Volume 14 Number 21 1986 Nucleic Acids Research

The alpha-amylase gene in *Drosophila melanogaster:* nucleotide sequence, gene structure and expression motifs

Poppo H.Boer and Donal A.Hickey

Department of Biology, University of Ottawa, Ottawa, Ontario K1N 6N5, Canada

Received 7 August 1986; Revised and Accepted 30 September 1986

ABSTRACT

We present the complete nucleotide sequence of a Drosophila α-amylase gene and its flanking regions, as determined by cDNA and genomic sequence analysis. This gene, unlike its mammalian counterparts, contains no introns. Nevertheless the insect and mammalian genes share extensive nucleotide similarity and the insect protein contains the four amino acid sequence blocks common to all α-amylases. In Drosophila melanogaster, there are two closely-linked copies of the α-amylase gene and they are divergently transcribed. In the 5'-regions of the two gene-copies we find high sequence divergence, yet the typical eukaryotic gene expression motifs have been maintained. The 5'-terminus of the α-amylase mRNA, as determined by primer extension analysis, maps to a characteristic Drosophila sequence motif. Additional conserved elements upstream of both genes may also be involved in amylase gene expression which is known to be under complex controls that include glucose repression.

INTRODUCTION

Amylase enzymes are widely distributed in nature and genes coding for α-amylases have been cloned and sequenced from a variety of organisms, including mammals (1, 2), plants (3), fungi (4), and bacteria (5, 6). However, no sequence information is currently available for any invertebrate α-amylase gene; the Drosophila sequence presented here fills this gap. Drosophila α-amylase genes are of particular interest because (i) a large array of electrophoretic variant proteins are observed in nature (7, 8), (ii) the gene is duplicated (9, 10), and (iii) there is a complex set of gene regulatory elements controlling amylase expression (11, 12). Moreover, Drosophila amylase genes are unusual among higher eukaryotic genes in that their expression is glucose repressible (13).

In Drosophila melanogaster there are two closely-linked copies of the α-amylase gene (9, 10). Both copies are presumed to be actively expressed since many strains produce a duplicated banding pattern on enzyme activity gels (7, 8). Here we report α-amylase sequences from the two common

wild-type strains; Oregon-R and Canton-S. In the Oregon-R strain a single isozymic band AMY1 is observed, while in the Canton-S strain there are two isozyme bands, AMY1 and AMY3, the most abundant of which comigrates with the Oregon-R band (10). We have sequenced regions of genomic clones encoding the two Amy genes in the Oregon-R strain, as well as derived cDNAs; we also analyzed cDNA and genomic Canton-S clones. This allows us to assess sequence divergence between strains at a given locus, and also the divergence between gene copies within a strain. A comparison of the sequences of the duplicated genes reveals that the coding regions are very similar, but not identical, whereas much lower levels of sequence homology are observed in the upstream non-coding regions. Both Amy genes contain typical transcriptional regulatory motifs located at the appropriate positions relative to the site of transcription initiation. Our primer extension experiments support the view (14) that in Drosophila, the initiation of transcription occurs at a conserved sequence motif. Additional conserved upstream elements may play a role in the regulation of amylase gene activity through glucose repression.

MATERIALS AND METHODS

The isolation and characterization of α-amylase cDNA and genomic clones from the Oregon-R and Canton-S strains of Drosophila melanogaster are described in detail elsewhere (15). The cDNA inserts, identified by their homology to the mouse amylase cDNA sequences (1), were transferred into pUC13 plasmids and the restriction maps were compared to the Canton-S genomic map (10). For DNA sequence analysis, subfragments were recovered from low melting point agarose and cloned into suitable M13 vectors (16). DNA sequences were obtained by the dideoxy chain termination technique (17) using LKB Macrophor electrophoresis equipment; the sequences were assembled and analyzed using a sonic digitizer and the Microgenie (Beckman) computer programs (18). Additional sequence information was obtained using a synthetic oligonucleotide, 5'-AGCGATGTCGTCCCACTTCC, as a primer; this oligomer is complementary to amylase mRNA (Fig. 2, positions 110-129).

To determine the position of the 5' terminus of the Drosophila amylase mRNA, the above oligonucleotide primer was annealed with Oregon-R poly A^+ larval RNA and elongated with AMV reverse transcriptase (Life Sciences) in the presence of α-32P-dATP (Amersham) under standard conditions (19). Products were resolved on sequencing gels with appropriate DNA sequence markers.

