Molecular analysis of the alcohol dehydrogenase 2 (Adh2) gene of maize

E.S.Dennis, M.M.Sachs¹, W.L.Gerlach, E.J.Finnegan and W.J.Peacock

CSIRO Division of Plant Industry, GPO Box 1600, Canberra 2601, Australia, and ¹Department of Biology, Washington University, St. Louis, MO 63130, USA

Received 29 November 1984; Accepted 2 January 1985

ABSTRACT

The Adh2 gene of maize has a nucleotide sequence closely related to that of the maize Adh1 gene indicating that the two genes arose from a progenitor gene by a duplication event. The coding regions are 82% conserved at the nucleotide level and 87% conserved at the amino acid level. Each gene has nine introns in identical positions but their nucleotide sequences and lengths differ. Adh2 encodes a single mRNA and has a possible polyadenylation signal, AATAAT, fifteen bases upstream from the polyadenylation site. The Adh2 gene together with the Adh1 gene is induced by anaerobic conditions and under these conditions produces an increased level of mRNA. The 5' untranslated regions of the transcripts of Adh1 (100 bp) and Adh2 (126 bp) have diverged in sequence except for a conserved region which may be important for their anaerobic-specific translation. The 3' and 5' flanking sequences of the two genes are also divergent except for 11 bp of precise homology around the TATA boxes and three other 8 bp sequences further upstream. One of the common 8 bp sequences (CACCTCCC) is in a similar position (-150 bp to -200 bp) in the two genes and may have a regulatory function in gene expression.

INTRODUCTION

In maize the two alcohol dehydrogenase genes, Adh1 and Adh2, are on separate chromosomes; Adh1 is located on chromosome 1 (1) and Adh2 on chromosome 4 (2). The polypeptides produced by both genes have a molecular weight of about 40,000 daltons although it has been suggested, on the basis of mobility in SDS-polyacrylamide gels, that the Adh2 encoded polypeptide (ADH2) has a molecular weight 2000 daltons greater than the Adh1 encoded polypeptide (ADH1) (3). Alcohol dehydrogenase enzyme (ADH) is a dimer, and both homodimers (ADH1.ADH1, ADH2.ADH2) and the heterodimer (ADH1.ADH2) are active enzymes (4, 5), emphasizing the similarities of the two polypeptides and suggesting that they are closely related and may be derived from an ancestral gene duplication. Antibodies to ADH1.ADH1 dimer partially cross-react with ADH2.ADH2 showing that they share some, but not all, antigenic determinants (5). The specific activity of the ADH1.ADH1 enzyme

is ten-fold higher than that of the ADH2.ADH2 enzyme (5) and the two genes have different patterns of expression in different tissues of the plant. ADH1 predominates in seeds and is the only alcohol dehydrogenase present in pollen while ADH2 is the major form in the tassel nodes and in mature node cells of the stem (2).

These two gene products are among a small number of polypeptides induced by anaerobic conditions (6, 7, 8). ADH1 is important for survival of the seedling during flooding but the function of ADH2 is less clear. ADH1⁺ ADH2⁻ plants seem as healthy as ADH1⁺ ADH2⁺ plants even under anaerobiosis, but ADH1⁻ ADH2⁻ seeds germinate poorly and the seedlings succumb to flooding more rapidly than their ADH1⁻ ADH2⁺ counterparts (2). Translatable mRNA from both genes increases during anaerobiosis (3) and anaerobic polypeptide mRNAs are specifically translated under anaerobic conditions despite the presence of aerobic mRNAs in the cell. Both genes are also induced by the synthetic auxin 2,4D (7).

We have previously cloned and sequenced the Adh1 gene (9), identified transcription and translation start sites, predicted the complete protein sequence and described the position and sequence of the nine introns interrupting the gene. We have also identified putative promoter elements (TATA and CCAT boxes) which may be important for expression of the gene. In this paper we report the cloning and sequencing of the Adh2 gene and discuss its relationship to the Adh1 gene. The nucleotide sequences of the coding regions show that the Adh1 and Adh2 genes are homologous and must have had their origin from the duplication of an ancestral gene. The 3' and 5' flanking sequences have diverged except for short regions in the 5' region; one conserved region involves the presumptive TATA box which is known to be important for transcription and one or more of the others may be of significance in the induction of the Adh genes by anaerobic conditions. A conserved region in the 5' untranslated region of the mRNAs may control the preferential translation of anaerobic specific mRNAs.

MATERIALS AND METHODS

1. Maize lines

Cromac: Commercial hybrid line (9); BKF: Inbred line (8) containing standard Adh1-1F and Adh2-N alleles; P78: Rust resistant line (10); WFK: Wilbur Flint Knobless - Maize Genetics Co-Op Stock Centre no. 71-900(x); MS904: A line with the null allele Adh2-33 [a naturally occurring Adh2 null allele from Wilbur Flint Knobless] crossed into BKF background; REV: A

revertant to wild type ADH activity from Adh1-1Fm335 (11).

2. cDNA cloning

cDNA clones for Adh2 were prepared from the Cromac line of maize as described previously (9). The insert from the Adh1 cDNA clone (pZML84) (10) was used as a probe in colony hybridization under standard conditions (hybridized in 3 X SSC/50% formamide at 42°C and washed in 0.1 X SSC at room temperature). Two classes of hybridizing colonies were detected: strongly hybridizing Adh1 clones and weakly hybridizing Adh2 clones. The weakly hybridizing clone pZML841 was selected for further study.

3. Genomic cloning

A partial Sau3A1 digest of BKF DNA was sedimented through salt gradients (5-30% NaCl in 0.01M Tris, 0.001 M EDTA pH8.4) and the 12-20 kb fraction isolated and ligated into the BamH1 site of λ Charon 28 (12). The library was screened by plaque hybridization (modified from 13) using the insert of the Adh1 cDNA clone (pZML84) as a probe. Positive plaques were purified and hybridized to both pZML841, an Adh2 cDNA clone, and to pZML84, the Adh1 cDNA clone, to discriminate between Adh1 and Adh2 genomic clones.

4. Southern and Northern transfer and hybridization were as described previously (9).

5. Sequencing

The cDNA clone pZML841 was sequenced using the method of Maxam and Gilbert (14), DNA segments were labelled by "filling in" recessed 3' ends with the large fragment of DNA polymerase 1.

