The EMBO Journal vol.3 no.6 pp.1307–1310, 1984

Correlation of exons with structural domains in alcohol dehydrogenase

Carl-Ivar Brändén*, Hans Eklund, Christian Cambillau and Anthony J. Pryor[1]

Department of Chemistry and Molecular Biology, Swedish University of Agricultural Sciences, S-750 07 Uppsala, Sweden, and [1]Division of Plant Industry, CSIRO, Canberra, ACT 2601, Australia

*To whom reprint requests should be sent
Communicated by C.-I. Brändén

The intron/exon arrangement in the gene sequence of maize alcohol dehydrogenase has been compared to the three dimensional structure of liver alcohol dehydrogenase. The coenzyme binding domain is separated from the catalytic domain by introns four and nine. Intron seven separates the coenzyme binding domain into two structurally similar mononucleotide binding units. The first of these units is divided by introns five and six into three structurally similar $\alpha\beta$ modules. Implications of these results for protein evolution is discussed. All splice junctions map close to or at the surface of the domains, and several of these cannot be identified by distance maps.
Key words: alcohol dehydrogenase/intron/3-D structure/evolution

Introduction

The subunits of NAD^+-dependent dehydrogenases are divided into two structurally and functionally different domains (Rossmann *et al.*, 1975). One of these binds the co-enzyme and the other provides the residues necessary for substrate binding and catalysis. The NAD^+-binding domains of lactate dehydrogenase (Adams *et al.*, 1970), alcohol dehydrogenase (Brändén *et al.*, 1973) and glyceraldehyde-3-phosphate dehydrogenase (Buehner *et al.*, 1973) are all similar in structure and bind the co-enzyme in a similar way. The catalytic domains, on the other hand, have quite different structures. In spite of the similarities in structure there is no sequence homology between these enzymes.

When these relationships were discovered ~10 years ago it was suggested (Ohlsson *et al.*, 1974; Rossmann *et al.*, 1974) that they reflected evolutionary relationships and that these and other complex enzymes have evolved by reassorting and joining genes that coded for simple ancestral polypeptide chains. At that time no plausible genetic mechanism was known that could easily account for such rearrangements. Some years later after the discovery of intron-exon arrangements in eucaryotic genes Gilbert (1978) proposed that exons might correspond to units of protein function and that this feature would be a mechanism for increasing the rate of evolution.

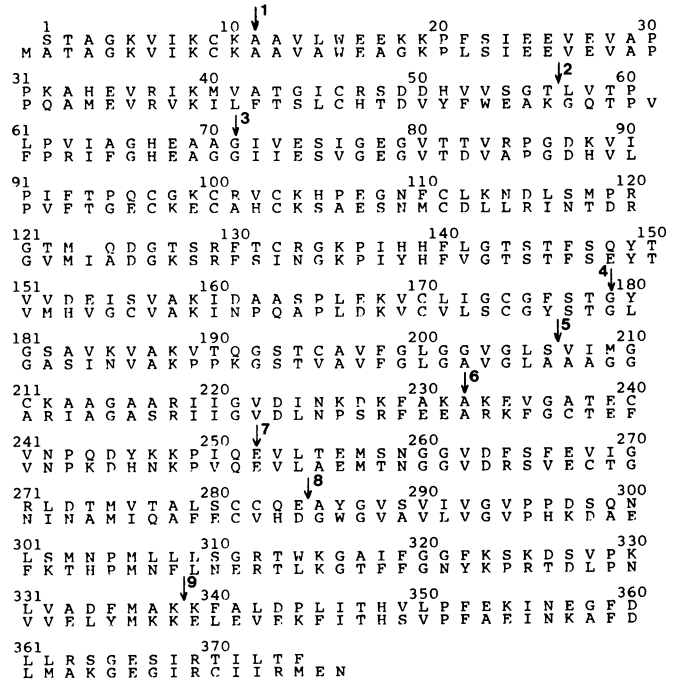
Here we compare the intron-exon arrangement in the recently determined gene sequence of maize alcohol dehydrogenase (Dennis *et al.*, 1984) with the three dimensional structure of liver alcohol dehydrogenase (Eklund *et al.*, 1976) to examine if there is any clear relation between the nine intron positions in the gene and structural or functional modules in the protein.