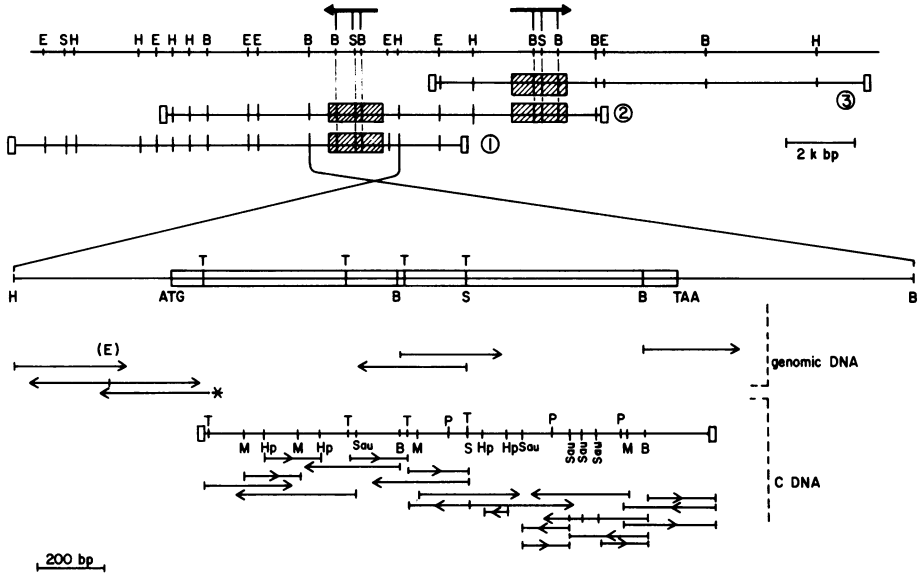
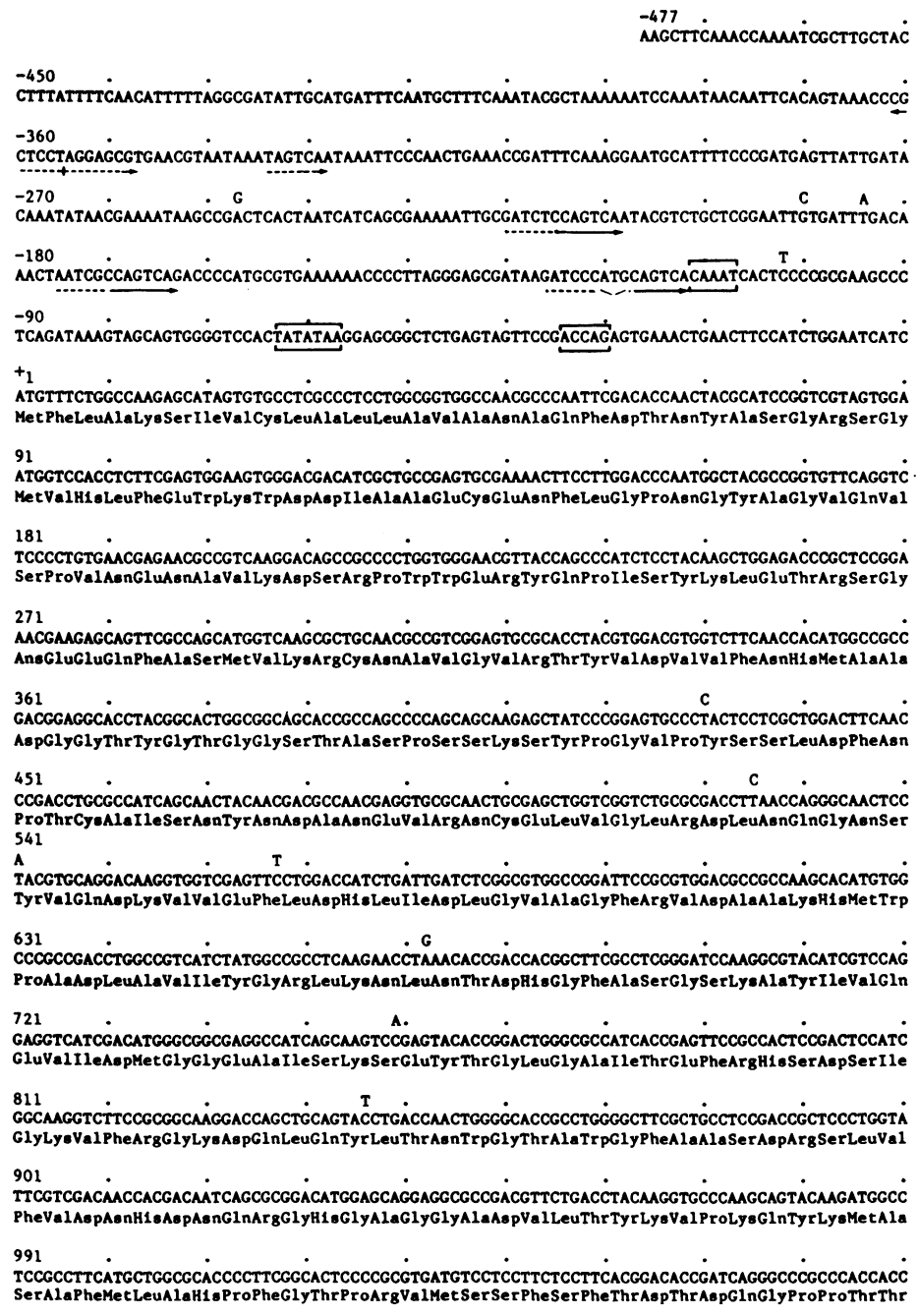


Figure 1. Restriction map of the α-amylase gene region. Overlapping genomic DNA fragments from Oregon-R (2, 3) and Canton-S (1) are indicated with the EcoRI linkers shown as small open boxes. The duplicated α-amylase genes and their direction of transcription are shown as divergent arrows over the B-S-B sites. Sequencing strategy is shown in the lower part of the figure; the open box represents the coding region; arrows indicate direction and extent of sequence determined; (E) is a polymorphic restriction site. The longest Oregon-R cDNA obtained is shown below. B, BamHI; E, EcoRI; H, HindIII; Hp, HinPI; M, MspI; P, PstI; S, SalI; Sau, Sau3A; T, TaqI. Note that the orientation of the "proximal" α-amylase gene (upper) is reversed (lower) and that the genomic sequencing strategy applies to all gene-copies studied. Asterisk marks the position of the oligonucleotide primer used.

