Acetyl-blocked N-terminal structures of sorbitol and aldehyde dehydrogenases

Thomas Fairwell, Henry Krutzsch, John Hempel, Jonathan Jeffery and Hans Jörnvall

Molecular Disease Branch, NHLBI and Laboratory of Experimental Carcinogenesis, NCI, Bethesda, MD 20205, USA; Department of Biochemistry, University of Aberdeen, Marischal College, Aberdeen AB9 1AS, Scotland and *Department of Chemistry I, Karolinska Institutet, S-104 01 Stockholm, Sweden*

Received 26 March 1984

Two new dehydrogenase structures, the 354-residue polypeptide chain of sorbitol dehydrogenase (from sheep liver) and the 500-residue polypeptide chain of cytoplasmic aldehyde dehydrogenase (from human liver), have blocked N-termini. The N-terminal peptides were purified by reverse-phase high-performance liquid chromatography and submitted to mass spectrometry after derivatization. They were also analyzed by dipeptidyl carboxypeptidase digestion, utilizing gas chromatography-mass spectrometry for dipeptide identifications. Results are consistent and establish that sorbitol dehydrogenase has N-terminal acetylalanine and aldehyde dehydrogenase N-terminal acetylserine in amino acid sequences that are compatible with estimates from chemical analyses. The two N-terminal residues found are typical of acetylated proteins in general, extend the group of known acetylated dehydrogenases, and show that these intracellular proteins are frequently N-terminally acetylated.

Acetylated N-terminus *Amino acid sequence* *Mass spectrometry* *Dipeptidyl carboxypeptidase*

1. INTRODUCTION

N-terminal acetylation is a post-translational modification found in many proteins [1], and the structures involved have some common properties. Thus, the N-terminal acetylated residue itself is a small residue, most often serine (about half of all cases) or alanine (about half of the remainder), while the N-terminal region has on average a distinctive distribution of charged residues (frequently several adjacent) and of branched-chain residues (Ile over-represented in relation to Leu, Val). All these properties suggest a general function for acetylation. A protective role against N-terminal proteolysis appears possible, although, in special cases, other roles have been observed [1] and considerable resistance to proteolysis even without N-terminal acetylation has been noticed [2].

N-terminal acetylation appears especially common among several structural proteins [1], such as viral capsid proteins, but many intracellular proteins, including dehydrogenases, are also acetylated at the N-terminus. Lactate dehydrogenase [3,4], glyceraldehyde-3-phosphate dehydrogenase [5] and alcohol dehydrogenase [6] are acetylated in several species. The blocking acetyl groups complicate direct sequence analysis, though incomplete stoichiometry of acetylation may sometimes permit low-yield N-terminal sequence determinations [7].

Two further types of dehydrogenase were recently characterized; sorbitol dehydrogenase [8] and cytoplasmic aldehyde dehydrogenase [9] were then found also to be N-terminally blocked. Specificities of proteolytic enzymes, total compositions of small blocked peptides, and results of limited acid hydrolysis [8,9] indicated preliminary amino acid sequences even for the N-terminal regions of these proteins and suggested the blocking groups to be acyl moieties, probably acetyl groups.

Published by Elsevier Science Publishers B.V.
00145793/84/$3.00

Small N-terminal peptides from sorbitol and cytoplasmic aldehyde dehydrogenases have now been purified and fully analyzed. The dipeptidyl carboxypeptidase/gas chromatography-mass spectrometry method [10], as well as mass spectrometry of the esterified peptide derivatives, establish the presence of native, N-terminal acetyl groups and confirm the amino acid sequences deduced. The results also permit further comparisons of acetyl-blocked dehydrogenase structures and show that enzymes of this group are often acetylated, terminal residues being typical of acetylated proteins in general.

2. MATERIALS AND METHODS

Sorbitol dehydrogenase and aldehyde dehydrogenase were prepared, carboxymethylated, and cleaved by proteolytic enzymes and CNBr as described [8,9,11]. Peptides were purified by reverse-phase high-performance liquid chromatography (HPLC) on μBondapak C18 using a Waters instrument [11]. Total compositions were determined with a Beckman 121 M analyzer after acid hydrolysis, and carboxypeptidase Y digestions were performed in 0.1 M pyridine acetate (pH 6.0) monitoring the amino acid release [12].

Esterification and mass-spectrometric analysis of the peptides were performed as in [7]. Analysis by gas chromatography-mass spectrometry of dipeptidyl carboxypeptidase treated samples was carried out after trimethylsilylation as in [10]. Peptides were treated for 4 h at 37°C with dipeptidyl carboxypeptidase. The digests were freeze-dried, heated for 10 min with bis(trimethylsilyl)trifluoroacetamide–acetonitrile (1:1, v/v), and analyzed for the presence of dipeptides by gas chromatography-mass spectrometry, using an LKB model 2091, a 25 μm SE-54 fused silica column and splitless injections.

