Analysis of the Gene Encoding the Multifunctional Alcohol Dehydrogenase Allozyme ADH-71k of Drosophila melanogaster

Karel Th. Eisses, * † Alexandra J. Andriesse, † A. Douwe de Boer, † George E. W. Thörig, * and Peter J. Weisbeek †

*Department of Population and Evolutionary Biology and †Department of Molecular Cell Biology, Rijksuniversiteit at Utrecht

The nucleotide sequence of the alcohol dehydrogenase gene Adh71k has been determined. The Adh71k allele encodes the thermostable and multifunctional ADH-71k allozyme of Drosophila melanogaster. Comparison with the sequences of Adh5, Adh3, and AdhFCHD reveals differences in the coding and noncoding regions of the gene. Conceptual translation of the Adh71k sequence indicates that ADH-71k shares with ADH-F and ADH-FCHD an amino acid replacement at residue 192 and with ADH-FCHD an additional replacement of serine for proline at residue 214. Three unique differences were found in the nontranslated regions. It is proposed that a nucleotide deletion in the adult intron is related to the difference in expression level of the Adh71k allele, relative to the other alleles. An insertion of five nucleotides, additional to a single base deletion at that site, was detected in one of the larval enhancer regions in the 5' flanking region of the Adh71k allele, creating a palindromic structure in that area.

Introduction

The alcohol dehydrogenase gene-enzyme system of Drosophila melanogaster has been studied from the population level to the DNA level (Chambers 1988). Population genetic studies revealed two worldwide existing electromorphic enzyme variants, ADH-SLOW and ADH-FAST, which differ structurally by a single amino acid residue (Thatcher 1980). In flies homozygous for these alleles there is a twofold difference in protein concentration (Laurie-Ahlberg and Stam 1987). A close correlation has been found between enzyme activity and alcohol tolerance. Flies and larvae homozygous for the AdhF allele show the highest ADH activity and tolerate higher ethanol concentrations in general (Kamping and Van Delden 1978). Because D. melanogaster frequently encounter environmental alcohol in their rotting-fruit diet, the Adh polymorphism has been used to study natural selection (e.g., see van Delden 1982).

Besides other electrophoretic variants and null mutants, an aberrant ADH-F variant showed up in laboratory strains containing Notch mutants (Thörig et al. 1975). Unlike ADH-FAST, this ADH-71k variant is very thermostable (Thörig et al. 1975; Eisses et al. 1985a). ADH-71k differs also in substrate specificity (Eisses 1989a, 1989b) and has multifunctional properties with respect to the oxidation of substrates normally
oxidized by flavoproteins (Eisses et al. 1985b). Moreover, this enzyme is present in a 30% higher protein concentration relative to ADH-F (Eisses et al. 1985a; Heinstra et al. 1987). The multifunctional property of the ADH-71k allozyme and high expression of the Adh71k allele provide a selective advantage over the other Adh alleles in combination with the Notch8 mutation, which can account for its fixation in Notch mutant strains (Eisses et al. 1985b, 1986; Eisses 1986). Other heat-resistant ADH-F variants—ADH-FAST RESISTANT (ADH-FR) and ADH-FAST CHATEAU DOUGLAS (ADH-FCHD)—have been detected, respectively, in field populations in the United States and in Australia and China. The allele frequencies were found to vary from a few percent in the United States (Sampsell 1977) to 16% in Australia (Gibson et al. 1982) and to as much as 35% in China (Jiang et al. 1989).

We determined the base sequence of the genomic DNA fragment incorporating the Adh71k allele, in order to explain the differences in substrate specificity, kinetic parameters, thermostability, and quantity of ADH-71k, relative to ADH-F, ADH-S, and ADH-FCHD. Collet's (1988) sequence of the AdhFChD allele permits the comparison of two sequences of thermostable ADHF variants with those from various Fast and Slow alleles (Kreitman 1983). The ultimate aim is to evaluate the role of structural differences and differences in enzyme quantity in the selection pressure on the Adh polymorphism.