Another Adh2 cDNA clone pZM16.6 and the genomic clone λ 2N.1 were sequenced using the dideoxy method (15). Restriction fragments were cloned in the M13 vectors mp8 and mp9 (16).

7. S₁ mapping used probes labelled in a single strand and annealing and S₁ treatment were as described previously (9).

RESULTS

cDNA cloning

A cDNA clone (pZML841) encompassing approximately 1000 bp from the 3' terminus of the mRNA was selected by its weak hybridization to an Adh1 probe. The same probe hybridized strongly to a control Adh1 clone. The assumption that this was an Adh2 cDNA clone was supported by its failure to hybridize to mRNA isolated from a specific Adh2 null line (see section on Expression of the Gene) and by direct sequencing. A longer Adh2 cDNA clone, pZM16.6, was isolated from BKF RNA and identified using pZML841 as a probe. Figure 1 shows the sequence comparison between the Adh1 cDNA clone pZML793

Fig. 1 Comparison of the coding regions of *Adh1-1F* (top line) and *Adh2-N* (bottom line). The *Adh2-N* sequence presented is a composite derived from sequencing the two cDNA clones pZML841 and pZML6.6 and the coding region identified from the genomic clone λ 2N.1. The regions from each clone are indicated. The standard one letter amino acid code is used and where *Adh1* and *Adh2* differ the changed amino acid is noted. The positions of introns are indicated by a vertical line together with a Roman numeral. The regions of homology in the 5' untranslated region are underlined.

TABLE 1
Analysis of nucleotide changes in the *Adh1* and *Adh2* coding sequences

A. Percentage difference between ADH1 and ADH2 coding sequences (380 codons)					
	Nucleotides	Amino acids			
	17.8	12.6			
B. Proportion of changes resulting in changed amino acid sequence					
	No. of codons with changes				
	EXPRESSED	SILENT	TOTAL		
At position					
1 only	14	1	15		
2 only	10	0	10		
3 only	3	126	129		
1+2	3	0	3		
1+3	12	2	14		
2+3	3	0	3		
1+2+3	3	0	3		
none			203		
Changes: 72 expressed, 131 silent.					
C. Analysis of frequency of transitions and transversions					
111 Transitions and 93 Transversions					
CODON POSITION		1	2	3	TOTAL
Transitions	- silent	1	0	79	80
	- expressed	14	10	8	32
Transversions	- silent	2	0	49	51
	- expressed	18	9	13	40

(9) and a composite of the two *Adh2* cDNA clones (pZML841 and pZML6.6) together with the presumed coding region of the genomic clone (λ 2N.1, see next section). These sequences have been aligned at a stop codon (TAG) which is the C-terminus of the ADH1 polypeptide (10) and it seems likely that it is also the C-terminus of the ADH2 polypeptide. Upstream of this codon the sequences are 82% homologous but downstream they diverge completely. In the coding region a majority of the base changes (131 out of 203) are silent (Table 1) but 72 do change the amino acid sequence. The clone pZML841 has a polyA tail and the length of the 3' untranslated region of the *Adh2* mRNA to this site is 310 bases. The sequence AATAAT lies 15 bp upstream of the A residues and could be the operative polyA addition signal analogous to the animal consensus polyA addition signal (AATAAA) (17).

Restriction and sequence analysis of the genomic *Adh2* segment

Southern DNA blot analysis of DNA from different maize lines cut with a number of restriction enzymes (Fig. 2) showed simple patterns with the *Adh2*

Fig. 2 Genomic organization of *Adh2* genes. DNA of three maize lines was digested with three restriction enzymes and hybridized to a nick translated 3 kb Bam-HindIII genomic fragment (2N.1L) spanning the coding region of the *Adh2* gene (Fig. 3).

cDNA probe, the small number of fragments and hybridization intensity indicating that, like *Adh1*, *Adh2* is present in the genome in a low or, more likely, single copy number.

Three different maize stocks BKF, WFK (containing a naturally occurring *Adh2* null allele) and the REV line all have different patterns when digested with each of three restriction enzymes (Fig. 2) showing there is restriction polymorphism around the *Adh2* gene. This polymorphism is comparable to that seen in and around the *Adh1* gene (18, 19).

Fig. 3 Restriction map of λ2N.1, genomic clone of the *Adh2* gene in λCharon 28. Fragments used as probes are indicated.

CCGACGATG CCGACGATCG CAGCGCG CAGCCAGCA CAGACGGCG TCGCTCTCG CAGCCACCA CAGTTGGCGA CTTGTGCGAC CCGACAGATG
TCTCTGATG TCTATTCAATA TCTCTTAATA CAGCCAGCA CAGACGCAAC TATACCAAGA CAGCCAGCA CAGTTGGCGA CAGTTGGCGA CAGTTGGCGA

CACCACTTCC CACGACGAGG AAAACACAGT CAGCCAGCA CCGGATATGTT CCACTCCAGG TGGAGTGCA GCCCCGGTTT CCGAGAGGCC GCAGCTGGTT
TAAAGACGA ATCGGACGAG GACAATGGT CACCTCTGCT CTTCTTCCCT TTTCTACCTG CAGTAAAA AACATCCGAG CTTTCTTCT CCGAGTCCGC

TTGTGCGCA CAGCGCGGCA ACCGACACCC CTTCTCCGCT CTTCTCCAT CTTCTCTCT TTAGACGTA CACATATAT CAGTACGGCT CATTCTTCGG
CAGCTAGCTG AGGGGTTGT GTTGCTCTCC TGTAGACTATA CAGCTCTAGC CACTATATAT AGFAGCTACG TCCCGGTTT CTTCTCCAGC ACCTCTCAT

CTCTCCAGC GCGCTCATCT CTTGGTCTG ATTTGGTTGT TAACTGGTGA AGCATACGAG CTTCCGAGT GGATCGATT GGGATCTGT TCGAAGATT TCGAAGATT
CATCAGGCA GCTTAAAGG TCGGAGGAAG CAGTGAATGA AGGATCTAT CTTCTGGTGT TGGAGTGGT CACATGACAT AGCTTCGTGA AGAGAGAGA

CGCGAGAGCA CCGGAGTGA CCGCGGGGCA GGTGATCAAG TCGAAGATC CCGCTGTTT CTTCTCTGT TTTGTATGTA CTAATCTCTGT GTTTATGAT
GAGAGACGA GATCTAGCA CCGGAGTGA GGTGATCAAG TCGAAGATC CCGCTGTTT CTTCTCTGT TTTGTATGTA CTAATCTCTGT GTTTATGAT