RESULTS AND DISCUSSION

Nucleotide Sequence of the α-amylase gene and its predicted protein

The two copies of the α-amylase gene are localized on the Oregon-R restriction map as shown in Fig. 1; they are present in an inverted orientation, about 4 kb apart. Most of the coding sequence was obtained from two overlapping Oregon-R cDNA clones, and it was extended with genomic sequences in order to gain insight into gene-expression signals. Coding regions in genomic clones were then partially sequenced to verify the absence of introns and to investigate sequence polymorphisms. Polymorphisms between the two gene copies allowed the unambiguous assignment of cDNA sequences to the "proximal" gene copy (using the nomenclature of ref. 9), see Fig. 1.


-477
AAGCTTCAAACCAAAATCGCTTGCTAC
-450
CTTTATTTTCAACATTTTTAGGCGATATTGCATGATTTCAATGCTTTCAAATACGCTAAAAAATCCAAATAACAATTCACAGTAAACCCG
-360
CTCCTAGGAGCGTGAACGTAATAAATAGTCAATAAATTCCCAACTGAAACCGATTTCAAAGGAATGCATTTTCCCGATGAGTTATTGATA
-270
CAAATATAACGAAAATAAGCCGACTCACTAATCATCAGCGAAAAATTGCGATCTCCAGTCAATACGTCTGCTCGGAATTGTGATTTGACA
-180
AACTAATCGCCAGTCAGACCCCATGCGTGAAAAAACCCCTTAGGGAGCGATAAGATCCCATGCAGTCACAAATCACTCCCCGCGAAGCCC
-90
TCAGATAAAGTAGCAGTGGGGTCCACTATATAAGGAGCGGCTCTGAGTAGTTCCGACCAGAGTGAAACTGAACTTCCATCTGGAATCATC
+1
ATGTTTCTGGCCAAGAGCATAGTGTGCCTCGCCCTCCTGGCGGTGGCCAACGCCCAATTCGACACCAACTACGCATCCGGTCGTAGTGGA
MetPheLeuAlaLysSerIleValCysLeuAlaLeuLeuAlaValAlaAsnAlaGlnPheAspThrAsnTyrAlaSerGlyArgSerGly
91
ATGGTCCACCTCTTCGAGTGGAAGTGGGACGACATCGCTGCCGAGTGCGAAAACTTCCTTGGACCCAATGGCTACGCCGGTGTTCAGGTC
MetValHisLeuPheGluTrpLysTrpAspAspIleAlaAlaGluCysGluAsnPheLeuGlyProAsnGlyTyrAlaGlyValGlnVal
181
TCCCCTGTGAACGAGAACGCCGTCAAGGACAGCCGCCCCTGGTGGGAACGTTACCAGCCCATCTCCTACAAGCTGGAGACCCGCTCCGGA
SerProValAsnGluAsnAlaValLysAspSerArgProTrpTrpGluArgTyrGlnProIleSerTyrLysLeuGluThrArgSerGly
271
AACGAAGAGCAGTTCGCCAGCATGGTCAAGCGCTGCAACGCCGTCGGAGTGCGCACCTACGTGGACGTGGTCTTCAACCACATGGCCGCC
AnsGluGluGlnPheAlaSerMetValLysArgCysAsnAlaValGlyValArgThrTyrValAspValValPheAsnHisMetAlaAla
361
GACGGAGGCACCTACGGCACTGGCGGCAGCACCGCCAGCCCCAGCAGCAAGAGCTATCCCGGAGTGCCCTACTCCTCGCTGGACTTCAAC
AspGlyGlyThrTyrGlyThrGlyGlySerThrAlaSerProSerSerLysSerTyrProGlyValProTyrSerSerLeuAspPheAsn
451
CCGACCTGCGCCATCAGCAACTACAACGACGCCAACGAGGTGCGCAACTGCGAGCTGGTCGGTCTGCGCGACCTTAACCAGGGCAACTCC
ProThrCysAlaIleSerAsnTyrAsnAspAlaAsnGluValArgAsnCysGluLeuValGlyLeuArgAspLeuAsnGlnGlyAsnSer
541
TACGTGCAGGACAAGGTGGTCGAGTTCCTGGACCATCTGATTGATCTCGGCGTGGCCGGATTCCGCGTGGACGCCGCCAAGCACATGTGG
TyrValGlnAspLysValValGluPheLeuAspHisLeuIleAspLeuGlyValAlaGlyPheArgValAspAlaAlaLysHisMetTrp
631
CCCGCCGACCTGGCCGTCATCTATGGCCGCCTCAAGAACCTAAACACCGACCACGGCTTCGCCTCGGGATCCAAGGCGTACATCGTCCAG
ProAlaAspLeuAlaValIleTyrGlyArgLeuLysAsnLeuAsnThrAspHisGlyPheAlaSerGlySerLysAlaTyrIleValGln
721
GAGGTCATCGACATGGGCGGCGAGGCCATCAGCAAGTCCGAGTACACCGGACTGGGCGCCATCACCGAGTTCCGCCACTCCGACTCCATC
GluValIleAspMetGlyGlyGluAlaIleSerLysSerGluTyrThrGlyLeuGlyAlaIleThrGluPheArgHisSerAspSerIle
811
GGCAAGGTCTTCCGCGGCAAGGACCAGCTGCAGTACCTGACCAACTGGGGCACCGCCTGGGGCTTCGCTGCCTCCGACCGCTCCCTGGTA
GlyLysValPheArgGlyLysAspGlnLeuGlnTyrLeuThrAsnTrpGlyThrAlaTrpGlyPheAlaAlaSerAspArgSerLeuVal
901
TTCGTCGACAACCACGACAATCAGCGCGGACATGGAGCAGGAGGCGCCGACGTTCTGACCTACAAGGTGCCCAAGCAGTACAAGATGGCC
PheValAspAsnHisAspAsnGlnArgGlyHisGlyAlaGlyGlyAlaAspValLeuThrTyrLysValProLysGlnTyrLysMetAla
991
TCCGCCTTCATGCTGGCGCACCCCTTCGGCACTCCCCGCGTGATGTCTCCTTCTCCTTCACGGACACCGATCAGGGCCCGCCCACCACC
SerAlaPheMetLeuAlaHisProPheGlyThrProArgValMetSerSerPheSerPheThrAspThrAspGlnGlyProProThrThr

