficult, due to the length of the corresponding tryptic peptide and to its N-terminal Gln.

Since DNA sequence analysis of the D-serine dehydratase structural gene has been started, results of the 2 methods should be complementary. So far 3 DNA fragments were sequenced confirming 1/3rd of the protein sequence: Positions Ser 100–Ser 147 including Lys 118, Asn 277′–Asp 304′ and Asp 329′–Val 404′ (E. McFall, personal communication).

An interesting question is how related the sequence of the dehydratase is to the other two PLP enzymes, pig heart mitochondrial aspartate aminotransferase, sequenced in [13,14] and the β-subunit of *E. coli* tryptophan synthase [15]. We used a computer program similar to program RELATE in [16] for detecting distant relationships and the scoring matrix for matched pairs (fig.84 of [16]) derived from accepted point mutations in closely related proteins. We searched for hexapeptides of D-serine dehydratase fitting best to hexapeptides of one of the 2 other enzymes. The best alignments were extended and a limited number of gaps (<15% of the alignment length) were introduced manually if this improved the score. A gap penalty parameter of −2 was used. Scores were calculated as mean valuation of the matched residues in an alignment. For each comparison at least 100 random alignments of the same composition were scored. The quality of fit describes where the pair of real sequences is located in the normal distribution of the random sequences. The difference from the mean of the random sequence is expressed in standard deviation or σ units as in [16]. In general, the score can be especially high for selected regions of 10–20 residues. On the other hand, the quality of fit increases when longer sequences are compared, because then the standard deviation of the scores of random sequences decreases. The procedure used is a compromise between high scores and limitation in alignment breaks. Since number, size and valuation of gaps remain arbitrary some variation in the alignment is possible, but the quality of fit of the examples in table 2 demonstrates the degree of relatedness that can be found between the 3 enzymes.

Besides the increase in chain length by at least 10% the most striking difference is the position of the PLP binding site: Lys 118 in the dehydratase, Lys 86 in the synthase [17] and Lys 250 in the aminotransferase [13,14]. The difference observed when comparing the two α,β-eliminating enzymes can be explained by an N-terminally extended D-serine dehydratase: In the segment comparison the highest score is obtained for the 9 N-terminal amino acids of the synthase when they are aligned with the dehydratase sequence starting at Thr 33 and when 2 residues of the longer enzyme (Gly 38 and Leu 39) are omitted. As a remarkable result the positions for the coenzyme binding lysines become identical (86 + 32 = 118). Alignment of the N-terminal 88 residues of the synthase subunit without gaps with the dehydratase starting at position 33 yields an insignificantly low score of −0.46. When 10 residues in each sequence are omitted, i.e., left unpaired in the alignment (corresponding to 13 gaps), a score of 1.3 is reached with a quality of fit of 7.8 σ (partially shown in table 2).

Continuous identity between the sequences of the

Table 2
Examples of related sequences between *E. coli* D-serine dehydratase (DSD), pig heart mitochondrial aspartate aminotransferase (AAT) [13,14] and *E. coli* tryptophan synthase B-subunit (TSB) [15]

	Positions	Sequences in one letter code[+]	Score*	Quality of fit* (σ units)
AAT	1–20	SSWWAHVEMGPPD—PILGVTE··	1.85	4.3
DSD	25–46	··TTWFNPGTTTLAEGLPYVGLTE··	3.22	3.7
TSB	1–9	TTLLN—PYFG··		
AAT	50–116	··SVRKAE–A–QIAA—KNLDKEYL–PIGG—L—AEFCKASAELALGENNEVLKSGRYVTVQTISGTGALRIGANFL··	1.32	8.0
DSD	75–145	··GIIESELV–AIPAMQKRLEKEYQQPISG—Q—LLLKKDS–HLPISGSI–(K)ARGGIYEVLAHA—EK–LALEAGLLTLDDD··	1.47	8.9
TSB	44–110	··FN–DLLKNYAGRPTALTK–CQN–ITAGTNTTLYLKRED–LLH–GGAH–(K)TN———QVLGQALLAKRMG–KTEIIA–ETG··		
AAT	217–262	··YQGFA–SGDGNKDAWAVR—HFIEQG—INVCL–CQSY———A(K)N–MGL——YGERVGAF··	1.58	6.7
DSD	252′–309′	··FLGYSVAGQRLKAQFAQ–QGRIVDADNPLFVYLPCGVGGGPGGVA–FGLKLAFGDHVHCF··	2.29	7.9
TSB	206–241	··MIGEETKAQILEREGRLPDA———VIAC–VGGGSNAIGMFA··		
DSD	382′–396′	··LAQEEGI—RLEPS–ALA··	2.27	4.7
TSB	338–355	··LCLHEGIIPALESSHALA··		
AAT	384–401	··VTSG–NVGYLAHAIHQ–VTK	1.11	4.2
DSD	430′–449′	··ATGG–GM–VPEEEMNQYLAKGR	1.56	4.1
TSB	377–396	··GRGDKDI–F——TVHDLL–KARGEI		

[+] No residues are omitted in the sequences shown; – is a direct link between 2 residues and indicates a gap in the comparison. Identical residues in 2 compared segments are underlined; PLP binding Lys are encircled