Material and Methods

DNA was extracted from 2 g of decapitated flies homozygous for the Adh71k allele, by using a modified method of Muskavitch and Hogness (1982). Density-gradient centrifugations, extractions, and alcohol precipitations were performed according to the method of Maniatis et al. (1982). An enriched plasmid bank was constructed: chromosomal EcoRI fragments 4–5 kb in length were isolated from an agarose gel after electrophoresis according to a modification of the method of Burns and Beacham (1983). These fragments were ligated into the vector pEMBL 9 (Dente et al. 1983) and were transformed into Escherichia coli JM109 according to the method of Mandel and Higa (Maniatis et al. 1982). These cells were plated out and blotted on nitrocellulose. Cells containing the vector with a 4.8-kb EcoRI fragment containing the Adh allele were detected in the library by colony hybridization with a 4.8-kb probe, according to the method of Hanahan and Meselson (1980). This probe was derived from the plasmid sAF2, which consists of an 11.8-kb SacI genomic fragment around the AdhF gene (Goldberg 1980) inserted into the vector pSV2. From this probe a 4.8-kb EcoRI subclone in pEMBL 9 was constructed. The 4.8-kb EcoRI fragment, containing the Adh71k allele, was subcloned in the vectors pEMBL 8, 9, 18, and 19 by using various restriction sites. As well as the 4.8-kb EcoRI/EcoRI clone, 2.1-kb and 2.6-kb BamHI/EcoRI, 1.3-kb SalI/BamHI, 0.72-kb SalI/HindIII, 0.6-kb BamHI/HindIII, 0.16-kb HindIII/HindIII, and 0.66-kb SalI/BamHI subclones with fragments in two orientations were constructed. The various clones were transformed into E. coli JM 101, recA, and hsdS. Superinfection with phage IR1 was used to produce ssDNA. M13 primer or primers matched against the consensus sequence of Adh8 were used as initiation sites of the polymerase reaction in the dideoxy-chain termination method of Sanger et al. (1977), with 35SdATP. The subclones were each sequenced three to five times with good resolution, without contradictory results. The parts of the sequence comprising unique differences were sequenced twice in one direction and once in the other direction. Gel electrophoresis was performed with a Bio-Rad
Sequi-Gen® nucleic acid sequencing system with wedge spacers. Sequence comparisons and secondary-structure calculations were performed with the MICROGENIE® sequence analysis program of SciSoft, Inc. (of Beckman Instruments, Inc.).

Results

An enriched plasmid bank in pEMBL vectors was made from the Drosophila melanogaster genome by digestion of the DNA with EcoRI. The Drosophila strain was homozygous for the Adh71k allele. By means of colony hybridization with a probe containing the AdhF allele, some positive clones were isolated. The clone that gave restriction-fragment patterns identical to that of the probe was used for subcloning and sequence analysis (fig. 1).

The nucleotide sequence of Adh-71k was determined for 1.9 kb spanning the Adh coding region and for an additional 380 bases spanning part of the 5' flanking region (figs. 1–3).

In other studies the Adh-71k sequence from base −63 to base +1929 has been compared both with the respective consensus sequences of six Adh-Slow and five Adh-Fast haplotypes (Kreitman 1983) and with Adh-FChD (Collet 1988), which encodes another thermostable ADH variant (table 1). The nomenclature of these studies is different from ours. The 5' flanking region, bases −1297 to −916 5' upstream of the distal start site (fig. 3), was compared with sequences from African (Af-s) and Japanese (Ja-f) strains, which are identical for this part of the sequence (Kreitman and Aguadé 1986).

Adh-71k has three unique mutations in the noncoding regions: two single site deletions between positions 580/581 and 1700/1701 and a single base substitution at position 886. The G deletion between positions 580 and 581 destroys one of the GC(A/T)GC tandem repeats (Heberlein et al. 1985) and reduces the stability of a

