The EMBO Journal vol.3 no.12 pp.2809–2816, 1984

Multiple non-allelic genes encoding pancreatic alpha-amylase of mouse are expressed in a strain-specific fashion

M.Tosi[1], R.Bovey, S.Astolfi[2], S.Bodary, M.Meisler[3] and P.K.Wellauer

Swiss Institute for Experimental Cancer Research, 1066 Epalinges, Switzerland
Present addresses: [1]Department of Immunology, Institut Pasteur, 25, rue du Dr Roux, 75724 Paris Cedex 15, France, [2]Departmento de Biologia Celular, Universidade de Brasilia, Brasilia, DF 70910 Brasil and [3]Department of Human Genetics, University of Michigan, Ann Arbor, MI 48109, USA

Communicated by P.K.Wellauer

The number of active *Amy*-2 genes has been estimated in strain CE/J mice which produce four distinct electrophoretic forms of alpha-amylase in their pancreas. cDNA cloning and DNA sequence analysis discloses five distinct mRNA sequences which differ by ~1% of their nucleotides. Two of these mRNAs specify the same protein. Changes in the nucleotide sequences result in amino acid replacements that alter the net charges of the deduced proteins. This has allowed a tentative assignment of individual mRNAs to isozymes detected by electrophoresis. Quantitative Southern blot hybridization using a DNA probe specific for the first exon of *Amy*-2 reveals the presence of > 10 *Amy*-2 related sequences per haploid CE/J genome. Models which could account for the mouse strain-specific differences with respect to the number of pancreatic alpha-amylase isozymes and their variable but genetically determined quantitative ratios are discussed.

Key words: electrophoretic protein variants/mRNA sequences/gene number/expression of multigene family

Introduction

The two genetic loci *Amy*-1 and *Amy*-2, which are closely linked on chromosome 3 of mouse, carry the determinants for parotid and pancreatic alpha-amylase, respectively (Eicher and Lane, 1980). Surveys of laboratory strains and of wild mouse populations have shown that qualitative and quantitative variants exist for both the parotid and the pancreatic form of alpha-amylase (Nielsen and Sick, 1975). Most inbred strains of mouse, including the A/J strain, express a single isozyme species in the pancreas. Multiple electrophoretic forms of pancreatic alpha-amylase have been detected in mice of inbred strains such as CE/J and YBR and found to differ in amino acid sequence (Hjorth *et al.*, 1980; Bloor *et al.*, 1981; Strahler and Meisler, 1982). Models assuming expression of multiple structural genes have been proposed to explain the qualitative and quantitative interstrain differences of pancreatic alpha-amylase isozymes (Bloor *et al.*, 1981; Owerbach *et al.*, 1981; Nielsen, 1982; Meisler *et al.*, 1983).

The pancreatic, the parotid gland and the liver alpha-amylase mRNAs of strain A/J mice have been compared at the nucleotide sequence level using cDNA clones specific for the corresponding tissues (Schibler *et al.*, 1980; Hagenbüchle *et al.*, 1980, 1981; Tosi *et al.*, 1981). The genes specifying the various alpha-amylase mRNA species have been isolated and characterized. *Amy*-1, which specifies both the parotid gland the liver-type mRNAs, is a single copy gene (Young *et al.*, 1981). In contrast, *Amy*-2 exists as multiple, very similar copies in the genome of A/J mice (Schibler *et al.*, 1982; Hagenbüchle *et al.*, 1984). The total number of transcriptionally active *Amy*-2 genes has not so far been determined for any inbred strain of mouse.

We determined and compared the nucleotide sequences of five distinct alpha-amylase mRNAs. This demonstrates that at least five different *Amy*-2 genes are expressed in the pancreas of CE/J mice.

Results

Homozygous mice of the inbred strain CE/J produce multiple electrophoretic variants of pancreatic alpha-amylase. An example is shown in Figure 1 which compares the electrophoretic profiles of alpha-amylases present in strains CE/J and A/J. High resolution non-denaturing polyacrylamide gel electrophoresis reveals a single species in A/J and four distinct forms of CE/J. The separation into two electrophor-