CGTAGTAAT TTTGGGAGA AGCTTGCTG ACAGTTTGT TTTGTAGAA CAGTGCAGCA GTGGCGCTGA TCTTGATGTC TATCTCGCAA TCGTGGTGAA
CTAATATCA TGGATTATT TTTGTCTTG ATCTCTCTG AG

CTTATGCTT TTTATCTT CACTACATG AAAAGACTAG TATCTTCTT CGATGTAA CAGTGCAGCA TCGTATGTA CCGTGGTCC ATCCGACAT

CTGGTAGAC ACATCATAG ATATTGAGA AAGATATCT TTCCCTGTT TTTATGAA GAGCTATT TCACTAGT GATCTAAGAA TGTGCAACT

TGCGAGAGG GDTTCTTCT TTGAATTATA ACTACTGCT TGAGTGGCC TGTTTCTCG AGTAAAGCT TTGTGCTG CACACATGT CATTGGAAT

TTACGTTGT TACGAAGCG AAAAGATTG CATCTTACG ATTTAGCTG ACTATGCT TGCITTTCT GACCCGTGCA GTGCGGGG CAGGAGGAG
TATATATAT TATGTTGTTT

CGGACAGCA CTTGTCTAG AGAGGTGSGA GTTAGGCGCT CCGCAGCGCA TGGAGTGGG TATCAAGAT CTTCTACAT CCGTCTGCGA CAGCAGCGT
CGGCAACCG TGTCTCATG AGAGGTGSGA GTTAGGCGCT CCGCAGCGCA TGGAGTGGG TATCAAGAT CTTCTACAT CCGTCTGCGA CAGCAGCGT

TACTCTTGG AGGCAAAAT ACTATCAAT CATTCCATCT TGTGATCTT TCGATAGAT ATGATACAC CAGCTGGAT CTTCTGCGC CTTCTTGGCG
TACTCTTGG AGGCAAAAT CTTCTGCTG TACTCTATCA CCGCGAAGC GATATAGT ATGATACAC CAGCTGGAT CTTCTGCTG CATTGTTGGC

GGTATTCT CTGAC GCGACATCC GTGTTGCT CCGATCTGT GCGATAGTG GCGATAGTG GCGAGATAT GTTCTATCT CCGATATCT TCACTATGT
TGAACTGTT CTTGTGAGT GCGAACCC GTGTTGCTG AGGATCTAG CAGCAGAG GCGAGATAT GTTCTATCT CCGTCTGCT TCTGTGCTGT

GCTGACTATA GATGTGCTGT GTTTATATT TACATATTA TTTATGTTT T GCGTGGAT TATGAGGAT GTTGGTGGT GTTCTTTAT GTTTTACTA
ACTGTTTCT TACTGCTG TACTGAGTC ATCACTTCT TCCACTGTT TTTTGGAG TGCAGTATG TGTCTATTA TATGAGAT TCACTAGT

GATCATCTG GAAAGTCAG AAGAAATAT TTTTGTGTA AAGTATGAT CTTTGGTGG ATATTATAT ATCTGTGCG ACATGAAAG TCTTGGGTA
ATATTAGGCG TCGTTGTTGA ACATAGTCT GTCAATCTT CAGTCAGCA CCGTGGTGG CAGTGATT TAAAGTCAAT ACTAAGAAAG ACAGCCTGC

CTGGAGGCT GTTGTGAGC ATTTATCTAT GTTCTTTCTA TTGAGCTTT CTTGCACTTA CTTGCACTTA CATTTTCTA GATACAAAG ACAGTACAG
ATATTATTT GATTAATATG GGTATAAAT AAGATGCTT AAGTCTAGT AAGGACACA CAGCTGAGC CAGTCAGCA GATACGCAAG ATCTGACTT

CTAGGCTGT GATGAATGT TTAAGATCT TTTTCCATATA ATAGAGGCA CAAATATGAT TTTATTAGC CAGCTGAGT TCTTGATT TCGGAG
ACGATGTTTA TTTCTTAT TTTTCCATATA ATAGAGGCA CAAATATGAT TTTATTAGC CAGCTGAGC AAGCTATCT ATAGGCTGT TTAATATCT CATTGCGAG

TATCAT AGAGATGTT GAGAGAGGCG TGACTAGCT CCGCGCAGG GACCATCTG TCTGTGTT CATTGCGAG
TAAAGACCC GCTCTTGGT CAGATATGT GAGAGAGGCG TGACTAGCT CCGCGCAGG GACCATCTG TCTGTGTT CATTGCGAG

TGCAAGAT GCGCCACTG CAGTGCBGCA GAGAGACGA TGTGTGAT CCGCAGGAT AACACTGAG CCGGGTGTAT GATTGGCAT GGCATGAT
TGCAAGGAT GCGCCACTG CAGTGCBGCA GAGAGACGA TGTGTGAT CCGCAGGAT AACACTGAG CCGGGTGTAT GATTGGCAT GGCATGAT

GGTTTTCAT CAAATGGAAG CATTCTATC ACTTTTGGT GAGCTCTCAC TTGACGAGAT ACAGCTGAT CAGTGTGGT TGTGTGCGA AGATCAACG
GGTTTTCAT CAAATGGAAG CCGATCTAT ACTTTTGGT GAGCTCTCAC TTGACGAGAT ACAGCTGAT CAGTGTGGT TGTGTGCGA AGATCAACG

TCAGCGCTCG CTTGAAGA TTGGGCTGT TACGTTGGT ATTCTACG GTAAATGTA TTTACTATCT TTTGTTGGA TCGTGAGTA CATTATTCT
CGCGGCGCT CTTGAAGA TTGGGCTGT TACGTTGGT ATTCTACG GTAAATGTA TTTACTATCT TTTGTTGGA TCGTGAGTA CATTATTCT

GAGATGTA GTTACACAA TTTCTTTCT CCGCTTGT TTTGCTGCGA CAAATATG TTTGAAAC CCGAAGGTT TCGACAGTG
AGATGTA GTTACACAA CTTAGACTT CCGCTTGT TTTGCTGCGA CAAATATG CAAATATG TTTGAAAC CCGAAGGTT TCGACAGTG