3. RESULTS

3.1. *Purification of acyl-blocked N-terminal fragments*

Material from the acyl-blocked N-terminal region of sorbitol dehydrogenase was prepared from the N-terminal CNBr fragment (CB2:2 in [11]). Subsequent re-cleavage with Glu-specific staphylococcal protease and purification by reverse-phase HPLC gave the blocked N-terminal hexapeptide (P1 in fig.3 of [11]). Data from chemical analyses of this peptide are shown in table 1. The tentative amino acid sequence shown was established by carboxypeptidase Y digestions (liberating Glu and Ala) and by the fact that the peptide is not cleaved by trypsin [11] (suggesting lysine to be in a Lys-Pro structure).

Material from the acyl-blocked N-terminal region of cytoplasmic aldehyde dehydrogenase was prepared from the N-terminal tryptic peptide (T1 in [9]). Different samples of this peptide were cleaved with *Astacus fulvatillus* protease, limited acid hydrolysis, thermolysin, or pepsin, and the blocked N-terminal peptides in each case were purified by reverse-phase HPLC as shown in fig.1. Data for the chemical analysis of these peptides are given in table 1. The tentative amino acid sequence shown was established by the total compositions of L1 and L2, and by Edman degradations of other peptides starting at positions 3,4 and 5 (from digestions with *Astacus fulvatillus* protease [9]).

3.2. *Mass spectrometric analysis*

Peptides P1 from sorbitol dehydrogenase and L2 from aldehyde dehydrogenase were acetylated and esterified as in [7]. They were then analyzed by mass spectrometry after introduction of the sample through the solid inlet probe. Results are shown in fig.2 (top panels), together with the sequence assignments of those peaks that could be identified to correspond to the sequence determining fragments.

Peptide P1 of sorbitol dehydrogenase gives a fairly good and conclusive fragmentation pattern. The important sequence determining peaks are the ones at m/z 86, 114, 284, 382, 453 and 524, which are identified as shown in fig.2. Some of these peaks are the results of pyrolysis and the actual peak may be off from the expected value by m/z 1, because of protonation.

Peptide L2 of aldehyde dehydrogenase does not give a good spectrum, as shown in fig.2 (bottom). Because of the 3 serine residues (cf. table 1), peptide L2 undergoes extensive dehydration and rearrangements (plus possibly some polymerization). Consequently, some of the peaks are difficult to interpret. However, the important ions are m/z 88, 89, 144 and 172. These peaks prove that the C-

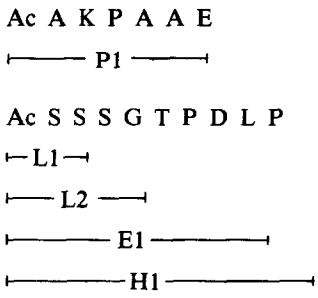
Table 1

Data for acyl-blocked N-terminal peptides of sorbitol dehydrogenase (SDH) and cytoplasmic aldehyde dehydrogenase (AldDH)

(A)

Peptide	SDH P1		Ald DH E1		Ald DH H1		Ald DH L1		Ald DH L2	
Composition										
Asp	–	–	1.0	(1)	1.0	(1)	–	–	–	–
Thr	–	–	1.0	(1)	0.9	(1)	–	–	–	–
Ser	–	–	2.9	(3)	2.5	(3)	1.8	(2)	2.9	(3)
Glu	1.2	(1)	–	–	–	–	–	–	–	–
Pro	1.0	(1)	1.0	(1)	2.2	(2)	–	–	–	–
Gly	–	–	1.1	(1)	1.0	(1)	0.4	–	1.0	(1)
Ala	3.0	(3)	–	–	–	–	–	–	–	–
Leu	–	–	1.0	(1)	0.9	(1)	–	–	–	–
Lys	0.9	(1)	–	–	–	–	–	–	–	–

(B)



(A) Compositions; (B) Structures. Peptide P1 was obtained by digestion with the extracellular, Glu-specific protease from *Staphylococcus*, E1 by digestion with pepsin, H1 by digestion with thermolysin, and L1, L2 by limited acid hydrolysis

terminus is Gly and the N-terminus Ser. Together with the amino acid composition of peptide L2, these results are fully conclusive.

Two synthetic peptides corresponding to the sequences of P1 and L2 were also analyzed by mass spectrometry and the spectra were found to be identical to those of the native peptides, further confirming their structures.