Results and Discussion

Both liver and maize alcohol dehydrogenase are dimeric enzymes, each composed of two polypeptide chains of 374 and 379 amino acids respectively. They exhibit 52% sequence identity using the alignment shown in Table I. The amino acid sequence of the maize subunit has been deduced by Dennis *et al.* (1984) from the base sequence of the maize *Adh*1 gene. The high sequence identity is a strong indication that the three dimensional structures of the subunits are very similar. Furthermore, an analysis of the sequence differences (unpublished results) has shown that all structurally important residues are either identical or conservatively substituted. We can thus safely assume that the polypeptide fold and subunit arrangement of maize alcohol dehydrogenase are the same as in the liver enzyme. A schematic diagram of the polypeptide fold with the intron positions is shown in Figure 1.

The intron positions of the maize gene occur within the triplets coding for residues 11, 71, 179 and 232 and after the triplets coding for residues 56, 206, 252, 284 and 338 (Dennis *et al.*, 1984). All numbers refer to the horse liver alcohol dehydrogenase sequence. The positions along the polypeptide

Table I. Comparison between liver and maize alcohol dehydrogenase sequences



Arrows show the positions of introns in the maize gene.
Line 1: horse liver alcohol dehydrogenase; line 2: maize alcohol dehydrogenase.

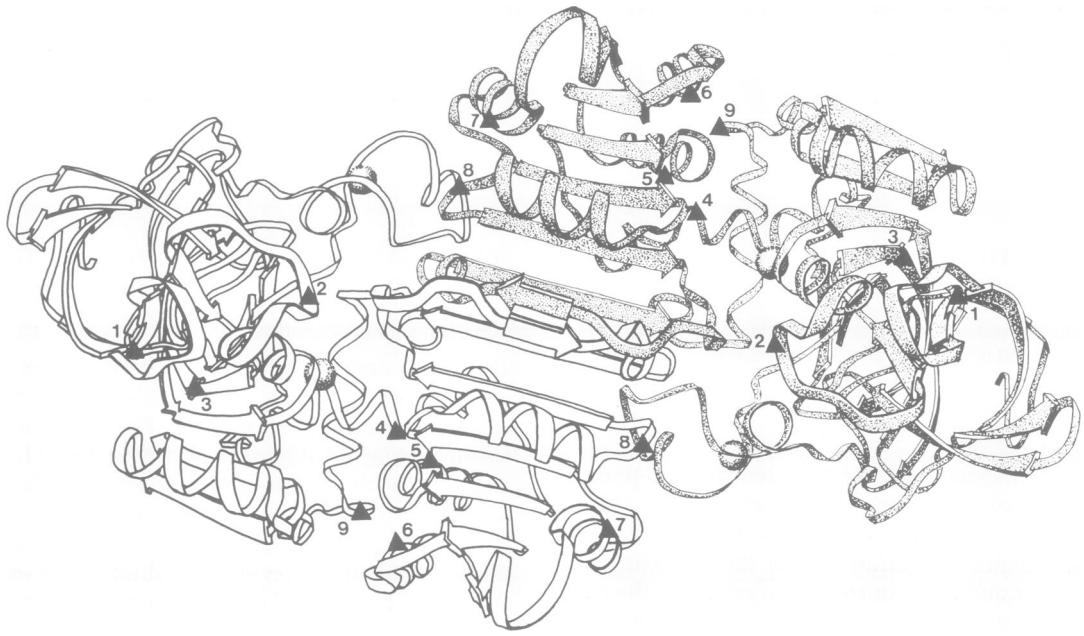


Fig. 1. Schematic diagram of the polypeptide fold of the alcohol dehydrogenase dimer. One subunit is shaded. The intron positions are marked and numbered from the amino end. (We are indebted to Bo Furugren for this drawing).

