Multiplicity of N-terminal structures of medium-chain alcohol dehydrogenases
Mass-spectrometric analysis of plant, lower vertebrate and higher vertebrate class I, II, and III forms of the enzyme

Lars Hjelmqvist[a], Murray Hackett[b], Jawed Shafqat[a], Olle Danielsson[a], Junko Iida[b], Ronald C. Hendrickson[b], Hanspeter Michel[b], Jeffrey Shabanowitz[b], Donald F. Hunt[b,c], Hans Jörnvall[a,*]

[a]*Department of Medical Biochemistry and Biophysics, Karolinska Institutet, S-171 77 Stockholm, Sweden*
[b]*Department of Chemistry, University of Virginia, Charlottesville, VI 22901, USA*
[c]*Department of Pathology, University of Virginia, Charlottesville, VI 22901, USA*

Received 16 May 1995

Abstract Ten different alcohol dehydrogenases, representing several classes of the enzyme and a wide spread of organisms, were analyzed for patterns of N-terminal structures utilizing a combination of conventional and mass spectrometric peptide analysis. Results show all forms to be N-terminally acetylated and allow comparisons of now 40 such alcohol dehydrogenases covering a large span of forms and origins. Patterns illustrate roles of acetylation in proteins in general, define special importance of the class I N-terminal acetylation, and distinguish separate acetylated structures for all classes, as well as a common alcohol dehydrogenase motif.

Key words: Enzyme family; Alcohol dehydrogenase; Mass spectrometry; N-terminus; Acetyl group

1. Introduction

Alcohol dehydrogenases constitute an enzyme system of at least six classes in vertebrates [1] and additional forms of separate duplications in other lines [2,3]. Several of the classes have been structurally characterized in a wide spread of species, class I (the classical ethanol-active liver enzyme) from mammals to bony fish, and class III (glutathione-dependent formaldehyde dehydrogenase) from mammals to prokaryotes, while class II and remaining classes thus far have been structurally characterized only in single species, mainly human [2]. Additional isozyme heterogeneity has also been analyzed.

All or most of these eukaryotic enzymes are N-terminally blocked by acetylation. The residue acetylated early appeared to form a pattern, with acetyl-Ser being a common N-terminus of class I, acetyl-Ala of class III, and perhaps acetyl-Gly of class II [4]. In addition, they start at slightly differently shifted positions and exhibit further structural characteristics in the N-terminal regions. The pattern has even been utilized to detect novel multiplicity in enzymes that were suspected to reflect the duplicatory class origins [4].

In our efforts to trace these enzymes and their functional relationships, we have purified and structurally analyzed a number of novel alcohol dehydrogenases from vertebrate and other life forms. Throughout, we have characterized the N-termini by mass spectrometry, allowing rapid and sensitive assignment of the N-terminal structures, confirming the usefulness of predictions based on these terminal structures, in agreement with data thus obtained for other protein systems [5]. In this study, we report on the analysis of ten different alcohol dehydrogenase N-terminal peptides from three classes of the enzyme in seven species, ranging from plant alcohol dehydrogenase, through cyclostome alcohol dehydrogenase, to vertebrate alcohol dehydrogenases of classes I-III. Results are consistent in establishing distinct N-terminal patterns.

2. Materials and methods

Alcohol dehydrogenase from livers of ostrich, kiwi, the lizard *Uromastix hardwickii* (higher vertebrates), cod (bony fish), dogfish (cartilaginous fish) and Atlantic hagfish (cyclostome) was purified essentially as described for the class I (when present) and III enzymes [4,6,7]. In addition, two novel alcohol dehydrogenases were detected and purified, one constituting a class II form from ostrich, detected by separation of a second ethanol-active fraction upon DEAE chromatography, the other constituting a pea class III form obtained by purification in a manner similar to that for animal class III forms [7]. Final purifications were in the order of 100-fold for the class I and II enzymes, and 200–2,000-fold for the class III enzyme. Preparations were checked by SDS/polyacrylamide gel electrophoresis and then carboxymethylated, before submission in separate batches to proteolytic treatments with Lys- and Glu-specific proteases, all as described [6,7]. N-terminal peptides were identified from their approximately known HPLC elution positions, derived from largely conserved Lys~5 and Glu~16, respectively, in alcohol dehydrogenases [6,7], and were checked for identity by amino acid compositions and attempts at direct sequencer degradations verifying the blocked N-terminus.

In several cases, the blocked peptides were also submitted to sequencer analysis after they had been deblocked in TFA/methanol [8]. Molecular masses were determined by either fast atom bombardment on a tandem quadrupole Fourier transform instrument [9] or by electrospray ionization on a triple quadrupole mass spectrometer [10]. Peptide sequences determined by mass spectrometry were deduced from collision activation dissociation (CAD) spectra recorded on a triple quadrupole mass spectrometer [11].