3.3. *Gas chromatography-mass spectrometry of samples treated with dipeptidyl carboxypeptidase*

Peptides L2, E1 and H1 of aldehyde dehydrogenase (cf. table 1) were analyzed using the dipeptidyl carboxypeptidase sequence methodology. Results obtained are shown in fig.3. Peptide L2 yielded identifiable peaks corresponding to Ac-Ser-Ser and Ser-Gly. The blocking group could then be identified as acetyl and together with the total composition (table 1) the sequence of L2 could be deduced to be Ac-Ser-Ser-Ser-Gly. Similar dipeptidyl carboxypeptidase treatment of the other two polypeptides yielded Ac-Ser-Ser, Ser-Gly, Thr-Pro and Asp-Leu for E1 and Ac-Ser, Ser-Ser, Gly-Thr, Pro-Asp and Leu-Pro for H1. Combined with the total compositions, these dipeptides give the whole sequence of H1 and prove the structures to be as shown in table 1.

Results by the dipeptidyl carboxypeptidase method and the analysis by direct inlet probe fully

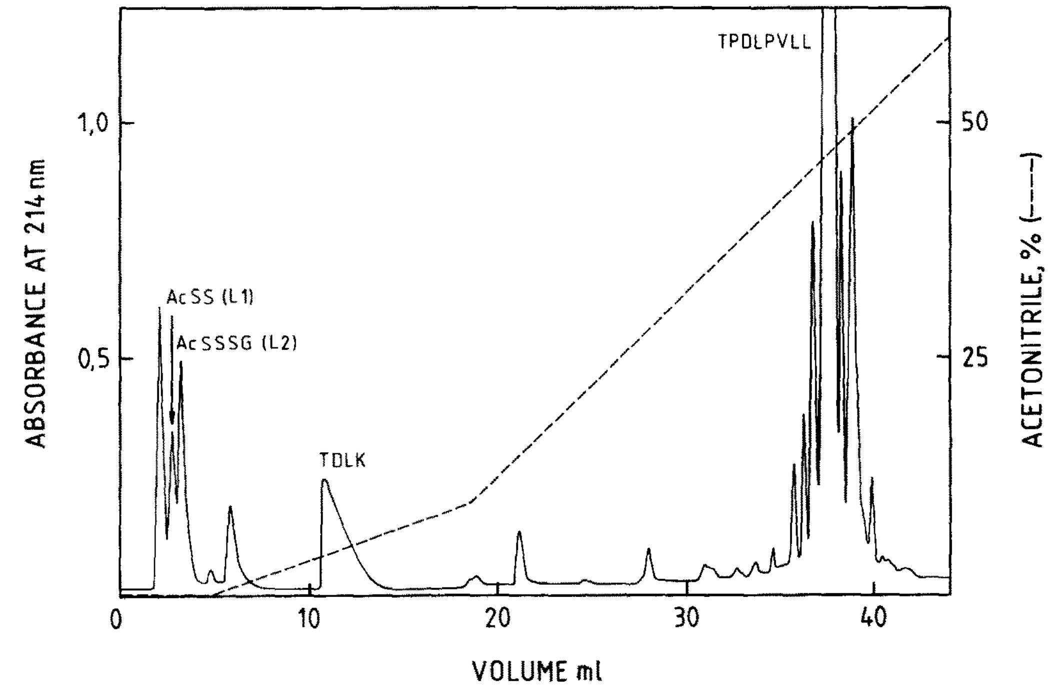
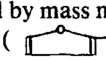


Fig.1. Purification by reverse-phase HPLC of the N-terminally acyl-blocked peptides L1 and L2 (cf. table 1) from cytoplasmic aldehyde dehydrogenase. Remaining peaks in the figure correspond to the non-acetylated peptides from a larger tryptic peptide. Structures are shown adjacent to each peak.

agree for the aldehyde dehydrogenase peptides studied by both methods. They also agree with the tentative assignment from chemical analysis (table 1). In the case of peptide P1 from sorbitol dehydrogenase analyzed only by direct inlet probe, the results from mass spectrometry also agree with the data from chemical analysis (cf. fig.2 and table 1), as well as with results from a synthetic replicate of the structure.