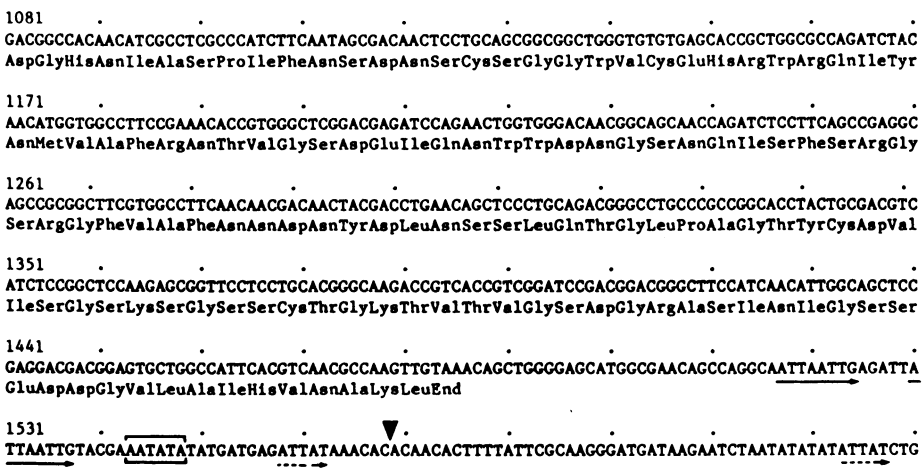


Figure 2. Nucleotide sequence of the Drosophila α-amylase gene region and derived amino acid sequence, numbered from the initiation codon. Differences in Canton-S DNA (-477 to 91 and 179 to 975) are indicated above the Oregon-R sequence; they result in amino acid changes at 430 (His) and 541 (Asn). Regulatory motifs are blocked; repeated sequences are indicated with arrows, the poly A tail is added at C(A)1565 (arrowhead).

For those portions of the transcribed region which were not sequenced from both cDNA and genomic clones, the absence of introns was established by two methods: (i) fine structure restriction mapping of both cDNA and genomic DNA clones (i.e. comigrating Taq 1 and Bam Hl fragments) and, (ii) by the hybridization of amylase cDNA probes to Southern blots of total genomic DNA which was cleaved by tetranucleotide-recognizing restriction enzymes (BanI, HhaI, HaeIII, HinfI, AluI, TaqI, Sau3a, MspI and Sau961) and electrophoresed on polyacrylamide "sequencing" gels.

The nucleotide sequence of the Drosophila α-amylase gene is shown in Fig. 2. The Amy gene encodes a protein of 493 amino acids, of which the N-terminal 18 amino acids likely represent the signal peptide of the secreted protein. Proteolytic cleavage is predicted to occur between Ala-18 and Gln-19, the conserved residues at which processing is known to occur in the mouse amylase precursor protein (20). In the N-terminal regions, the mouse and insect proteins share a single positively-charged residue, namely Lys at positions 2 and 5 respectively; moreover, the motif Leu-Ala-Val, which is found in many Drosophila signal peptides (21), is present at positions 13 to 15 (Fig. 3). The mature α-amylase protein from Drosophila thus is predicted to be 475 amino acids long. Its N-terminus is blocked (unpubl.), presumably