3. Results

Ten different alcohol dehydrogenases of three classes (I, II, and III) and representing six different life forms (avian, reptil-

*Corresponding author. Fax: (46) (8) 33 74 62.

SSDI 0014-5793(95)00572-2

ian, bony fish, cartilaginous fish, cyclostome, and plant) were purified for analysis of the N-terminal structural patterns. Attempts at direct sequencer analysis failed to yield productive results, suggesting a uniform presence of N-terminal blocking groups. Hence, peptides derived from the N-terminal segments were prepared and analyzed. Generally, the proteins were then reduced and carboxymethylated (to stabilize a cysteine residue frequently present at a position corresponding to Cys-9 of class I alcohol dehydrogenases [4]), and were then treated in separate batches with Glu-specific and Lys-specific proteases, respectively. The former often gives a 16-residue peptide, the latter generally a 5- or 6-residue peptide, eluting midpoint and early, respectively, in reverse phase separation of the proteolytic digest [4,6]. Correct purifications were checked by total compositions, sequencer analysis (giving no result, compatible with the N-terminal derivation), and in some cases also by deblocking with TFA in methanol and subsequent sequencer analysis [8,12]. Thus identified, the N-terminal structures were submitted to mass spectrometric analysis. Frequently, data were also obtained from traditional sequencer analysis of other fragments covering the N-terminal regions.

For mass spectrometry, the peptides were first submitted to fast atom bombardment or electrospray ionization mass spectrometry, establishing the total mass values as shown in Table 1. When further analyzed for sequence assignments, collision-activated dissociation was utilized, frequently after acetylation, methyl esterification, and further subdigestion, to establish all structures. Results obtained are summarized (top) and detailed (bottom) in Table 1. For sequencer degradations, results are also given in Table 1. Complementary data were further obtained by sequencer analysis after chemical deblocking and by ongoing cDNA analyses in a few cases.

Results by the different methods are consistent, reproducible, and establishing the structures shown in Table 1. Notably, all peptides are N-terminally acetylated, and the most successful hydrolytic deblocking of the N-terminus was obtained with the Hagfish enzyme (Table 1) which, significantly, has acetylated Ser, in contrast to the more stable acetylated Ala present

Table 1
Results of combined sequence analysis of 10 different acetylated alcohol dehydrogenase N-terminal peptides

No.	Species	Structure
I	Ostrich ADH I	AcSTAGKVIKCKAAVLWEPKKPFSIEE
II	Ostrich ADH II	AcTTEGK
III	Kiwi ADH I	AcSTAGKVIKCKAAVLWEPKKPFSIEE
IV	*Uromastix* ADH III	AcASGVIKCKAAVAWE
V	Cod ADH III *l*	AcATAGQVIKCKAAVAWE
VI	Cod ADH III *h*	AcATVGKTIRCRAAVAWE
VII	Dogfish ADH III a	AcAGTVGK
VIII	Dogfish ADH III b	AcAGTVGK
IX	Hagfish ADH III	AcSKMDGQVIHCKAAVAWE
X	Pea ADH III	AcATQGQVITCKAAVAWEPNKPLTIEDVE

Data obtained by mass spectrometry

No.	Parent ion	Molecular mass		Sequence identification	
		Experimental[a]	Calculated[a]	CAD	Edman degradation (this or other peptide)
I	$(M + H)^+$	504	504.5	Ac(1–5)	–
	$(M + 3H)^{+3}$	1,489	1,488.8	Ac(1–14)[b]	6–8 and 11–14
	$(M + 3H)^{+3}$	1,389	1,389.6	15–25	15–25
	$(M + 4H)^{+4}$	2,861[b]	2,860.4[b]	–	6–8 and 11–25
II	$(M + H)^+$	576	576.6	Ac(1–5)	4–5
III	$(M + H)^+$	2,861[b]	2,860.4[b]	–	1–10 (cDNA) and 11–25
IV	$(M + 2H)^{+2}$	1,532	1,532.8	Ac(1–14)[b]	9–14
V	$(M + 2H)^{+2}$	830	829.0	Ac(1–8)	–
	$(M + H)^+$	646	645.7	11–16	11–16
	$(M + 2H)^{+2}$	1,746	1,746.0	Ac(1–16)[b]	11–16
VI	$(M + H)^+$	516	516.6	Ac(1–5)	–
	$(M + H)^+$	646	645.7	11–16	11–16
	$(M + 2H)^{+2}$	964	963.1	9–16[b]	9–16
	$(M + 2H)^{+2}$	1,832	1,832.1	1–16[b]	6–16
VII	$(M + H)^+$	574	573.7	Ac(1–6)	–
VIII	$(M + H)^+$	574	573.7	Ac(1–6)	–
IX	$(M + 2H)^{+2}$	1,973	1,973.3	Ac(1–17)[b]	1–17
X	$(M + H)^+$	1,148	1,148.3	Ac(1–10)[b]	–
	$(M + 2H)^{+2}$	1,880	1,882.1	11–27	11–27
	$(M + 2H)^{+2}$	3,014	3,012.4	Ac(1–27)[b]	11–27