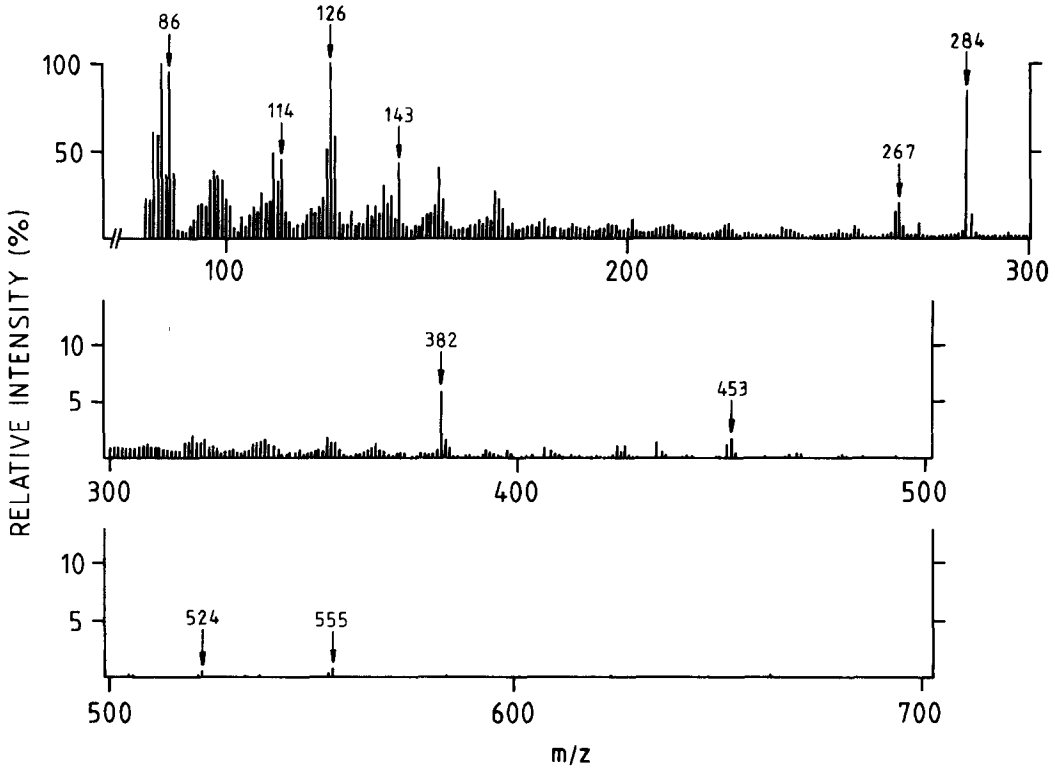
Fig.2. Results of mass spectrometric analyses of the derivative of peptide P1 of sorbitol dehydrogenase (top panels) and peptide L2 of aldehyde dehydrogenase (bottom panel). The peptides were purified by reverse-phase HPLC, for P1 from a digest with staphylococcal Glu-specific protease [11], and for L2 from a limited acid hydrolysis (fig.1). Mass peaks of special importance in the identification of the fragments are indicated by mass numbers and the appropriate fragment identifications are for P1: 86 (Ac-Ala-minus CO), 114 (Ac-Ala-), 126 ( $COOCH_3$), 143 ($Glu(OMe)_2$ minus CH_3OH), 267 (-Lys-Pro-), 284 (Ac-Ala-Lys-), 382 (Ac-Ala-Lys-Pro-), 453 (Ac-Ala-Lys-Pro-Ala-), 524 (Ac-Ala-Lys-Pro-