![Fig. 1.](image-url)
Fig. 2.—DNA sequence and conceptual translation of the Adh<sup>71k</sup> haplotype from base -63 to base +1929. The coding region of Adh comprises three exons separated by two introns. A proximal TATA box is located from positions +680 to +687, and the proximal leader (from position +712) is contiguous with the first coding region. The distal TATA box is located at positions -32 to -26, and a short transcribed region of 87 bases (exon 1) is spliced into a site 36 bases 5' to the ATG residue. The 658 bases between exon 1 and the splice site in the proximal leader comprise intron 1. Numbering of the amino acid residues begins with the first serine (Thatcher 1980). Amino acid replacements compared with other allozymes are underlined. Triple underlining (≡) indicates sites with unique differences between Adh<sup>71k</sup> and Adh<sup>FCD</sup>. Base deletions are indicated by a capital delta (Δ). Insertions are indicated by an inverted capital delta (▽), with double-shaft arrows (‖) pointing to the sites of insertion. Single-barb arrows (- - -) indicate (inverted) repeats in the P region (Heberlein et al. 1985).
DNA Sequence of Thermostable ADH-71k of Drosophila

**FIG. 2. (Continued)**

potential secondary structure within the region 571-601, as indicated by Kreitman (1983, fig. 3b): $\Delta G = -38.5$ kJ/mol for Adh-71k instead of $\Delta G = -81.2$ kJ/mol for Adh-Slow.

The base substitution at position 1494 results in the replacement of lysine in ADH-S by threonine at residue 192 in ADH-71k, ADH-F, and ADH-FCHD. This mutation is responsible for the higher electrophoretic mobilities of the latter enzymes. A single base substitution at position 1559 results in the replacement of proline in ADH-F and ADH-S by serine at residue 214 in ADH-71k. This mutation is identical to the one in ADH-FCHD (Collet 1988) and is responsible for the high thermostability of these allozymes (Thöring et al. 1975; Gibson et al. 1980; Chambers et al. 1984; Eisses et al. 1985a). No differences between Adh-71k and Adh-FChD were found in the coding regions.

Adh-71k shares with Adh-FChD some differences from Adh-Fast: (1) a synonymous substitution in exon 2 at position 820, (2) a single site substitution at position 521, (3) absence of a length polymorphism at position 555/556, (4) an insertion of a homopolymer repeat at position 1701, (5) a single site insertion (T) at position 1741, (6) a single site deletion (T) at position 1770/1771, and (7) 17 adenine residues
at the homopolymer repeat within the 3' noncoding region of Adh-71k and FChD, compared with 15 and 11 residues present in the consensus Fast and Slow sequences, respectively. The sequence between positions 482 and 886 is characteristic of the Slow consensus sequence. Differences from the consensus Slow sequence are similar to those described for Adh-Fast and Adh-FChD (Kreitman 1983; Collet 1988).

The 5' flanking region of Adh-71k differs at least at two positions from those of Adh-Fast (=Ja-f) and Adh-Slow (=Af-s) (Kreitman and Aguadé 1986) (fig. 3). A base substitution at position 3476 (-1294) destroys a HindIII restriction site very close to the EcoRI site. A single base deletion at position 3660 (-1110), with a simultaneous insertion of five bases (ATGTA), creates a 4-base direct repeat (TATG) which is tandemly reiterated three times. An identical 4-base repeat is located 3' adjacent to the point of insertion. The insertion alters the sequence such that a 10-bp inverted complementary repeat arises, which can form a secondary structure from position 3653 to position 3686, although with low stability: $\Delta G = -41$ kJ/mol.

**Discussion**

The genomic DNA sequence analysis of Adh-71k gives clues to several aspects of the phenotypic expression of the gene. The lysine-to-threonine mutation at residue 192 changes the electrophoretic mobility of ADH-71k from slow to fast as in ADH-F (Thatcher 1980). No charge change is involved with the proline-to-serine replacement at residue 214, resulting in a "cryptic" thermostable allele only detectable by heat treatment of the gel, prior to staining (Thörig et al. 1975; Eisses et al. 1985a). The serine in ADH-71k alters the conformation of the enzyme and its thermostability and substrate specificity relative to ADH-F and ADH-S (Eisses et al. 1985a, 1985b; K. Th. Eisses, unpublished data). These alterations are possibly due to a less rigid structure of the substrate-binding domain: the proline-to-serine replacement is positioned in the turn between two $\alpha$ helices (Thatcher and Sawyer 1980; Chambers et al. 1981; Benyajati et al. 1983).