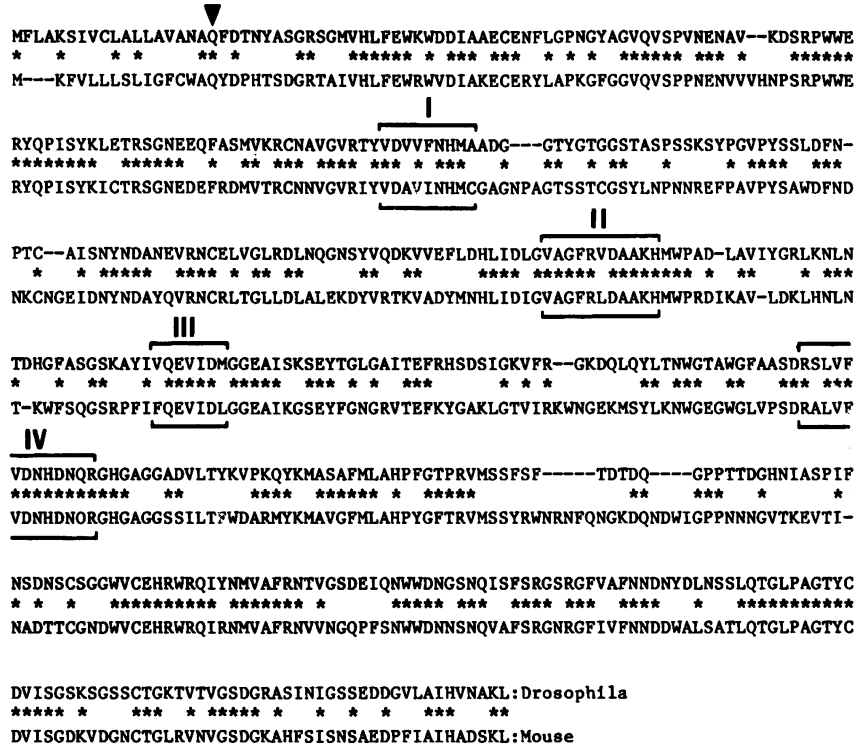


Figure 3. Amino acid alignment of the Drosophila and mouse α-amylase proteins. Identities are marked with asterisks and dashes indicate deletions. The four amino acid sequence motifs common to all α-amylases are blocked. Arrow indicates the presumed N-termini of the mature proteins.

with a derivatized Gln residue, as in the mouse protein (20). The derived amino acid composition of the Drosophila α-amylase protein closely matches experimentally-determined values (unpubl.), with 46% of the residues being polar amino acids. The α-amylase coding region is very GC rich (62.8%) relative to the flanking region (average GC is 40%); this results in a significantly high number of codons with C in the third position (60% of the codons). The codon usage patterns of a number of abundantly expressed Drosophila proteins have been examined (22) and the α-amylase protein follows their general pattern. For example: within the Leu codon family TTA is not used and CTG is preferred over CTC; the codons AAA (Lys) and AGA (Arg) are also not used in the α-amylase gene and, as in most Drosophila genes, GCC is the preferred Ala codon. This high GC content results in additional reading

frames within the Amy gene; a second reading frame of 963 nucleotides (positions 585-1547) is also open; however it lacks the distinctively Drosophila codon usage pattern as well as a typical translational initiation site. The presumptive initiation codon of the Amy gene is the first AUG encountered in progressing from the 5'-end of the mRNA (see below for precise mapping of the 5'-terminus) and the surrounding nucleotides conform to the mammalian consensus (23) (e.g. C, T, A in positions -1 to -3). We conclude from the sequence analysis that the gene possesses the proper signals for translation.

When the GenBank and NBRF protein data banks were searched for sequences homologous to the Drosophila α-amylase gene, significant levels of homology were found with those of the mammalian amylases and considerably lower levels with those of plants (barley) and bacteria (B.subtilis). The nucleotide sequence identity between the Drosophila and mouse coding regions is 57%. Because similarity extends over the the entire coding region, there is no doubt that the insect and mammalian amylases have descended from a common ancestral gene. The amino acid sequence of the mouse pancreatic α-amylase precursor protein is aligned with that of D. melanogaster in Fig. 3, with amino acid matches indicated by asterisks. The mature proteins share 55.4% amino acid identity. This value is high compared to other insect-mammal sequence relationships (cf. 22% amino acid identity between the Drosophila and bovine rhodopsin proteins, ref. 24). Figure 3 shows that there are extensive stretches of complete identity, for example those that are 14, 17 and 16 residues long (at positions 71, 299, 437 respectively). Conserved amino acid sequence motifs which have been identified in other (plant, animal and microbial) α-amylases (25, 26, 27) were noted, and we have blocked these regions in Fig. 3. These motifs are thought to be important for α-amylase function (28). We note that these "conserved blocks" do not correspond exactly to the most highly conserved regions that have been shared between the Drosophila and mouse α-amylase proteins during the 700 million year period since the insect-mammal divergence (29).