* The calculation of relatedness is based on a matrix of relative amino acid substitution frequencies in closely related proteins [16]. The score is defined as the mean valuation of each possible pair formed by an amino acid in the upper sequence with its corresponding residue in the lower sequence. For gaps a penalty parameter of −2 was used. The quality of fit is the difference between the score of the comparison shown and the mean score of at least 100 random comparisons of the same composition, divided by the standard deviation of the random comparisons

two α,β-eliminating enzymes is limited to dipeptides [12], tripeptides [5] and 1 tetrapeptide. Nevertheless, related sequence segments are distributed all over the 2 polypeptide chains, as shown in table 2 for the N-terminal, C-terminal and PLP binding regions as well as for another identity-rich fragment. The high quality of fit for the PLP binding sequences together with the occurrence of a short palindrome proximal to the PLP binding Lys of both β-eliminating enzymes makes it unlikely that this is a chance event. The similarity between these 2 enzymes might not be unexpected, since the isolated β_2-subunit of tryptophan synthase can also deaminate serine [18]. But differences exist which must be accounted for, such as the binding site for the other β-subunit and for the α-subunit. The increased chain length of the monomeric dehydratase, in particular the N-terminal extension, might compensate the absence of a second subunit found in all other vitamin B_6-dependent enzymes. Similarly, the relatively large number of amino acid doublets [26] and triplets [6], which involves 70 of the 449 residues (15.6% in comparison to 8.1% in the synthase and 9.6% in the transaminase) might be a consequence of the relatively small overall size of the dehydratase.

It is surprising that the aminotransferase can be aligned to the dehydratase with a similar high quality of fit (table 2). In an overall alignment of the 2 enzymes (not shown) 75 out of 401 residues (18.7%) could be placed to become identical. This corresponds to 83.5% difference (alignment length−identities/alignment length) as defined in [16]. The unpaired terminal positions are not counted. For the two β-eliminating enzymes an alignment resulted in 87 identical residues (22%) and in 80.8% difference. We conclude that the 3 PLP-dependent enzymes belong to the same or a closely related sequence superfamily as defined in [16].

How related are the 3 proteins in their tertiary structure? The cytosolic and the mitochondrial aspartate aminotransferases have 48% sequence identity and their 3-dimensional structures are closely related [19]. Examples exist for less closely related proteins like protease B and chymotrypsin (13% identity), hemoglobulin α-chain and myoglobin (27% identity) where the chain folds can be superimposed and differences occur only at the surface of the molecules [20]. However, the fact that the PLP binding lysines occupy quite different positions in the chains of the α,β-eliminating enzymes and the aminotransferases does not support the idea of overall structural similarity. Only X-ray crystallography of other PLP-dependent enzymes can show whether or not there exists a common domain for PLP binding in different amino acid-metabolising enzymes.

Acknowledgements

We wish to thank Mrs Monika Kirsch, Ms Rosi Blatterspiel and Ms Birgitt Sorg for their excellent technical assistance. We are endebted to Dr Gabriele Lachmann for developing computer programs for the sequence comparisons. We are grateful to Dr Klaus Schnackerz for his encouraging interest in our work and to Professor Ernst Helmreich for his support and critical review of the manuscript. This work was supported by grants from the Deutsche Forschungsgemeinschaft (Schi 145/1–5) and from the Volkswagen Stiftung. Part of this work is from the PhD thesis of W. S.

References

[1] Maas, W. K. and Davis, B. D. (1950) J. Bacteriol. 60, 733–740.
[2] Dowhan, W. jr and Snell, E. E. (1970) J. Biol. Chem. 245, 4618–4628.
[3] Schiltz, E. and Schnackerz, K. D. (1976) Eur. J. Biochem. 71, 109–116.
[4] Kazarinoff, M. N. and Snell, E. E. (1976) J. Biol. Chem. 251, 6179–6182.
[5] Schnackerz, K. D., Feldmann, K. and Hull, W. E. (1979) Biochemistry 18, 1536–1539.
[6] Heil, A., Müller, G., Noda, L., Pinder, Th., Schirmer, H., Schirmer, I. and Von Zabern, I. (1974) Eur. J. Biochem. 43, 131–144.
[7] Schiltz, E., Blatterspiel, R. and Untucht-Grau, R. (1979) Eur. J. Biochem. 102, 269–278.
[8] Schächtele, K.-H., Schiltz, E. and Palm, D. (1978) Eur. J. Biochem. 92, 427–435.
[9] Gross, E. and Witkop, B. (1962) J. Biol. Chem. 237, 1856–1860.
[10] Jacobson, G. R., Schaffer, M. H., Stark, G. R. and Vanaman, Th. C. (1973) J. Biol. Chem. 248, 6583–6591.
[11] Gibbons, I. and Perham, R. N. (1970) Biochem. J. 116, 843–849.
[12] Schmitt, W. (1980) PhD thesis, Universität Würzburg.
[13] Kagamiyama, H., Sakakibara, R., Wada, H., Tanase, S. and Morino, Y. (1977) J. Biochem. 82, 291–294.
[14] Barra, D., Bossa, F., Doonan, S., Fahmy, H. M. A., Hughes, G. J., Kakoz, K. Y., Martini, F. and Petruzelli, R. (1977) FEBS Lett. 83, 241–244.
[15] Crawford, I. P., Nichols, B. P. and Yanofsky, Ch. (1980) J. Mol. Biol. 142, 489–502.

[16] Schwartz, R. M. and Dayhoff, M. O. (1978) in: Atlas of Protein Sequence and Structure, vol. 5, suppl. 3, pp. 353–358, fig.84, Natl. Biomed. Res. Found., Washington DC.
[17] Higgins, W., Miles, E. W. and Fairwell, Th. (1980) J. Biol. Chem. 255, 512–517.
[18] Kumagai, H. and Miles, E. W. (1971) Biochem. Biophys. Res. Commun. 44, 1271–1278.
[19] Ford, G. C., Eichele, G. and Jansonius, J. N. (1980) Proc. Natl. Acad. Sci. USA 77, 2559–2563.
[20] Schulz, G. E. and Schirmer, R. H. (1979) in: Principles of Protein Structure, pp. 166–205, Springer, New York.