The ADH activity per fly depends on both the rate of turnover and the number...
<table>
<thead>
<tr>
<th>SITE NO.</th>
<th>5' Flanking Sequence</th>
<th>Intron 1</th>
<th>Exon 2</th>
<th>Intron 2</th>
<th>Exon 4</th>
<th>3' Noncoding Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>447</td>
<td>555</td>
<td>580</td>
<td>820</td>
<td>1700</td>
</tr>
<tr>
<td>Adh</td>
<td></td>
<td>1770</td>
<td>1742</td>
<td>1740</td>
<td>1697</td>
<td></td>
</tr>
<tr>
<td>71k</td>
<td>T G C A G A T C V1 C</td>
<td>-Δ1 T</td>
<td>A G T</td>
<td>C T C T</td>
<td>C C</td>
<td>Δ2 17A V3 G Δ3</td>
</tr>
<tr>
<td>FChD</td>
<td>T G C A G A T C V1 C</td>
<td>G T</td>
<td>G T C T</td>
<td>T C T C</td>
<td>C C</td>
<td>C 17A V3 G Δ3</td>
</tr>
<tr>
<td>Fast</td>
<td>T G C A G A T C V1 G</td>
<td>G V2 G</td>
<td>G T C C</td>
<td>T C C C</td>
<td>C C</td>
<td>C 15A - G T</td>
</tr>
<tr>
<td>Slow</td>
<td>C G C A T G G C C</td>
<td>-Δ1 G</td>
<td>T C C A</td>
<td>C T C A</td>
<td>A C</td>
<td>11A - C T</td>
</tr>
</tbody>
</table>

NOTE.—Insertions and deletions are measured relative to Adh. Insertion V1 is 5 bp longer than the deletion at this site in Adh. The numbering is according to that of the presented Adh sequence.

* Implies amino acid polymorphisms.
of active enzyme molecules produced. In both aspects the homozygous genotypes differ. The amino acid differences between ADH-71k, ADH-F, and ADH-S can explain, in principle, differences in turnover (Winberg et al. 1982; Hovik et al. 1984). In the Utrecht strains, homozygous for the Adh71k allele, the ADH protein quantity is 30% higher than that in AdhF homozygotes, while in AdhF flies and larvae the ADH quantity is almost double that of AdhS flies and larvae (Heinstra et al. 1987). However, in the Australian AdhFChD strains the ADH content in 7-d-old males is at intermediate levels between those from AdhF and AdhS homozygotes (Chambers et al. 1984). These differences can hardly be due solely to variable degradation rates but rather must be due to differences in transcription and/or translation. In P-element transformation experiments Laurie-Ahlberg and Stam (1987) recombined at the HpaI site at position +325 the Adh-Fast distal promoter, including larval and adult enhancer regions (Posasony et al. 1985; Corbin and Maniatis 1989a, 1989b), with the Adh-Slow coding region 3' downstream from this site, and vice versa. The protein levels were related to the 3' downstream region. However, the two large insertions in the Adh-Fast intron 1 were not involved in the recombination experiment (insertion VI is 117 bp 3' downstream from the HpaI site). Nevertheless, absence of the two insertions in the Adh-Fast sequence of the Florida strain (Kreitman 1983) did not change the protein level (Laurie and Stam 1988). Moreover, the average mRNA ratio equals one for AdhF and AdhS alleles of several geographic origins, but the cross-reacting material differs by 50%, suggesting a difference in translation rate. Adh-Slow differs from Adh-Fast, Adh-71k, and Adh-FChD in the coding region 1477–1532 by three single base substitutions (Kreitman 1983; Collet 1988), one of them being responsible for the lysine-to-threonine replacement at residue 192. Only in the Adh-Slow mRNA can a secondary structure with a ΔG = −46.9 kJ/mol be formed in this region. Translation into the ADH-S protein might be retarded at this hypothetical structure.