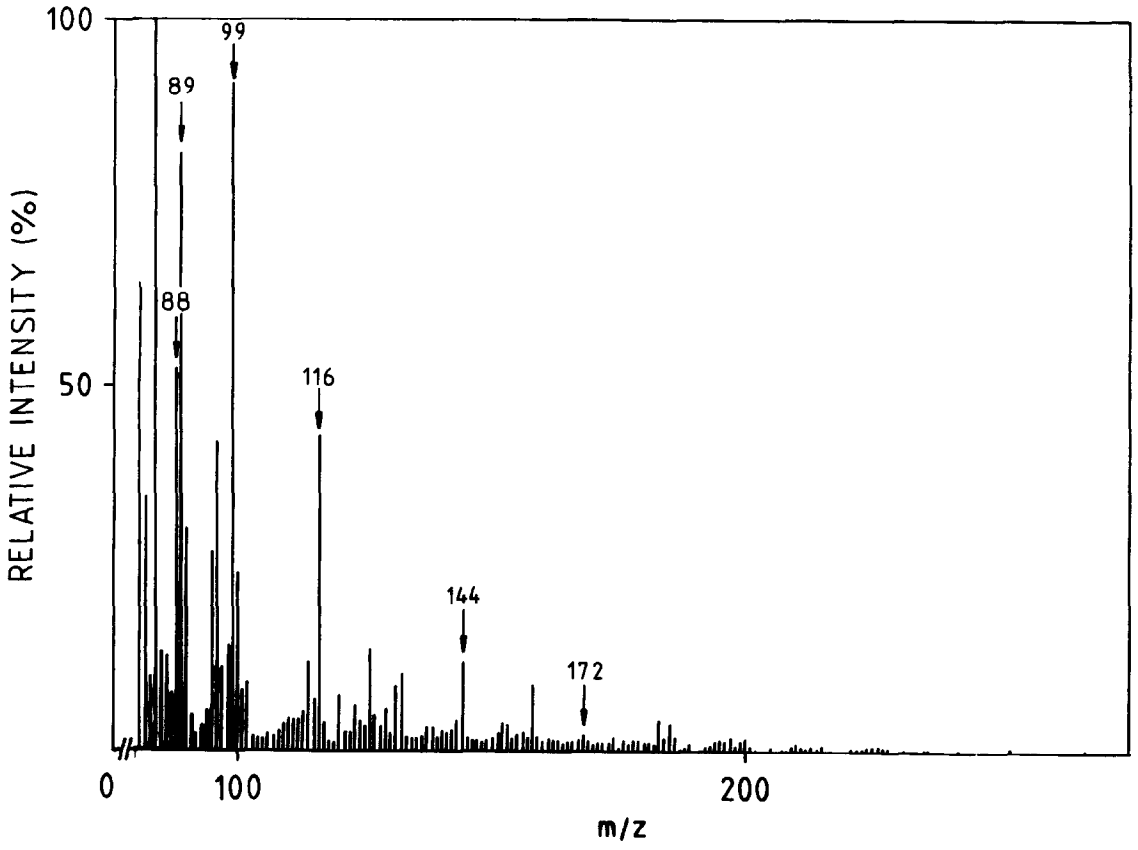
Ac Ac Ac Ac Ac

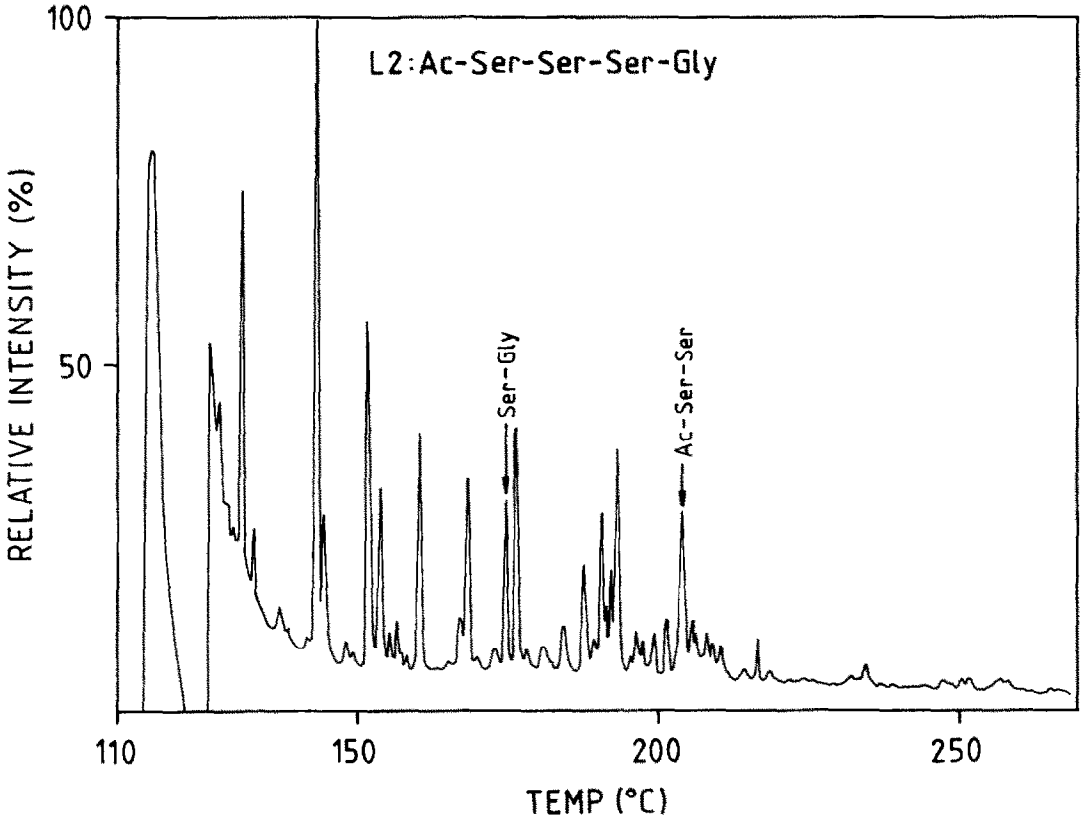
Ala-Ala) and for L2: 88 (Gly-OMe minus H), 89 (Gly-OMe), 99 ($CH_3CONHCH=C=O$ from N-terminal Ser), 116 ($O=\overset{+}{C}$-NH-CH_2-$COOCH_3$ from C-terminal Gly), 144 (Ac-Ser minus CO), 172 (Ac-Ser-). Fragment patterns agree with