CTGTTTGT TTTAGAGAG GTTGGCTG CTTAATGT TGAACATGT GTTGTGTT GTTATTTCT ATTGTCAGT AAGATGTTGT TGTGGTGT
CTGTTTGT TTTAGAGAG GTTGGCTG CTTAATGT TGAACATGT GTTGTGTT GTTATTTCT TTTTCTAAT TGGTGAAT TATGATGGA

ATCTGATG CTTGCGCAT GTTAGG CTTGAGAGT TCGAGAGT TCGAGAGT CCGAGATCAT CCGGTGCGA
GACATCTA CATTATCA TATATGTTT AAGCTATCA ATGATCAAT CTTGAGAGT TCGAGAGT TCGAGAGT CCGGTGCGA

CTGACGCA CAGATATGA AAGATGTA CTTACACAT CTTATATAT GTATGATGA TCCCTTGT CAGGAGCAT CTTGTGTA CTTAGTAT
ATACCCAG CAAATAGCA AGATGTA CTTACACAT TGCATCTA AGGCTGTG TTTTACCA ATCACTAT GATTAATAT CTTAGTAT

AGCTATTA TTTACTTGT TTAGCAAGT AGGATGTT CATTGTGCT ATTTTTCTA TTTGTTGAG TTTGTTGAG TGGATCTG TCTAATAT
GATGATCT TCTAATAT TATCCAAA CATTCTAGT CATTATAT TTTTATTC ATCTGAGT GCTGATTAT ATACAGCAT TGTGTGGGA

TAAATGAT TTTAATAGT TATGAAGC ATGATGTTA TTTAATCTA AGCAATCTA GAACTAAAT ATAAATAT TATATATTA TCCATATAT

ACTGATTAT CATTITATA CATTGTGCG AACAGTACT AGCTCATCA TCTGTCCAT TTTTATG
AAAAATAT AGGTTGAA GATACATTA TTAATGATG GATTATCT CAGATGAGA TTGACTGT AATACCCAG TTCAAGACC GATTAAATCT

ATGATGAT GATTATCT CATTGCA CAGCAAAAC ATTATATAT ATCTGCAAA TGTGATT TTTACTTGT GATTCTGT GCTCGCGCA

CTGAGATCTT ACTCTTAT GTTCTTGT TTTGTA TAAATATGTA AGTTTATG TCGACTAAT TCGTAAAG CAAAGACAG CAGACAGCG TCGAAGAT

TCTGTTTCTT AGCAGGCA AAAAAGTT ATCAAGCTT AGATTTAGC TGGCATAT ATCTAATAT CTTTATGAG TCTTATGAG TCTAG
GTGCTAGAT TGTGCGAT ATATAATTA CAGTGTGCA CCGTTTGT ATGACAAAGT CATTAAATCT CTTGATCTGA CTGCTGTTG

TTACTTGT GAGATCAAGA CCGGAGGTT CAGCCGAC GCGTGAAT CCGTCAAT CTTATGAT CTTATGAT CTTATGAT CTTATGAT

GTGCTGCT GTGCTATG TGTGCTAT GAGTCACTA AGCGGGGCT GACCGCAG CCGTGAAT CCGTCAAT CTTATGAT CTTATGAT

TGATGCTGT TCGATGTA ATATATTA TACACTCT GATCTTCTA TTTCTACG TCTTCTCT CAGCATCTA TGTTTATC CCGCGCTG
TCGATGAT CAGAGTA TGTCTGCT TCGCTCAAT CTTGTGAT TTTCTACG CATTCTAT TCTTCTCT TCTTCTCT TCTTCTCT

TGATGCTGT TCGGTTGT CCGCTGCTG TGGGTTGCT CAGTAAAGC GAGCAGTAT AGACCAAC CAGATGCT CTTAGCAAA GAGCCTTGA
AATAGGCTG TCGGTTGT CCGCTGCTG TGGGTTGCT CAGTAAAGC GAGCAGTAT AGACCAAC CAGATGCT CTTAGCAAA GAGCCTTGA

GGGACCTCT TTTGCAACT ATAGGACCG CATTGATCT CCAAGTGTG TGGAGTGT TAAGAAAAT TTAATTTCTA AGTGTGTT CTTCTAGCT
GGGACCTCT TTTGCAACT ATAGGACCG CATTGATCT CCAAGTGTG TGGAGTGT TAAGAAAAT TTAATTTCTA AGTGTGTT CTTCTAGCT

CTTACTGCG CAGGCTTGT CTTAAAGAT GTTAGAATG GTTCTGCGA TTTCTGACAG TTTGTTGTT TTTGTTGTT TTTGTTGTT TTTGTTGTT
CTTACTGCG CAGGCTTGT CTTAAAGAT GTTAGAATG GTTCTGCGA TTTCTGACAG TTTGTTGTT TTTGTTGTT TTTGTTGTT TTTGTTGTT

AMGTATCA CCGACAGCT CCGCTTCTG GAGATCAACA AGGCTGTCA CTTGATGCG AGGGGGAG GCTATCGCT CATCTACG ATGAGAAAT
AGTATCACTA CCGACAGCT CCGCTTCTG GAGATCAACA AGGCTGTCA CTTGATGCG AGGGGGAG GCTATCGCT CATCTACG ATGAGAAAT

AGTATCTAT CTTGATGTT TTTGCTGCG TGTGATGAG GAGGCAAT GAGCTGCT CTTGATGAG GAGGCAAT CTTGCTGCT TGTATGAT
AGTATCTAT CTTGATGTT TTTGCTGCG TGTGATGAG GAGGCAAT GAGCTGCT CTTGATGAG GAGGCAAT CTTGCTGCT TGTATGAT

ATGCGCCG CTTGTGCT GCGTTCAT GTTCTTGT CAGGAGCAT GATGACTCT CCGTGGCTA ATTAATGAT AGATACAT ACCTTTTGGC
AGAGCAATA AAGATTGAT TTTGCTCT GCGGCTAT TTTATTTCT CTTCTAGCT TTTCTAGCT TTTCTAGCT TTTCTAGCT

CTTACTGCG CAGGCTTGT CTTAAAGAT GTTAGAATG GTTCTGCGA TTTCTGACAG TTTGTTGTT TTTGTTGTT TTTGTTGTT TTTGTTGTT
CTTACTGCG CAGGCTTGT CTTAAAGAT GTTAGAATG GTTCTGCGA TTTCTGACAG TTTGTTGTT TTTGTTGTT TTTGTTGTT TTTGTTGTT

Fig. 4 The genomic sequences of the two maize *Adh* genes. The sequence of *Adh1-S* (above) is compared to *Adh2-N* (below). Intervening sequences (IVS) are indicated by lines and Roman numerals. The regions of homology about the TATA box are underlined. The transcription and translation start, the translation stop and the 5' upstream homologous regions are indicated. Where the intervening sequences are different lengths in the two genes the shorter sequence is filled with a line until it reaches the length of the longer.