This leaves the question about the differences between the ADH protein levels of the two thermostable allozymes. Although ADH-71k and ADH-FChD are structurally identical, the DNA sequence differs at three positions in the noncoding regions. It is tempting to relate these unique sequence differences to differences in protein expression level. One difference [the single base deletion (G) at position +580/581 or −132 P, when counted 5' upstream from the proximal start site] is a likely candidate. This deletion disrupts one of the GC(T/A)GC repeats in this region—and, with that, an area with a larger inverted repeat, which is recognized by the transcription factor Adh-1 (Heberlein et al. 1985). In Adh-FChD, Adh-Fast, and Adh-Slow the stability of the secondary structure of this inverted repeat is ΔG = −81.2 kJ/mol. The stability has been reduced in Adh-71k to ΔG = −38.5 kJ/mol. A similar situation exists in the Adh-Fast gene of a Japanese strain in which the inverted repeat has been disrupted because of a lack of the V2 insertion and a single base substitution at position +590 or −122 P (G to T) (Kreitman 1983): ΔG = −38.5 kJ/mol. The Japanese AdhF strain shows the highest specific ADH activity (30% higher than that of the other Fast strains) (Laurie-Ahlberg and Stam 1987) and a slightly higher mRNA ratio relative to Adh-Slow (Laurie and Stam 1988). This high activity must be due to higher enzyme expression, because there are no amino acid replacements. The high expressions of ADH-71k and ADH-F (Japan) might therefore have the same cause. The so-called P1 region (+558 to +602 or −150 to −110 P; Heberlein et al. 1985), comprising the deletion and substitution, has been shown by deletion experiments to be important for induction of ADH activity (Sofer and Martin 1987; Shen et al. 1989; Kapoun et al. 1990). It is possible that lowered binding of transcription factors to the recognition
sites involved with proximal mRNA transcription allows more distal mRNA transcription in the larval stage of Adh71k.

A complicating factor is that the 5' flanking region of Adh-71k, part of the larval enhancer region (Corbin and Maniatis 1989b, 1990), is also different. Enhancer region polymorphisms supposedly interact differently with the various promoter regions, which hampers a conclusive analysis. The enhancer regions comprise clusters with an extremely high frequency of a 5-bp motif [ATAC(T/A), TATG(A/T), or their inversions], occurring on average every 29 bp in the region between -1300 to -800 bp 5' upstream from the distal start site. In the Adh-71k 5' flanking region (fig. 3) the frequency is increased to an average of every 20 bp, because of the single base deletion and insertion of a 5-base stretch, ATGTA, at position 3660 (-1110 D). The ATAC(T/A) motif is repeated in intron 1 10 times (every 50 bp) and four times in the 3' noncoding region (every 40 bp). The reiterated motif ATACT in the insertion V1 or P2 region is recognized by transcription factors (Heberlein et al. 1985). It is therefore tempting to ascribe functional significance to this motif sequence, both in the enhancer regions and in the noncoding regions of the gene.

The sequence of Adh71k is obviously most closely related to AdhFChD, and these two sequences are closely related to the AdhF haplotypes from the strains from France, Washington, and Africa (Kreitman 1983; Collet 1988). This suggests that the two thermostable alleles arose from an AdhF haplotype, with an additional intragenic double crossover with the AdhS haplotype between the positions +482 (insertion of ATACT repeat) and +884 (intron 2).

The Adh71k allele was fixed in laboratory strains of Drosophila melanogaster comprising Notch mutants, originating from the United States (Thörig et al. 1975). ADH-71k has a broader substrate specificity than ADH-F. It even acts on substrates normally handled by flavoproteins (Eisses et al. 1985b). Notch mutants are partly deficient for a range of flavoproteins (Thörig et al. 1981), which results in selection in favor of alternative pathways that improve their viability (Eisses et al. 1985b, 1986; Eisses 1986). ADH-71k seems to function as such an alternative pathway, whereas its bypass functioning is facilitated by the large quantity of ADH-71k.

Acknowledgments

We would like to thank Dr. J. Hageman for instructive help at various levels, Dr. D. Goldberg for supply of the AdhF probe, and Professor W. Scharloo for critical reading of the manuscript.

LITERATURE CITED


RICHARD K. KOEHN, reviewing editor

Received January 16, 1990; revision received April 2, 1990

Accepted April 3, 1990