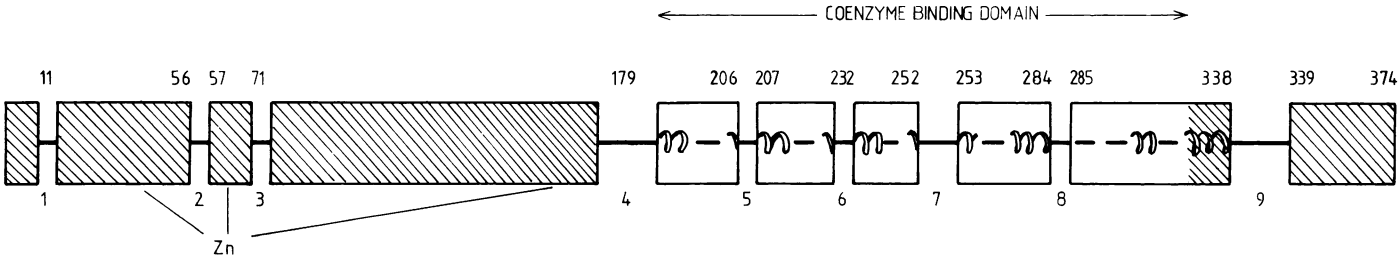


Fig. 2. Schematic representation of the polypeptide chain of alcohol dehydrogenase. Residue numbers refer to the sequence of the liver enzyme. Gaps represent intron positions in the maize gene. Residues belonging to the catalytic domain are shaded. Residues belonging to the helices and strands of pleated sheet in the co-enzyme binding domain are marked as well as the zinc ligands.

chain of the two domains and of the secondary structural elements in the co-enzyme binding domain are shown in Figure 2 as well as their relation to the intron-exon arrangement of the maize gene.

The domains are separated by introns

The catalytic domain of the liver subunit is built up from residues 1–174 and 318–374. The co-enzyme binding domain thus roughly comprises residues 175–317 with two covalent connections to the catalytic domain. From the physical separation of domains in the structure we would thus expect introns in the maize gene at approximately these positions if they separate the domains in the gene structure.

The co-enzyme binding domain is separated from the main part of the catalytic domain by intron four and from the extra carboxy terminal $\beta\alpha\beta$ module of the catalytic domain by intron nine. Intron four at postion 179 is in the expected region whereas intron nine at position 338 is 20 residues further along the chain. These 20 residues form a helix which connects the last parallel strand of the co-enzyme binding domain, βF, with the carboxy terminal structural unit of the catalytic domain.

In both lactate dehydrogenase and glyceraldehyde-3-phosphate dehydrogenase a helix of 15–20 residues connects βF of the co-enzyme binding domain with the catalytic domain (Rossmann *et al.*, 1975). If these domains are evolutionarily related to the corresponding domain in alcohol dehydrogenase, and if intron-exon arrangements reflect such relationship, we would expect that introns are present in the genes for these enzymes after this helix and not immediately after βF. This would correspond to positions 180 and 165 in lactate- and glyceraldehyde-3-phosphate dehydrogenase respectively.

The co-enzyme binding domain is separated by an intron into two mononucleotide binding domains

The co-enzyme binding domain is built up from six parallel strands of pleated sheet surrounded by helices on each side of the sheet. Rao and Rossmann (1973) showed that this domain could be divided into two roughly identical units each associated with a mononucleotide binding area. Intron seven after position 252 separates the domain into two such units.

The dinucleotide binding domain is more regular in alcohol dehydrogenase than in the other known dehydrogenases and consists of two $(\alpha\beta)_3$ units (Eklund *et al.*, 1976). Intron four is positioned after one turn in the first helix of the first of these units. Intron seven is positioned in the corresponding position in the second $(\alpha\beta)_3$ unit. Both exon five and exon eight thus start at structurally equivalent positions in these units.