TABLE 2
Length of introns in *Adh1* and *Adh2* genes of maize

Intron no.	Length (bp)	
	<i>Adh1-S</i>	<i>Adh2-N</i>
1	535	95
2	98	101
3	419	458
4	86	95
5	94	124
6	342	495
7	87	125
8	87	90
9	89	112

A maize DNA segment containing the *Adh2* gene from the BKF line was isolated from a *Sau3A* partial library in λ Charon 28. The clone, λ 2N.1, contains 13 kb of maize DNA and the restriction enzyme map of this clone (Fig. 3) is consistent with the data from the Southern analyses of genomic DNA, e.g. a 5.5 kb *HindIII* fragment containing the coding region is present in both the clone and the genomic DNA. Other internal sites are also consistent.

Coding region. The gene region occupies about 3.5 kb of the genomic clone λ 2N.1. This region was sequenced and compared to the of *Adh1* gene (Fig. 4). The coding regions in the genomic clones were identified from comparison with both *Adh2* and *Adh1* cDNA clones and used in Figure 1 to derive the entire protein coding region. The coding sequences are conserved (82% identity) and encode polypeptides of identical length. As in *Adh1* there are nine AT-rich introns interrupting the *Adh2* coding region. The positions of the introns are identical but their sizes and nucleotide sequences differ between the genes (Table 2; Fig. 4). The sequences at the intron/exon junctions are shown in Table 3 and resemble those in the *Adh-1S* gene; the coding sequences around the junctions are highly conserved. Within the *Adh1* gene there is a region of 104 bp spanning the intron 6-exon 7 boundary which is tandemly repeated three times (9, 19) but in *Adh2* only the exon 7 portion of this region is present, and in single copy.

5' regions of *Adh1* and *Adh2*. The 5' regions of the two genes can be aligned since the nucleotide sequences of the coding regions of *Adh1* and *Adh2* are conserved downstream from the presumptive points of initiation of translation. Immediately upstream from this point the two sequences diverge. The ATG codon at this position in *Adh2* lies within the

TABLE 3

		Adh2 Intron Donor Sites																			
Intron																					
1		A	A	G	T	G	C	A	G	A	G	G	T	G	C	G	T	C	T	T	C
2		G	G	A	G	G	C	C	A	A	G	G	T	C	T	G	T	G	C	C	C
3		A	A	G	C	G	G	G	A	G	G	G	T	A	T	G	C	T	T	C	C
4		A	T	C	T	C	G	A	C	C	G	G	T	A	A	G	C	A	A	G	A
5		A	G	G	C	C	T	T	G	C	T	G	T	A	A	G	T	G	A	T	A
6		T	A	C	G	A	G	C	A	A	G	G	T	A	T	A	A	C	T	A	A
7		A	G	T	G	C	A	A	G	A	C	G	T	G	T	G	C	T	A	A	G
8		C	G	T	C	C	A	C	A	G	A	C	G	T	A	T	G	T	C	C	C
9		A	C	A	T	G	A	A	A	A	G	G	T	A	A	C	C	G	C	C	G
Adh2-N consensus								<u>A</u>	A	G	G	T	A	<u>T</u>	G	<u>C</u>					Py
Adh1-S consensus								<u>A</u>	A	G	G	T	A	<u>A</u>	G	T	T				
Animal consensus								<u>A</u>	A	G	G	T	<u>A</u>	A	G	T					
		Adh2 Intron Acceptor Sites																			
Intron																					
1		C	C	C	T	C	T	G	C	A	G	C	C	G	C	C	G	T	G	G	G
2		T	C	T	T	G	T	G	C	A	G	G	G	G	C	C	A	A	A	A	C
3		T	T	G	G	T	C	G	C	A	G	C	C	A	T	C	T	G	G	G	A
4		T	G	T	G	G	T	G	C	A	G	G	C	C	C	T	T	G	G	G	C
5		G	A	T	C	G	A	T	C	A	G	G	C	C	T	G	C	A	G	A	A
6		C	G	T	T	T	T	T	C	A	G	C	T	A	A	A	A	A	A	T	T
7		T	T	G	C	A	T	T	G	A	G	G	T	A	C	T	G	A	T	T	T
8		G	G	A	A	A	T	G	C	A	G	G	G	C	A	C	T	G	G	G	T
9		A	C	G	T	A	C	G	T	A	G	G	A	G	C	T	G	G	A	G	G
Adh2-N consensus						<u>T</u>	<u>G</u>	C	A	G	G										
Adh1-S consensus						G	T	Py	C	A	G	G									
Animal consensus						<u>T</u>		<u>C</u>	A	G	G	<u>G</u>	<u>G</u>								

sequence GCAATGGC, identical to the sequence in Adh1. Since this conforms closely to the consensus sequence for translation initiation in eukaryotes (20), it is likely that it is also the translation start for ADH2. Formal proof that the two processed polypeptides are identical in length awaits the determination of the N-terminal amino acid in each but the presence of stop codons in this reading frame between this ATG and the next upstream ATG precludes a more 5' N-terminus for ADH2.

S_1 mapping with the Adh2 mRNA indicates that transcription initiates 126 bp upstream of the translation start, giving a 5' untranslated region 26 bp longer than for Adh1. At the transcription start there are regions of

homology - the sequence of which the longest is TCCTCAC. 36 bp 5' of the start of transcription there is a presumptive TATA box sequence (CTATATATA) (Fig. 4) which conforms to the plant consensus sequence $\begin{smallmatrix} C & & & & T \\ G & T & A & T & A \end{smallmatrix} A_{1-3} \begin{smallmatrix} C \\ T \end{smallmatrix} A$ (21) but there is no evidence of the CCAAT box like sequence seen in Adh1 or of another sequence postulated to be important in plant gene regulation, $\begin{smallmatrix} C & & & & G & & & & C \\ A & T & & & N & G & A & & C \\ T & A & -5 & T & & & & & T \end{smallmatrix} A_{2-4} TT$ (21).