Top panel shows the 10 structures from 7 species, bottom panel the mass and sequencer proofs for the structures determined. Regarding peptide designations, numbers joined by a hyphen indicate positions, and Ac indicates acetylation. [a]Average mass, experimentally determined masses were recorded at less than unit resolution and are rounded to the nearest whole number; [b]contains carboxymethylated Cys.

in most of the other alcohol dehydrogenase peptides analyzed (Table 1).

4. Discussion

The nature of the N-terminal segments of ten novel alcohol dehydrogenase forms has been determined (Table 1), defining the acetylation sites and N-terminal patterns for a total of now no less than 40 separate alcohol dehydrogenase structures, representing a large spread of organisms covering class I, II, III, and IV alcohol dehydrogenases (Table 2). Other alcohol dehydrogenases are also known, but excluded from the table because of unknown acetylation status (most plant alcohol dehydrogenases, derived from cDNA structures without protein analysis), non-acetylation (*E. coli*), or derivation from outside the ordinary class system (yeast alcohol dehydrogenase of the traditional type). The patterns obtained are quite informative and establish novel patterns of four different types. First, it is clear that the four analyzed classes of alcohol dehydrogenase (I-IV) have separate and distinct N-terminal structures, meaning that the classes can be easily identified by analysis of these single segments. The N-terminal class I and III segments are very different, except in one species, the cod, where they differ only marginally in the terminal parts. This N-terminal similarity for the cod class I and III enzymes, in spite of the general non-identity between class I and III forms, is probably another indication of the mixed properties of the cod class I enzyme [6]. Except for this case, consistent with the origin of bony fish from the time of the class I/III duplication, the acetylated N-terminal residue is largely class-specific. Thus, most class I enzymes have acetylated Ser, most class II enzymes acetylated Gly, and most class III enzymes acetylated Ala.

Second, independent of class, a consensus sequence may be discerned in many cases, XTXGKXIKCK, indicating considerable sequence conservation over this wide spread of enzymes and species. It defines an alcohol dehydrogenase motif often present and likely to represent an ancestral pattern.

Third, and highly surprising, the overall nature of the classes is reversed between the N-terminal segments and the rest of the molecules. Thus, regarding entire chains, class I is generally 'variable' and class III 'constant' in structure [7], whereas regarding the N-terminal segments, this class difference is now found to be reversed, with class I being the one with the most constant segment (only single exchanges at positions 1 and 3, cf. Table 2, top), and class III the one with the most variable segment (at all positions except Ile-7 and Cys-9, Table 2). Furthermore, not only the exchanges, but also the exact start positions, are variable for the N-terminal segments of class III (Table 2). The reason for this unexpected reversal of class variability between the N-terminal segments and the entire protein chains is not known, but suggests an extra importance of the N-terminal acetylation in class I (versus class III), forcing greater restrictions on this part of class I than on other parts of the molecule. Since N-terminal acetylation is likely to be involved in the regulation of protein degradation [14], it appears significant that it is the class with the least variable N-terminal structure (class I) that has the most typical [14] such structure, with Ser-1, and several Lys and branched-chain residues at positions 5–10, probably dictating the acetylation and hence producing correct susceptibility toward overall protein degradation.

Fourth, also details of the pattern of acetylation appear to be significant. Not only are all the analyzed structures acetylated, with Ser, Ala, Gly, and Thr at position 1, following the rules for protein acetylation in general [14], but also the residues at the position corresponding to position 3 of the class I enzymes have a special distribution. Although, this position is variable, with Ala and Val as the most frequent alternatives in class III (4 Ala cases, and 4 Val cases, Table 2), Val never occupies this position when it constitutes the actual N-terminus as in the mammalian class III forms (top of the class III alternatives in Table 2). Thus, when truncated structures do occur to yield 'position 3' as actual N-terminus of the mature protein, all cases keep the N-terminal residue as Ala (or Ser) but never Val, further establishing the lack of branched-chain residues at the position acetylated.