The mononucleotide binding domain is a structural module

Table II. Accessible surface area calculated for residues at the splice junctions

Intron no.	Residues	Residue numbers	Accessible surface area (Å²)
1	Ala	10 **11** 12	61 **0** 0
2	Lys Ala	55 **56** **57** 58	20 **35** **17** 115
3	Gly	70 **71** 72	0 **0** 46
4	Gly	178 **179** 180	18 **0** 15
5	Ala Ala	205 **206** **207** 208	0 **0** **0** 0
6	Ala	231 **232** 233	60 **0** 106
7	Glu Val	251 **252** **253** 254	58 **83** **2** 0
8	Asp Gly	283 **284** **285** 286	1 **41** **6** 23
9	Lys Gly	337 **338** **339** 340	56 **183** **127** 20

Intron positions are represented by bold type numbers.

which has been observed in a number of different proteins (Brändén, 1980). The fact that these modules are here so clearly separated by an intron lends considerable strength to the theory that introns may separate structural modules in proteins.

The first mononucleotide binding domain is divided into three exons which form repeating structural units

The DNA region for the first mononucleotide binding domain, residues 179–252, contains three exons. Corresponding intron positions five and six are after residue 206 and at residue 232. Structurally these positions are approximately after the first turn of helices αB and αC. The three exons each comprise an αβ unit with some additional residues after the β-strand. Exons five, six and seven thus roughly correspond to αAβA, αBβB and αCβC, respectively.

Blake (1983) has recently attempted to produce a comprehensive view on the origin and role of exons in protein evolution. He assumes that the fundamental early coding units of 20–40 residues might have been potential supersecondary structures such as ββ, αα or αβ units. These could then readily be assembled to form larger peptides representing early protein molecules. One such early molecule must have been a nucleotide binding protein. Granted that the present intron-exon arrangement in proteins to some extent reflects early stages of gene arrangement the intron positions described here for the first mononucleotide binding unit strongly support this view on protein evolution. The fact that the yeast genes for alcohol dehydrogenase (Bennetzen and Hall, 1982; Russel *et al.*, 1983) contain no introns, supports the suggestions by Doolittle (1978) and Blake (1983) for intron loss occurring in prokaryotes and yeasts during evolution.

There are two different ways to combine αβ units into globular domains (Brändén, 1980). One is by forming open sheets such as the nucleotide binding domain, the other by forming closed barrels from eight consecutive αβ units as found in triose-phosphate isomerase (Banner *et al.*, 1975), pyruvate kinase (Stuart *et al.*, 1979) and glycolate oxidase (Lindqvist and Brändén, 1980). When the gene sequences of these enzymes are known it will be very interesting to see if the pattern observed here of repetitive αβ exon units will also be found for these barrel structures. The second mononucleotide binding domain in alcohol dehydrogenase contains only one intron which is positioned after residue 284 in the loop before strand βE. There is thus no similar correlation between exons and structural repeat of αβ units in this half of the co-enzyme binding domain.

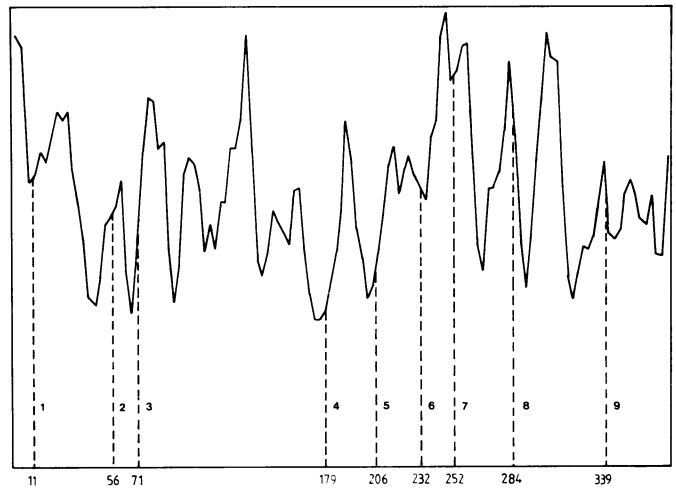


Fig. 3. Distance plot for horse liver alcohol dehydrogenase. The number of Cα-Cα distances larger than 27 Å is plotted against residue number. Intron positions in the maize gene are represented by vertical lines and numbered.