In the 5' untranslated region of the mRNAs at about 45 bp upstream from the translation start there is a sequence $G \begin{smallmatrix} G \\ T \end{smallmatrix} TCT \begin{smallmatrix} C \\ T \end{smallmatrix} GGAGTGG..ATCGA$ which is shared between the two genes (dots indicate bases present in Adh2 but not Adh1) (Fig. 1). Although the function of the 5' untranslated region of mRNAs is not clear this conserved sequence may be responsible for the selective translation of the two Adh mRNAs under anaerobic conditions.

A comparison of sequences 5' to the transcription start may demonstrate a consensus sequence(s) responsible for anaerobic (and hormone) induction similar to the sequence responsible for the Drosophila heat shock response (22). Five hundred bases 5' to the transcription start of the Adh1 gene (19) were compared to 300 bp 5' to the Adh2 gene transcription start using a dot matrix analysis. Very little homology between the two sequences was seen. With the criteria that there should be at least 7 bp of homology between the two sequences and that the conserved regions should be in roughly equivalent positions (± 100 bp) relative to the transcription start site, a few potential regulatory sequences were identified. The sequence AAATCCTA occurs at -410 bp in Adh1 and -280 bp in Adh2 and another 8 bp sequence (ATCCGAGC) at -300 bp for Adh1 and -130 bp for Adh2. Another possible regulatory sequence is an 8 bp sequence (CACCTCCC) which is at a similar position in the two genes [at -200 bp in Adh1 and -170 bp in Adh2]. There is also an 11 bp conserved sequence (ACCACTATATA) which includes the TATA box. The 11 bp perfect homology can be extended downstream to ACCACTATATA $\begin{smallmatrix} T \\ A \end{smallmatrix}.CAG$ (the dot indicates a base present in Adh1 but not Adh2). To determine whether any of these sequences confer the properties of hormone or anaerobic induction on the Adh genes, a functional analysis of the genes will be required.

3' region of transcript. S_1 mapping has shown the slow and fast variants of Adh1 (Adh1-1S and Adh1-1F) have multiple polyadenylation sites (4 and 7 respectively), with no obvious consensus sequence in the 50 bp upstream of these sites (19). In Adh2, there is a single polyadenylation site 315 bp from the end of the coding region, the S_1 mapping agreeing exactly with the 3' untranslated region seen in the Adh2 cDNA clone

Fig. 5 Length of the mRNA from *Adh2*. RNA was prepared from two maize lines, BKS (*Adh1-S*, *Adh2-N*) and MS904 (*Adh1-S*, *Adh2-O*) subjected to Northern analysis using as a probe nick-translated pZML793 (*Adh1*-probe, left 2 tracks) or pZML841 (*Adh2*-probe, right 2 tracks) under high stringency.

(pZML841) which has a polyA tail. Fifteen bases upstream from the polyA tail there is a sequence AATAAT which is similar to the animal consensus sequence AATAAA (17) and may be significant in RNA processing. In both *Adh1* and *Adh2* genes there is a AATAAA sequence about 200 bp upstream of the polyA tail, close to the stop codon, but this is not a functional polyA addition recognition signal.

Expression of the genes

If the complete *Adh1* or *Adh2* cDNA clone is used as a probe, conditions

Fig. 6 Anaerobic induction of *Adh2* mRNA. Total RNA was isolated from maize roots following various times of anaerobic treatment and subjected to Northern analysis. The same filter was hybridized to nick translated *Adh1* specific probe (pZML133) and *Adh2* specific probe (pZML841R).

1	.STAGKVIKCAAVLWEEKKPFSEIEVEVAPPKAHEVRIKMVATGICRSD	49
1	MATAGKVIKCAAVAWAEGKPLSIEIEVEVAPQAMEVRKILFTSLCHTD	50
1	MATAGKVIKCAAVTWEAGKPLSIEIEVEVAPQAMEVRKILYLTALCHTD	50
50	DHVVSGTLVTP.LPVIAGHEAAGIVESIGEGVTTVRPGDKVIPFTQCG	98
51	VYFWEAKGQTPVFPRIFGHEAGGIIESVGEVTDVAPGDHVLVFTGECK	100
51	VYFWEAKGQTPVFPRI LGHEAGGIVESVGEVTDVAPGDHVLVFTGECK	100
99	KCRVCKHPEGNFCLKNDLSMPRGT.MQDGTSRFTCRGKPIHHFLTSTFS	147
101	ECAHCKSAESNMCDLLRINTDRGVMADGKSRFSINGKPIYHFVGTSTFS	150
101	ECAHCKSEESNMCDLLRINVDRGVMIGDKSRFTISGQPIHFHVGSTFS	150
148	QYTVVDEISVAKIDAASPLEKVCLIGCGFSTGYSAVKVAKVTQGSTCAV	197
151	EYTMVHVGCVAKINPQAPLDKVCVLSGISTGLGASINVAKPPKGSVAV	200
151	EYTVIHVGCLAKINPEAPLDKVCILSGISTGLGATLNVAKPAKGSVAV	200
198	FGLGGVGLSVIMGCKAAGAARIIGVDINKDKFAKAKEVGATECVNPQDYK	247
201	FGLGAVGLAAAGGARIAGASRIIGVDLNPSRFEEARKFGCTEFVNPQDHN	250
201	FGLGAVGLAAMEGARLAGASRIIGVDINPAKYEQAKKFGCTEFVNPQDHD	250
248	KPIQEVLTMSNGGVDFSFEVIGRLDTMTALSCCQEAYGVXVIVGVPD	297
251	KPVQEVLAEMTNGGVDRSVECTGNINAMIQAFECVHDGMGVAVLVGVPKH	300
251	KPVQEVLIETLNGGVDRSVECTGNVNAMISAFECVHDGMGVAVLVGVPKH	300
298	SQNLSMNPMLLLSGRTWRGATFGGFKSKDSVPKLVADFMAKKFALDPLIT	347
301	DAEFKTHPMNFINERTLKGTFFGNYKPRTDLPNVVELYMKKELEVEKFIT	350
301	DDQFKTHPMNFLSEKTLKGTFFGNYKPRTDLPNVVEMYMKKELEVEKFIT	350
348	HVLPFEKINEGFDLLRSGESIRTILTF...	374
351	HSVPFAEINKAFDLMKGGGIRCIIRMEN	380
351	HSVPFSEINTAFDLMKGESLRICIRMED	380