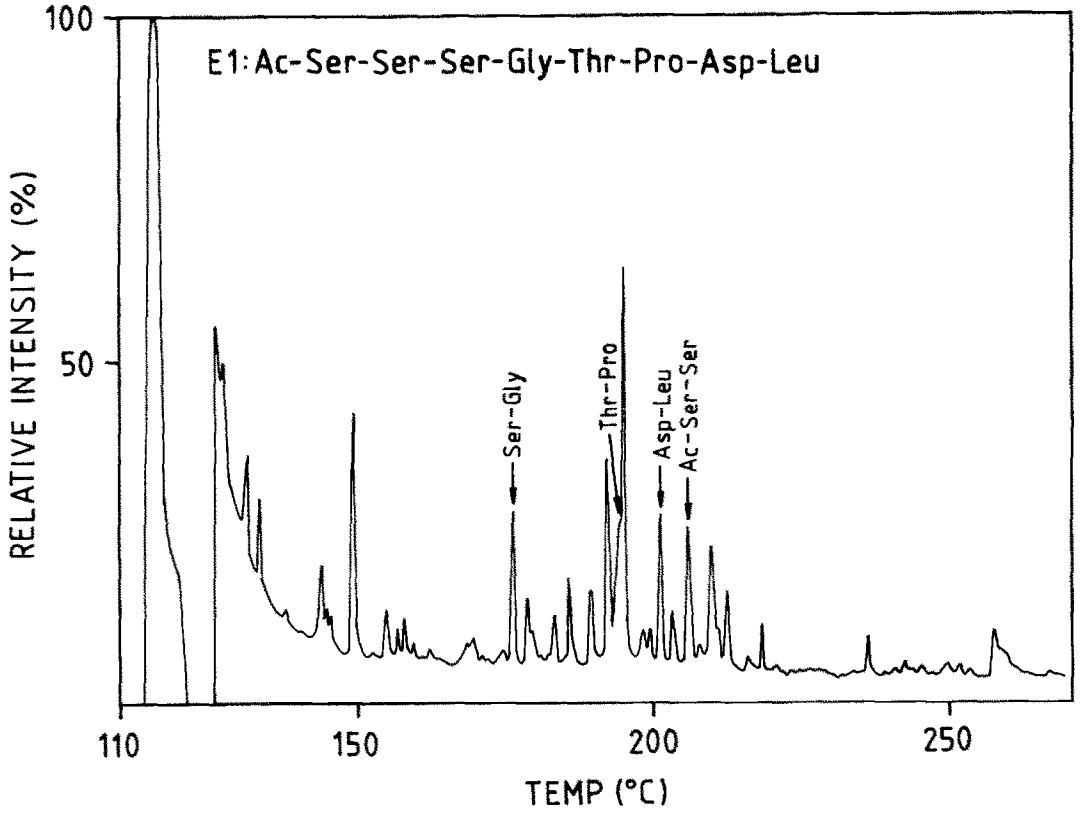
Ac Ac

results from analysis of synthetic peptides with structures corresponding to those of P1 and L2.


86
126
284
114
143
267
382
453
524
555
100
50
10
5
RELATIVE INTENSITY (%)
100
200
300
400
500
600
700
m/z


99
89
88
116
144
172
100
50
0
RELATIVE INTENSITY (%)
100
200
m/z


L2:Ac-Ser-Ser-Ser-Gly
RELATIVE INTENSITY (%)
100
50
Ser-Gly
Ac-Ser-Ser
110
150
200
250
TEMP (°C)


E1:Ac-Ser-Ser-Ser-Gly-Thr-Pro-Asp-Leu
RELATIVE INTENSITY (%)
100
50
Ser-Gly
Thr-Pro
Asp-Leu
Ac-Ser-Ser
110
150
200
250
TEMP (°C)