In summary, N-terminal alcohol dehydrogenase structures define distinct rules in relation to both the enzyme pattern and the role of acetylation. The structures are largely identical in

Table 2
N-terminally acetylated structures of 40 alcohol dehydrogenases

Class I	
Human α β γ	STAGKVIKCK
Baboon β*, Rhesus α*	STAGKVIKCK
Horse E S	STAGKVIKCK
Rabbit*	STAGKVIKCK
Rat, Mouse*, Deer mouse*	STAGKVIKCK
Chicken	**ST**VGKVIKCK
Quail, Ostrich, Kiwi	STAGKVIKCK
Alligator	STAGKVIKCK
Uromastix a	**G**TAGKVIKCK
Uromastix b	STAGKVIKCK
Frog	**A**TAGKVIKCK
Cod	**AT**VGKVIKCK
Class II	
Human	GT**K**GKVIKCK
Rat*	GT**Q**GKVI**T**CK
Ostrich	**TTE**GKVIKCK
Class III	
Human	A**NE**VIKCK
Horse	**SAE**VIKCK
Rat, Mouse*	A**NQ**VI**R**CK
Uromastix	A**SG**VIKCK
Cod *l*	ATAG**Q**VIKCK
Cod *h*	ATVGK**TIRCR**
Dogfish a b	A**G**TVGKXXXXX
Hagfish	S**KMDGQ**VI**H**CK
Octopus	**TDATG**K**P**IK**CM**
Drosophila	SAT**E**GKVI**T**CK
*Candida**	S**ES**TVGK**P**I**T**CK
Saccharomyces	SAATVGK**P**IKC**I**
Pea	AT**QGQ**VI**T**CK
Class IV	
Human*	**GTA**GKVIKCK
Rat	**SNRV**GKVIKCK

In all cases, the residue at position 1 is deduced to be acetylated, as proven by mass spectrometry or other direct identification (most cases), presence of blocked N-terminus, or identity (marked by asterisk) with known acetylated forms. Bold type letters indicate residues deviating from the common alternatives in each class. Two further classes of mammalian alcohol dehydrogenase exist [1] but are not listed since they have thus far been little analyzed and not yet proven to be acetylated, although they are likely to be so. Structures from data bases and this study, complemented with the rat class II structure [13] and acetylation data for *Saccharomyces* class III (Fernández et al., unpublished).

class I, to a large part different in class III, and with class II and IV occupying intermediate positions.

Acknowledgements: This study was supported by grants from the Swedish Medical Research Council (13X-3532), Peptech (Australia), Karolinska Institutet, and NIH (GM37537).

References

[1] Jörnvall, H. and Höög, J.-O. (1995) Alcohol and Alcoholism, in press.
[2] Jörnvall, H. (1994) in: Toward a Molecular Basis of Alcohol Use and Abuse (Jansson, B., Jörnvall, H., Rydberg, U., Terenius, L. and Vallee, B.L. Eds.) pp. 221–229, Birkhäuser Verlag, Basel.
[3] van Ophem, P.W. and Duine, J.A. (1993) in: Enzymology and Molecular Biology of Carbonyl Metabolism 4 (Weiner, H., Crabb, D.W. and Flynn, T.G. Eds.) pp. 605–620, Plenum Press, New York.
[4] Hjelmqvist, L., Ericsson, M., Shafqat, J., Carlquist, M., Siddiqi, A.R., Höög, J.-O. and Jörnvall, H. (1992) FEBS Lett. 298, 297–300.
[5] Flinta, C., Persson, B., Jörnvall, H. and von Heijne, G. (1986) Eur. J. Biochem. 154, 193–196.
[6] Danielsson, O., Eklund, H. and Jörnvall, H. (1992) Biochemistry 31, 3751–3759.
[7] Danielsson, O., Atrian, S., Luque, T., Hjelmqvist, L., Gonzàlez-Duarte, R. and Jörnvall, H. (1994) Proc. Natl. Acad. Sci. USA 91, 4980–4984.
[8] Gheorghe, M.T. and Bergman, T. (1995) in: Methods in Protein Structure Analysis (Atassi, M.Z. and Appella, E. Eds.) Plenum, in press.
[9] Hunt, D.F., Shabanowitz, J., Yates III, J.R., Zhu, N.Z., Russell, D.H. and Castro, M.E. (1987) Proc. Natl. Acad. Sci. USA 84, 620–623.
[10] Hunt, D.F., Henderson, R.A., Shabanowitz, J., Sakaguchi, K., Michel, H., Sevilir, N., Cox, A.L., Appella, E. and Engelhard, V.H. (1992) Science 255, 1261–1263.
[11] Cox, A.L., Skipper, J., Chen, Y., Henderson, R.A., Darrow, T.L., Shabanowitz, J., Engelhard, V.H., Hunt, D.F. and Slingluff Jr., C.L. (1994) Science 264, 716–719.
[12] Parés, X., Cederlund, E., Moreno, A., Hjelmqvist, L., Farrés, J. and Jörnvall, H. (1994) Proc. Natl. Acad. Sci. USA, 91, 1893–1897.
[13] Höög, J.-O. (1995) FEBS Lett., in press.
[14] Persson, B., Flinta, C., von Heijne, G. and Jörnvall, H. (1985) Eur. J. Biochem. 152, 523–527.