Polymorphisms in α-Amylase Genes from Different Drosophila Strains

In addition to determining the sequence of the α-amylase "proximal" gene copy in Oregon-R, we also obtained sequence data for the analogous gene in Canton-S. A cDNA sequence was determined (pos. 179-702, Fig. 2) and it was extended with genomic sequence both upstream (pos. -477 to 91) and in the coding region (pos. 625-909). Again, regions of sequence overlap showed that the cDNA was derived from the "proximal" gene-copy. In the regions compared

(1441 bases), we observed eleven differences relative to the Oregon-R sequence; these are shown in Fig. 2. This level of divergence, approximately one per cent between the Amy genes of the two Drosophila strains is comparable to what was found in the rhodopsin genes of these strains(24). In the upstream region, one of these polymorphisms affects an EcoRI restriction site (Fig. 1), while in the coding region three of the five changes are silent, occurring in the third position of codons. Two changes, however, result in amino acid substitutions between the Oregon-R and Canton-S derived proteins (Tyr to His at position 430 and Tyr to Asn at position 541). Thus these amino acid substitutions would not alter the charge of the protein on enzyme activity gels and this is consistent with the gene product in the two strains having the same electrophoretic mobility. This observation contrasts with the finding for the Adh gene of Drosophila where alleles with the same electrophoretic mobility code for identical polypeptides (30). The indication of greater variability in α-amylase sequences correlates with the generally high levels of α-amylase enzyme variation seen in nature (7, 8). We have analysed two cDNAs from Oregon-R and one from Canton-S and all derive from the "proximal" gene copy; the absence of cDNA clones corresponding to the second gene copy suggests that this gene is less active in these Drosophila strains.

Transcription of the Drosophila α-Amylase Gene

Examination of the DNA sequence upstream of the amylase initiation codon reveals a potential transcription initiation site. At position -35 we find an ACCAG motif that resembles a conserved sequence, ATCA$^{G}_{T}$T$^{C}_{T}$ present at the extreme 5'-end of many Drosophila genes (14, 22). Primer extension analysis confirms that this is also the case for α-amylase mRNAs (Fig. 4, arrowhead). A single discrete band was found and it positions the 5'-terminus at the conserved C in this "APyCAG" motif. Initiation of transcription is thus likely to occur within this box. At -64, that is 29 nucleotides upstream from this site, we find the TATA box and it is preceded by the CAAT pentanucleotide motif some 40 nucleotides further upstream. Thus, the typical eukaryotic promoter elements are present and appropriately spaced; furthermore, the α-amylase CAAT box conforms to the Drosophila consensus A$^{A}_{T}$GCA$^{A}_{T}$N$^{A}_{T}$N (22).

The 3'-non-translated region of the α-amylase mRNA could be determined unambiguously by comparing cDNA and genomic sequences. The polyA tail of the mRNA is added to C(A), position 1565(6) in Fig. 2. Most polyadenylation sites (32) contain the highly conserved hexanucleotide AATAAA at -10 to -30

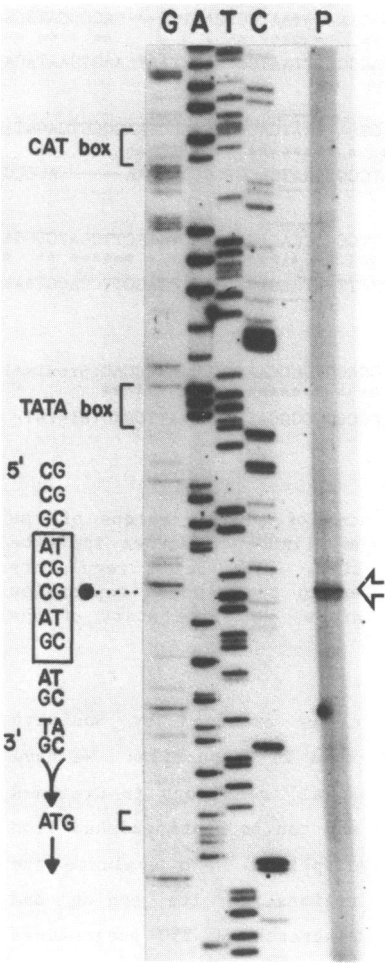


Figure 4. Precise mapping of the 5'-end of the α-amylase mRNA. Primer extension products and a DNA sequence marker, using the synthetic 20-mer described in the text, were resolved on a sequencing gel. The DNA sequence motifs for Amy gene expression are indicated at left and the Drosophila cap site motif is blocked. The primer extension product is shown at right, marked by an arrowhead.

from the A-tail, as well as other structural motifs, sometimes including a less conserved TG cluster (33). We observe an AATATA motif at positions 1540-1545, preceded by eight-nucleotide repeats. We conclude that the Drosophila α-amylase mRNA has short non-coding regions: a leader of 33 nucleotides and a non-coding tail of 86 nucleotides for a total length of 1598 nucleotides, plus the added polyA tail. This is consistent with an estimated messenger length of 1.65 kb from Northern blot analysis (15).

Alpha-Amylase Gene Duplication: Conservation of Coding Sequence and Regulatory Motifs