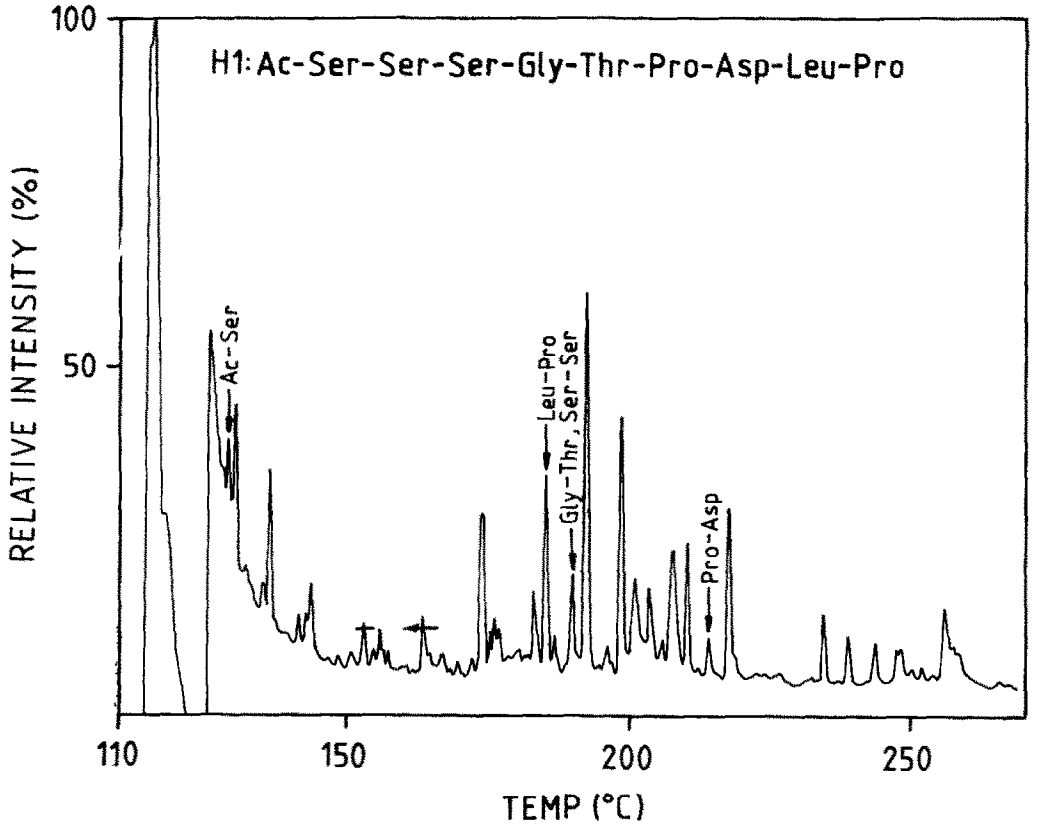


Fig.3. Results of gas chromatography-mass spectrometry of peptides L2, E1 and H1 of aldehyde dehydrogenase after treatment with dipeptidyl carboxypeptidase. Fragment identifications are shown at positions proven by analysis of synthetic peptides with these structures.

Table 2

Structures of N-terminal regions of dehydrogenases (DH) with acetyl-blocked subunits

Structure	Protein
Ac Ala Lys Pro Ala Ala Glu	Sheep sorbitol DH
Ac Ser Ser Ser Gly Thr Pro	Human cytoplasmic aldehyde DH
Ac Ser Phe Thr Leu Thr Asn	*Drosophila* alcohol DH
Ac Ser Thr Ala Gly Lys Val	Horse, rat, human alcohol DH
Ac Ser Ile Pro Glu Thr Gln	Yeast alcohol DH
Ac Ala Thr Leu Lys Glu Lys	Chicken lactate DH B
Ac Ser Leu Lys Asp His Leu	Chicken lactate DH A
Ac Thr Ala Leu Lys Asp Lys	Dogfish lactate DH A
Ac Ala Thr Leu Lys Asp Gln	Pig lactate DH A
Ac Ala Thr Leu Lys Glu Lys	Pig lactate DH B
Ac Ser Thr Val Lys Glu Glu	Mouse, rat lactate DH C
Ac Ser Lys Ile Gly Ile Asp	Lobster glyceraldehyde-3-phosphate DH
Ac Ala Gly Lys Lys Val Cys	Rabbit glycerol-3-phosphate DH
Ac Ala Gln Ala Asp Leu Ile	Sheep 6-phosphogluconate DH
Ac Ser Asn Leu Pro Ser Glu	*Neurospora* glutamate DH
Ac (Ser Glx Pro) Ile Arg Val	Pig cytoplasmic malate DH

The 6 first residues are listed. Sorbitol and aldehyde dehydrogenases are from this work, remaining structures are assembled from [1,13–18]

4. DISCUSSION

4.1. *N-terminal structures*

The results establish that the N-terminal residues are acetylated in sorbitol dehydrogenase and aldehyde dehydrogenase. In both cases, the amino acid sequences are also fully established for the blocked N-terminal regions, all results being consistent. The acetyl group is directly discernible from the ordinary mass spectrometric fragmentation (fig.2) and from the results with dipeptidyl carboxypeptidase degradation (fig.3). Sequences obtained also agree with the total compositions (table 1). Although the pretreatment in the mass spectrometric analysis involves acetylation (cf. section 2), the acetyl group cannot be derived from that treatment, because the native peptides are already N-terminally blocked (table 1), and because the acetyl group is also detectable in the samples from the dipeptidyl carboxypeptidase degradation that does not have an acetylation step (fig.3). Furthermore, even in the mass spectrometric analysis, any other, original, acyl group would have been readily identified, because of different m/z values for the fragments.