Splice junctions map close to or at the surface of the domains

Intron-exon junctions usually map to amino acids at the protein surface (Craik *et al.*, 1982). The situation is somewhat different in alcohol dehydrogenase, where five of the splice junctions are at residues which are non-accessible to solvent. Four of these are, however, close to the surface since a neighbouring residue is accessible (Table II). The fifth splice junction at residues 206–207 is deeply buried in the subunit. It is, however, situated in the interface region between the domains close to the surface of the separated co-enzyme binding domain. The amino acids that dominate at the splice junctions in maize alcohol dehydrogenase are Gly and Ala. Craik *et al.* (1982) found that extremely hydrophilic amino acids were over-represented on the 5′ side of the splice junctions. There are four examples of this type in alcohol dehydrogenase at residues 56, 252, 284 and 338.

Correlation of exonic regions with distance maps.

Distance maps are useful to localize separated structural elements (Liljas and Rossmann, 1974). For lysozyme and haemoglobin these have been correlated to exonic regions (Gô, 1981, 1983). In Figure 3 we have plotted the number of Cα-Cα distances larger than 27 Å *versus* residue number.

This is essentially a simplified representation of distance maps (Blake, 1983), where minima would correspond to the separation of structural elements according to Gô. However, due to the nature of these calculations they are also biased to show minima at central regions of the protein subunit. In alcohol dehydrogenase this is close to the active site between the domains. This is clearly seen in Figure 3, where some of the minima correspond to intron positions whereas others do not, but instead occur at active site residues. Furthermore, some intron positions such as intron numbers three and eight do not occur at minima and can thus not be identified in distance maps.

Conclusions

Previous work (Dennis *et al.*, 1984) recognized that the catalytic and co-enzyme binding domains of the alcohol dehydrogenase subunit are separated by intron sequences in the maize gene. Here we show that the co-enzyme binding domain in addition is divided by an intron into two structurally similar $(\alpha\beta)_3$ units, the two mononucleotide binding domains. One of these units is further subdivided by introns into three similar $\alpha\beta$-supersecondary structural elements. These results support the following view for the evolution of alcohol dehydrogenase. The genes for the two domains evolved independently of each other and were later fused together as was first suggested by Ohlsson *et al.* (1974) and Rossmann *et al.* (1974). The basic building block for the co-enzyme binding domain was a short $\alpha\beta$ unit of $\sim$20 residues. Three such blocks were fused together to form a mononucleotide binding domain, which later duplicated to form the basic structure of the NAD^+-binding domains, as seen today with minor variations in all known NAD^+-dependent dehydrogenases.

Alternatively one could suggest a different evolutionary history for the two halves of the co-enzyme binding domain. It is on the one hand possible that there has been conservation of intron position in the first half without such restraint in the second mononucleotide binding domain. However, it is also possible that the two nucleotide binding domains have different evolutionary origins with the second domain providing the additional function of the alpha-helical connecting region between the co-enzyme binding domain and the carboxy-terminal structural unit of the catalytic domain.

Materials and methods

The amino acid sequence of maize alcohol dehydrogenase (Dennis *et al.*, 1984) was aligned to the horse liver alcohol dehydrognase sequence (Jörnvall, 1970) based on the high homology. A model of maize alcohol dehydrogenase based on this homology was built on a computer controlled display system UG 3404 using the program FRODO (Jones, 1982). Distance plots were calculated from the refined coordinates of horse liver alcohol dehydrogenase. Accessibility calculations (Lee and Richards, 1971) were made by a programme supplied by T. Richmond and F.M. Richards.