The α-amylase gene is duplicated in many strains of D. melanogaster (9,

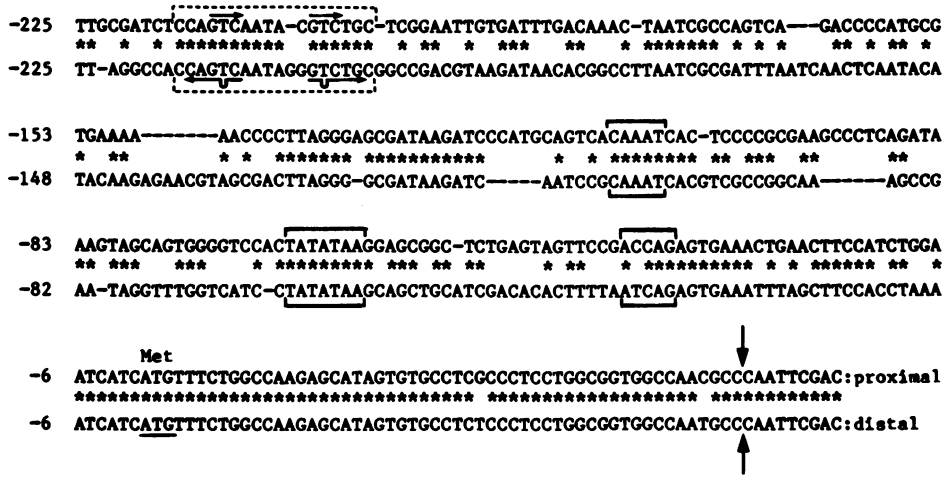


Figure 5. Comparison of the nucleotide sequences of the 5'-regions of the two α-amylase gene-copies from Oregon-R. In the alignment, dashes indicate nucleotide deletions and asterisks show identity. The boxed regulatory motifs and the dotted box, which may be related to the glucose repression effect, are discussed in the text. Vertical arrows mark the start of the sequence for mature gene products.

unpubl.) and gene organization and structure as examined by Southern hybridization studies appear to be highly conserved in Drosophila. We have direct evidence for the expression of the "proximal" gene copy in Oregon-R since the genomic and cDNA copies match exactly and can be distinguished from the second gene copy by the presence of polymorphisms. To evaluate the status of the "distal" copy, we examined regions of its coding and 5'-flanking sequences. Within a gene-internal stretch of 350 nucleotides (corresponding to positions 625-975, Fig. 2) we observed two nucleotide substitutions. These are silent third position changes; T and C at positions 846 and 899 respectively. The former is also polymorphic between the "proximal" gene copies of Oregon-R and Canton-S (see Fig. 2). Additional sequence was obtained using the primer mentioned above. Here, comparing the coding regions of the two gene-copies, we found two differences. A first position codon change (pos. 31, Fig. 3) generates an Ala-Ser difference in the signal peptide and a silent change occurs at position 51.

The 5'-non-coding regions of the two Amy genes are much less conserved; an alignment is shown in Figure 5. In the region preceding the TATA boxes they were found to share only 57% nucleotide identity. Among the sequence motifs that are strictly conserved between the upstream regions of the two

genes are the transcription initiation site, APyCAGagtgaaa, the classical TATA box, cTATATAAg, and the CAAT motif, CAAATcac; the nucleotides conserved within Drosophila genes are boxed in Fig. 5. Thus, typical RNA polymerase II signals are conserved at the appropriate positions, although their exact location relative to the site of transcription initiation varies. The observations suggest that both copies of the genes are transcriptionally active in the Oregon-R strain. The sequence alignment in Fig. 5 also predicts that only a single primer extension product will be obtained from mRNAs transcribed from the two gene copies; this is, in fact, observed (Fig. 4).

Since the two Drosophila Amy genes, unlike most higher eukaryotic genes, are glucose repressible (13), we were interested in comparing their common upstream sequences with those of microbial glucose repressible genes. In addition to the shared TATA and CAAT motifs, we noted a region that shows significant homology (21 out of 29 nucleotides) with the 5'-sequence (-190 region) of the glucose-repressible Adh III gene of yeast (33). The yeast and Drosophila genes share GGCCAC$\overset{agg}{\underset{-c}{}}$AGTCAA$_t^-AGG_g^c$TPyT$_g^tCG_g^-$CC and in the Amy 5'-regions, beginning at position -216, 16 out of 18 nucleotides are identical (dotted box in Fig. 5). This region contains the hexanucleotide CAGTCA, that is repeated (Fig. 2). This is of interest because in yeast upstream regions that possess (imperfectly) repeated sequence motifs have been implicated in glucose repression of, for example, iso-1-cytochrome c (34) and Adh II (35). It is also known that certain Drosophila regulatory elements are recognized in yeast (e.g., heat shock elements, ref. 36). Nevertheless, confirmation of the significance of the Amy gene flanking sequence motifs must await their functional analysis. We are currently investigating the involvement of cis-acting elements in the glucose repression of the Amy gene using the P element-mediated embryo transformation system of Drosophila.

In summary, the data presented here provide the first sequence information for Drosophila α-amylase genes. Amylase coding sequences and their predicted proteins are highly conserved among vertebrates and invertebrates. However, in contrast to the mammalian α-amylase genes, those in Drosophila contain no introns: within the animal kingdom Amy introns are optional. Two copies of the α-amylase gene, both apparently functional, are present in D. melanogaster, including strains that produce a single isozymic form of amylase. Coding sequences are highly conserved between the duplicated genes, whereas many substitutions are observed in the non-coding regions. The organization of the coding sequences indicates that the

duplicated genes are divergently transcribed, their promoters being approximately 3.7 kb apart. A number of conserved motifs in the upstream regions are observed and will be of value in determining cis-acting elements involved in the regulation of amylase gene expression. Of particular interest is the possibility that controls involved in glucose repression may be shared between higher eukaryotic genes and glucose-repressible microbial genes.