With the presently determined structures, aldehyde dehydrogenase and sorbitol dehydrogenase are added to the list of known acetylated dehydrogenases. They have blocked N-terminal residues that are typical of acetylated proteins in general, one with acetyl-Ser (aldehyde dehydrogenase) and one with acetyl-Ala (sorbitol dehydrogenase). In fact, the acetylated nature as well as the residue actually acetylated could have been predicted already from the compositions of the peptides (table 1) and some of the common properties of acetylated proteins [1]. A list of several known acetylated structures in dehydrogenases (assembled from [1,13–18] and this work) is shown in table 2. Although these structures do not fully display all properties previously shown for acetyl-blocked proteins in general, they further illustrate that acetylated proteins frequently start with serine or alanine, have over-representation of charged residues at about position 4 or 5, and often have other characteristic features [1].

4.2. *Methodological aspects*

The usefulness of the dipeptidyl carboxypeptidase method in combination with the specificity of different proteases in peptide generation is well illustrated. Since peptides E1 and H1 differ by only one residue in length, the two sets of fragmentation patterns from dipeptidyl carboxypeptidase digestions directly overlap to give the entire structure of the N-terminal region in aldehyde dehydrogenase. Finally, it may be noted that although the mass spectra obtained with both the gas chromatographic-mass spectrometric method and the direct inlet method have artifactual peaks, in part due to the presence of serine and other residues non-optimal for clean mass spectra, the results are clearly interpretable in combination with the total compositions, thus allowing conclusive assignments.

ACKNOWLEDGEMENTS

Support by the Swedish Medical Research Council (project 13X-3532) is acknowledged, as well as a fellowship from USPHS (AA 05150) to J.H., and travel grants from the Swedish Medical Research Council to J.J. (project 13V-6471).

REFERENCES

[1] Jörnvall, H. (1975) J. Theor. Biol. 55, 1–12.
[2] Driessen, H.P.C., Ramaekers, F.C.S., Vree Egberts, W.T.M., Dodemont, H.J., De Jong, W.W., Tesser, G.I. and Bloemendal, H. (1983) Eur. J. Biochem. 136, 403–406.
[3] Sanborn, B.M., Brummel, M.C., Steginк, L.D. and Vestling, C.S. (1970) Biochim. Biophys. Acta 221, 125–128.
[4] Brummel, M.C., Sanborn, B.M. and Steginк, L.D. (1971) Arch. Biochem. Biophys. 143, 330–335.
[5] Davidson, B.E., Sajgó, M., Noller, H.F. and Harris, J.I. (1967) Nature 216, 1181–1185.
[6] Jörnvall, H. (1970) Eur. J. Biochem. 14, 521–534.
[7] Jörnvall, H., Fairwell, T., Kratofil, P. and Wills, C. (1980) FEBS Lett. 111, 214–218.
[8] Jeffery, J., Cederlund, E. and Jörnvall, H. (1984) Eur. J. Biochem., in press.
[9] Hempel, J., Von Bahr-Lindström, H. and Jörnvall, H. (1984) Eur. J. Biochem., in press.
[10] Krutzsch, H.C. (1983) Methods Enzymol. 93, 5290–5296.
[11] Jörnvall, H., Carlquist, M. and Jeffery, J. (1983) Pharm. Biochem. Behav. 18, Suppl.1, 67–71.

[12] Von Bahr-Lindström, H., Galante, Y.M., Persson, M. and Jörnvall, H. (1983) Eur. J. Biochem. 134, 145–150.
[13] Thatcher, D.R. (1980) Biochem. J. 187, 875–883.
[14] Torff, H.-J., Becker, D. and Schwarzwälder, J. (1977) in: Pyridine Nucleotide-Dependent Dehydrogenases (Sund, H. ed) pp.31–39, De Gruyter, Berlin.
[15] Li, S.S.-L., Fitch, W.M., Pan, Y.-C.E. and Sharief, F.S. (1983) J. Biol. Chem. 258, 7029–7032.
[16] Otto, J., Argos, P. and Rossmann, M.G. (1980) Eur. J. Biochem. 109, 325–330.
[17] Carne, A. and Walker, J.E. (1983) J. Biol. Chem. 258, 12895–12906.
[18] Holder, A.A., Wootton, J.C., Baron, A.J., Chambers, G.K. and Fincham, J.R.S. (1975) Biochem. J. 149, 757–773.