Acknowledgements

This work was supported by a grant from the Swedish Natural Science Research Council. C. Cambillau is a CNRS research fellow.

References

Adams,M.J., Ford,G.C., Koekok,R., Lentz,P.J.,Jr., McPherson,A.,Jr., Rossmann,M.G., Smiley,I.E., Schevitz,R.W. and Wonnacott,A.J. (1970) *Nature,* **227**, 1098-1103.

Banner,D.W., Bloomer,H.C., Petsko,G.A., Phillips,D.C., Pogson,C.I. and Wilson,A.I. (1975) *Nature,* **255**, 609-614.

Bennetzen,J.L. and Hall,B.D. (1982) *J. Biol. Chem.,* **257**, 3018-3025.

Blake,C. (1983) *Nature,* **306**, 535-537.

Brändén,C.-I. (1980) *Q. Rev. Biophys.,* **13**, 317-338.

Brändén,C.-I., Eklund,H., Nordström,B., Boiwe,T., Söderlund,G., Zeppezauer,E., Ohlsson,I. and Åkeson,Å. (1973) *Proc. Natl. Acad. Sci. USA,* **70**, 2439-2442.

Buehner,M., Ford,G.C., Moras,D., Olsen,K.W. and Rossmann,M.G. (1973) *Proc. Natl. Acad. Sci. USA,* **70**, 3052-3054.

Craik,C.S., Sprang,S., Fletterick,R. and Rutter,W.J. (1982) *Nature,* **299**, 180-182.

Dennis,E.S., Gerlach,W.L., Pryor,A.J., Bennetzen,J.L., Inglis,A., Llewellyn,D., Sachs,M.M., Ferl,R.J. and Peacock,W.J. (1984) *Nucleic Acids Res.,* in press.

Doolittle,W.F. (1978) *Nature,* **272**, 581.

Eklund,H., Nordström,B., Zeppezauer,E., Söderlund,G., Ohlsson,I., Boiwe, T., Söderberg,B.O., Tapia,O. and Brändén,C.-I. (1976) *J. Mol. Biol.,* **102**, 27-59.

Gilbert,W. (1978) *Nature,* **271**, 501.

Gô,M. (1981) *Nature,* **291**, 90-93.

Gô,M. (1983) *Proc. Natl. Acad. Sci. USA,* **80**, 1964-1968.

Jones,T.A. (1982) in Sayre,D. (ed.), *Computational Crystallography*, Oxford University Press, pp. 303-317.

Jörnvall,H. (1970) *Eur. J. Biochem.,* **14**, 521-534.

Lee,B. and Richards,F.M. (1971) *J. Mol. Biol.,* **55**, 379-400.

Liljas,A. and Rossmann,M.G. (1974) *Annu. Rev. Biochem.,* **43**, 475-507.

Lindqvist,Y. and Brändén,C.-I. (1980) *J. Mol. Biol.,* **143**, 201-211.

Ohlsson,I., Nordström,B. and Brändén,C.-I. (1974) *J. Mol. Biol.,* **89**, 339-354.

Rao,S.T. and Rossmann,M.G. (1973) *J. Mol. Biol.,* **76**, 241-256.

Rossmann,M.G., Moras,D. and Olsen,K.W. (1974) *Nature,* **250**, 194-199.

Rossmann,M.G., Liljas,A., Brändén,C.-I. and Banaszak,L.J. (1975) in Boyer,P.D. (ed.), *The Enzymes*, 3rd ed. Vol. **XI**, Academic Press, NY, pp. 61-102.

Russel,D.W., Smith,M., Williamson,V.M. and Young,E.T. (1983) *J. Biol. Chem.,* **258**, 2674-2682.

Stuart,D.I., Levine,M., Muirhead,H. and Stammers,D.K. (1979) *J. Mol. Biol.,* **134**, 109-142.

Received on 12 March 1984