Fig. 7 Comparison of amino acid sequence between horse liver ADH and maize ADH1 and ADH2. The standard one letter amino acid code is used. Top line - horse liver ADH; Middle line - maize ADH1; Bottom line - maize ADH2. Numbers indicate the amino acid residues in the respective enzymes.

can be adjusted (50°C, 3 X SSC/50% formamide) to eliminate cross-hybridization and provide specific probes. The *Adh2* cDNA clone pZML841 was hybridized to RNA extracted from anaerobically treated maize roots under stringency conditions almost entirely specific for *Adh2* hybridization. A single mRNA species of length 1750 bases was observed (Fig. 5). This length compares with an mRNA of 1650 bases seen for *Adh1-1S* and the doublet of 1650 and 1750 bases seen for *Adh1-1F* (10). Northern analysis of genetically defined stocks also verified that the clone pZML841 was derived from an *Adh2* gene (Fig. 5). The line MS904 containing the *Adh2-33* null allele was known not to accumulate any translatable *Adh2* mRNA as measured by *in vitro* translation (3) and our analysis could not detect any *Adh2* mRNA in this line or other lines containing the *Adh2-33* allele.

ADH2 is one of the set of polypeptides induced by anaerobiosis (8). To

determine if the induction was associated with an increase in mRNA levels, we extracted RNA from roots following different periods of anaerobiosis and hybridized the RNA to pZML841R (a 3' Adh2-specific clone). The mRNA for Adh2 is induced about fifty-fold by anaerobic conditions (Fig. 6) a level similar to that for Adh1 (9). In this experiment the same filter was hybridized again to reveal Adh1 mRNA. The Adh1 and Adh2 probes were of approximately equal specific activity so there are roughly equal amounts of Adh1 and Adh2 mRNA in the root cells at 7-10 hours. The level of the Adh2 mRNA decreases after 10 hours whereas Adh1 mRNA remains at a plateau level for at least 48 hours.

Sequence of the ADH1 and ADH2 polypeptides

The coding sequences of Adh1 and Adh2 are 82% homologous and the two amino acid sequences are 87% homologous (Fig. 7, Table 1). The amino acid sequence of the ADH2 polypeptide predicted from the nucleotide sequence of the cDNA and genomic clones is identical in length to the ADH1 polypeptide, excluding any possible N-terminal processing. One of the most noticeable changes between ADH1 and ADH2 is the total charge difference between the two polypeptides. ADH2 has a total of 5 negative charges more than ADH1-F, in keeping with its more rapid anodal electrophoretic mobility (2).

Maize ADH1 is 50% homologous to horse liver ADH (9) and analysis of the sequence differences shows that all structurally important residues are either identical or conservatively substituted, so the 3D structure for the horse liver enzyme determined by X-ray crystallography (23) is a good approximation for ADH1 (9, 24). Since the horse liver enzyme is also 50% homologous to ADH2 (Fig. 7) it provides a basis for the ADH2 3D structure. Using the horse liver 3D structure we can identify the residues involved in the catalytic domain of the enzyme (1 to 175 and 319-374), the co-enzyme binding domain (176-318) and sites of subunit interaction. The amino acids important in the binding of zinc (cys 47, his 69 and cys 178) are conserved between ADH1 and ADH2 sequences, as are the four cysteine ligands (cys 99, 102, 105, 113) for the second zinc atom. These sequences are also conserved between ADH1 and horse liver ADH.

The first 150 amino acids which make up the catalytic domain of the enzyme are well conserved between ADH1 and ADH2 (139 out of 150 are identical). There are a number of changes clustered around amino acid 230 which is in the middle of the co-enzyme binding domain.

Within the first mononucleotide binding site in the co-enzyme binding domain (residues 179-252), at amino acid 192 there is a proline (ADH1) to

alanine (ADH2) change which could have considerable effects on the conformation of the enzyme and possibly on activity. The differences in charge (see earlier) are spread throughout the enzyme in both the catalytic and co-enzyme binding domains and probably also have an affect on activity.

Overall the amino acid changes are sufficient in number and type to explain the differences in electrophoretic mobility and catalytic properties between ADH1 and ADH2.

DISCUSSION

Because the two Adh genes of maize are so similar in nucleotide sequence it is probable that they arose by duplication of an ancestral gene. Apart from coding sequence stability, Adh1 and Adh2 show conservation in the number and position of introns. All monocots and most dicots examined to date have at least two Adh genes (25), so it seems likely that the initial Adh gene duplication took place before divergence of these two major plant groups which occurred at least 120 million years ago. If these assumptions are correct then the levels of change in amino acids (13%) and nucleotides (18%) in the coding region over 120 million years is lower than the 0.2-0.3% per million years which is quoted for mammalian nuclear genes (26). Cloning of both Adh genes of other monocots and dicots and sequence comparison both between and within species should prove whether the gene duplication did predate the monocot/dicot separation. The time of the duplication event is such that the 5' and 3' flanking regions of the two genes, which must have been identical, have now fully diverged. The 3' region and polyA addition signals have diverged completely, and the introns and 5' flanking sequences have no sequence identity except for small regions 5' to the coding region of the genes.

Analyses of a number of other eukaryote genes have shown that short nucleotide sequences can be important for the regulation of gene expression. One type of regulation involves sequences in the vicinity of the TATA box since upstream sequences can be deleted without affecting gene expression. An example of such a gene is the rabbit β globin gene where the sequences responsible for its induction in MEL cells are in the immediate vicinity of the TATA box. No more no more than 58 bp of DNA 5' to the transcription initiation site are required for this induction (27). The two maize Adh genes show considerable homology in the region of the TATA box, perhaps indicating that, similarly to the rabbit β globin gene, this region is important in the regulation of Adh.

A second type of control involves sequences upstream from the TATA box where a short nucleotide sequence in the 5' flanking region (normally between positions -20 bp and -300 bp) is important for gene expression. In Drosophila melanogaster the heat shock genes have a common sequence (CTNGAATNTTCTAGA) (22) in the -40 bp to -150 bp region which has been shown by functional analysis to cause the heat inducibility of the genes. Other genes, e.g. the metallothionin genes of rats, also have short nucleotide sequences located in the 5' region responsible for their induction (28). In the Adh genes three 8 bp sequences in the 5' flanking region are conserved. One of these sequences (CACCTCCC) is in a similar position in both genes (-200 bp in Adh1 and -170 bp in Adh2) and it may be important for gene regulation. Until a functional analysis of the genes can be made no definite statements can be made as to the importance of these conserved nucleotide sequences.