ACKNOWLEDGEMENTS

We thank Drs. T. Maniatis, V. Pirrotta and L. Kauvar for the Drosophila DNA libraries, and B. Benkel for help with the primer extension experiments. We are grateful to Dr. M. Yaguchi (N.R.C., Ottawa) for performing the amino acid analysis of purified α-amylase and to Dr. L. Bonen for a critical reading of the manuscript. Excellent technical assistance was provided by Yves Genest. This work was supported by grants from MRC Canada and NSERC Canada.

REFERENCES

1. Hagenbüchle, O., Bovey, R., Young, R.A. (1980) Cell 21, 179-187.
2. Nakamura, Y., Ogawa, M., Nishide T., Emi, M., Kosaki, G., Himeno, S., and Matsubara, K. (1984) Gene 28: 263-270.
3. Rogers, J.C. and Milliman, C. (1983) J. Biol. Chem. 258: 8169-8174.
4. Toda, H., Kondo, K. and Narita, K. (1982) Proc. Japan. Acad. 58: 208-212.
5. Takkinen, K., Pettersson, R.F., Kalkkinen, N., Palva, I., Soderlund, H. and Kaariainen, L. (1983) J. Biol. Chem. 258: 1007-1013.
6. Young, M., Galizzi, A. and D. Henner (1983) Nucl. Acids Res. 11: 237-249.
7. Hickey, D.A. (1979) Genetica 51: 1-4.
8. Singh, R.S., Hickey, D.A. and J. David (1982) Genetics 101: 235-256.
9. Gemmill, R.M., Schwartz, P.E. and Doane, W.W. (1985) Nucl. Acids Res. 14: 5337-5352.
10. Levy, J.N., Gemmill, R.M. and Doane, W.W. (1985) Genetics 110: 313-324.
11. Doane, W.W., Treat-Clemons, L.G., Gemmill, R.M., Levy, J.N., Hawley, S.A., Buchberg, A. and Paigen, K. (1983) In: Isozymes: Current Topics in Biological and Medical Research, Alan R. Liss, New York, Vol. 9, pp. 63-90.
12. Hickey, D.A. and Benkel, B.F. (1987) CRC Crit. Rev. Biotechnol. Vol. 5(2) in press.
13. Benkel, B.F. and D.A. Hickey (1986) Genetics, 114: 137-144.
14. Hultmark, D., Klemenz, R. and Gehring, W.J. (1986) Cell 44: 429-438.
15. Benkel, B.F., Abukashawa, S., Boer, P.H. and Hickey, D.A. (1986) Can. J. Genet. Cytol. manuscript submitted.
16. Messing, J. (1983) Methods Enzymol. 101: 20-78.
17. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. (USA) 74: 5463-5467.
18. Queen, C. and Korn, L.J. (1984) Nucl. Acids Res. 12: 581-599.

19. Williams, J.G. and Mason, P.J. (1985) In Hames, B.D. and Higgins, S.J. (eds), Nucleic Acid Hybridization, IRL Press, Oxford, pp. 139-160.
20. Karn, R.C., Petersen, T.E., Hjorth, J.P., Nicles, J.T. and Roepstorff, P. (1981) FEBS Lett. 126: 292-296.
21. Henikoff, S., Keene, M.A., Fechtel, K., and Fristrom, J.W. (1986) Cell 44: 33-42.
22. O'Connell, P. and Rosbash, M. (1984) Nucl. Acids Res. 12: 5495-5513.
23. Kozak, M. (1984) Nucl. Acids. Res. 12: 857-872.
24. Zuker, C.S., Cowman, A.F., and Rubin, G.M. (1985) Cell 40: 851-858.
25. Rogers, J.C. (1985) Biochem. Biophys. Res. Comm. 128: 470-476.
26. MacKay, R.M., Baird, S., Dove, M.J., Erratt, J.A., Gines, M., Moranelli, F., Nasim, A., Willick, G.E., Yaguchi, M. and Seligy, V.L. (1985) BioSystems 18: 279-292.
27. Nakajima, R., Imanaka, T. and Aiba, S. (1986) Appl. Microbiol. Biotechnol. 23: 355-360.
28. Matsuura, Y., Kusunoki, M., Harada, W. and Kakudo, M. (1984) J. Biochem. 95: 697-702.
29 Moore, G.W., Goodmann, M., Callahan, C., Holmquist, R., and Moise, H. (1976) J. Mol. Biol. 105: 15-37.
30. Kreitman, M. (1983) Nature 304: 412-417.
31. Birnstiel, M.L., Busslinger, M. and Strub, K. (1985) Cell 41: 349-359.
32. Nussinov, R. (1986) Nucl. Acids Res. 14: 3557-3571.
33. Young, E.T. and Pilgrim, D. (1985) Mol. Cell. Biol. 5: 3024-3034.
34. Guarente, L., Lalonde, B., Gifford, P. and Alani, E. (1984) Cell 36: 503-511.
35. Shuster, J., Yu, J., Cox, D., Chan, R.V.L., Smith, M., and Young, E. (1986) Mol. Cell. Biol. 6: 1894-1902.
36. De Banzie, J.S., Sinclair, L. and Lis, J.T. (1986) Nucl. Acids Res. 14: 3587-3601.