Anaerobic induction is associated with an increased level of mRNA for both genes. Initially the time course of induction is similar in the two genes but Adh2 mRNA is not sustained at high levels if anaerobiosis continues. We assume that co-ordinate induction is due to similar control sequences within both genes, but that mRNA stability differs. Since there is selective translation of Adh1 and Adh2 mRNAs under anaerobic conditions, perhaps the sequences conserved in the 5' untranslated region of the two mRNAs are responsible for the recognition of these mRNAs by ribosomes during anaerobiosis.

The amino acid sequences of the two ADH polypeptides are clearly related and the ADH2 enzyme can be divided into catalytic and coenzyme binding domains by analogy with horse liver ADH and maize ADH1. The conservation of intron position and number between the two maize genes confirms the separation of the co-enzyme binding domain from the catalytic domain by introns four and nine (9, 24) and the division of the co-enzyme binding domain into two structurally similar mononucleotide binding units by intron seven as seen in the Adh1 gene. The amino acid changes between ADH1 and ADH2 occur more frequently in the coenzyme binding domain than in the catalytic domain suggesting a different rate of evolution of the two domains. A similar phenomenon is seen between horse liver ADH and maize ADH1 (45% conserved for coenzyme binding vs. 57% for catalytic domains) and also for horse liver and yeast ADHs where the catalytic domain is more highly conserved than the coenzyme binding domain (29). Particular residues are invariant between horse liver ADH, maize ADH1 and maize ADH2, e.g. the

cysteine and histidine residues important in zinc ligand binding, suggesting that these interactions are essential for the catalytic activity of alcohol dehydrogenases.

In summary, the high degree of conservation between the protein and nucleotide sequence in the coding regions of the two Adh genes of maize leads to the conclusion that these genes arose by ancestral gene duplication. Radical changes in sequence are seen in non-coding regions although there are short regions of homology 5' to the coding sequence. These conserved regions may be the sequences responsible for co-ordinate induction by anaerobiosis or by the hormone 2,4D, for tissue specific expression of the genes and for mRNA recognition by anaerobic ribosomes, but their functional identification awaits the development of a gene expression system for monocots.

ACKNOWLEDGEMENTS

Sequencing of portion of the Adh2 cDNA was performed by Dr Rob Ferl of the University of Florida while a visitor in Canberra. We are grateful for excellent technical assistance from Yvonne Hort, Fiona Clay, Georgina Koci, Janice Norman and Kathryn Ferguson. Danny Llewellyn provided helpful discussion.

REFERENCES

1. Schwartz, D. (1971) *Genetics*, **67**, 411-425.
2. Dlouhy, S.R. (1980) Ph.D. Thesis, Indiana University.
3. Ferl, R.J., Brennan, and Schwartz, D. (1980) *Mol.Gen.Genet.* **18**, 681-692.
4. Freeling, M. and Schwartz, D. (1973) *Biochem.Genet.* **8**, 27.
5. Freeling, M. (1974) *Biochem.Genet.* **12**, 407-417.
6. Hageman, R. and Flesher, D. (1960) *Arch.Biochem. and Biophys.* **87**, 203-209.
7. Freeling, M. (1973) *Molec.gen.Genet.* **127**, 215-217.
8. Sachs, M.M., Freeling, M. and Okimoto, R. (1980) *Cell* **20**, 761-767.
9. Dennis, E.S., Gerlach, W.L., Pryor, A.J., Bennetzen, J.L., Inglis, A., Llewellyn, D., Sachs, M.M., Ferl, R.J., Peacock, W.J. (1984) *Nucleic Acids Research* **12**, 3983-4000.
10. Gerlach, W.L., Pryor, A.J., Dennis, E.S., Ferl, R.J., Sachs, M.M. and Peacock, W.J. (1982) *Proc.Natl.Acad.Sci.USA* **79**, 2981-2985.
11. Osterman, J.C. and Schwartz, D. (1981) *Genetics* **99**, 267-73.
12. Rimm, D.L., Horness, D., Kucera, J. and Blattner, F.R. (1980) *Gene* **12**, 301-305.
13. Benton, W.D. and Davis, R.W. (1977) *Science* **196**, 180-182.
14. Maxam, A.M. and Gilbert, W. (1980) In "Methods of Enzymology", Vol.65 (L. Grossman and K. Moldave, eds.) Academic Press, New York, 499-560.
15. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc.Natl.Acad.Sci.USA* **74**, 5463-5476.
16. Messing, J. and Vieira, J. (1982) *Gene* **19**, 269-276.

17. Proudfoot, N.J. and Brownlee, G.G. (1976) *Nature* **263**, 211-214.
18. Johns, M.A., Strommer, J.N., and Freeling, M. (1984) *Genetics* (in press).
19. Sachs, M.M., Dennis, E.S., Gerlach, W.L. and Peacock, W.J. (1984) (In preparation).
20. Kozak, M. (1981) *Nucl. Acids Res.* **9**, 5233-5252.
21. Messing, J., Geraghty, D., Heidecker, G., Hu, N-T., Kridl, J. and Rubenstein, I. (1983) In "Genetic Engineering of Plants" (T. Kosuge, C.P. Meredith and A. Hollaender, eds.) Plenum Publishing Corp., 211-277.
22. Pelham, H.R.B. (1982) *Cell* **30**, 517-528.
23. Eklund, H., Nordström, B., Zeppenbauer, E., Söderlund, G., Ohlsson, I., Boiwe, T., Söderberg, B-O., Tapia, O. and Bränden, C-I. (1976b), *J.Mol.Biol.* **102**, 27-59.
24. Bränden, C., Eklund, H., Cambillau, C. and Pryor, A.J. (1984) *EMBO J.* **3**, 1307-1310.
25. Gottlieb, L.P. (1982) *Science* **216**, 273-276.
26. Kimura, M. (1983) *The Neutral Theory of Molecular Evolution* (M. Nei and R.K. Koehn, eds.), Sinauer Associates, Sunderland, Massachusetts.
27. Wright, S., Rosenthal, A., Flavell, R. and Grosveld, F. (1984) *Cell* **38**, 265-273.
28. Brinster, R., Chen, H.J., Warren, R., Sarthy, A. and Palmiter, R.D. (1982) *Nature*, **296**, 39-42.
29. Eklund, H., Bränden, C-I. and Jornuall, H. (1976a) *J.Mol.Biol.* **102**, 61-73.