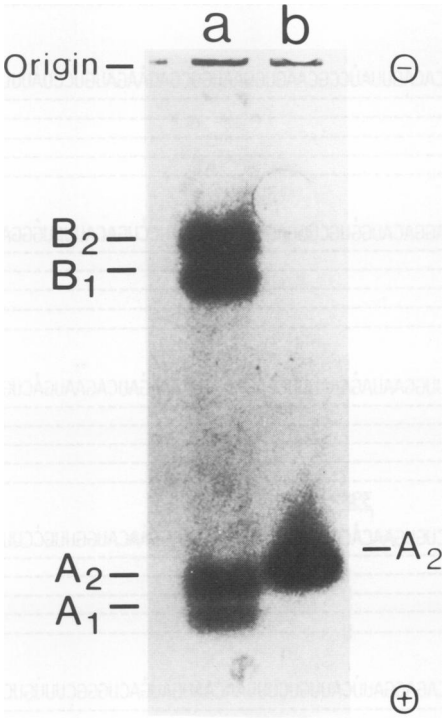
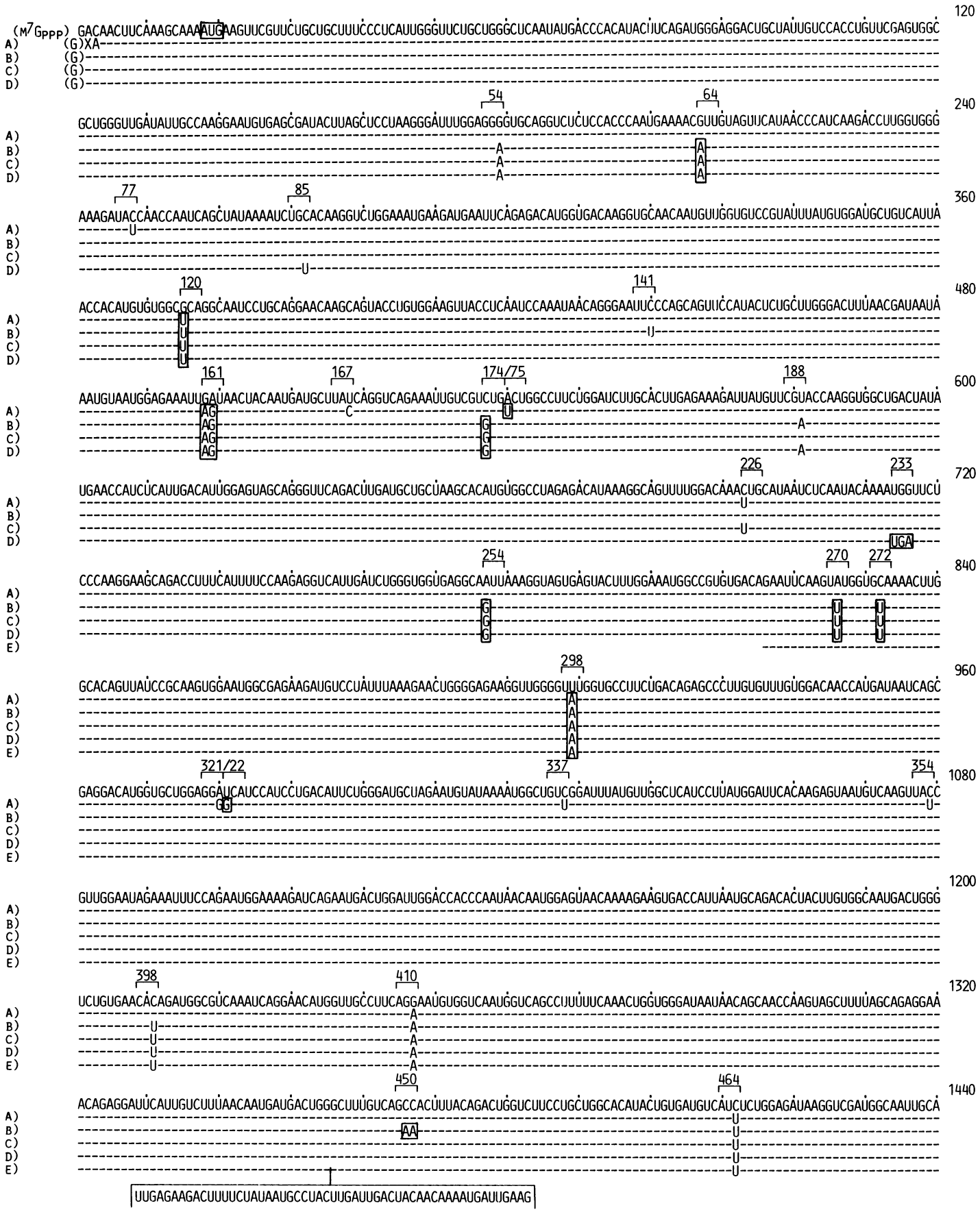


Fig. 1. Comparison of electrophoretic mobilities of alpha-amylase isozymes from CE/J (**a**) and A/J (**b**) pancreas. Amylase was purified by glycogen precipitation; and isozymes were separated by electrophoresis at pH 8.1 on a non-denaturing acrylamide gel and visualized by staining with Coomassie Blue. The various isozymes are designated according to the nomenclature used by Bloor *et al.* (1981). An alternate nomenclature based on electrophoresis at pH 9.0 has been proposed by Hjorth (1982).


(m7GPPP) GACAACUUCAAAGCAAAAUGAAGUUCGUUCUGCUGCUUUCCCUCAUUGGGUUCUGCUGGGCUCAAUAUGACCCACAUACUUCAGAUGGGAGGACUGCUAUUGUCCACCUGUUCGAGUGGC
A) (G)XA
B) (G)
C) (G)
D) (G)
120
54
64
240
GCUGGGUUGAUAUUGCCAAGGAAUGUGAGCGAUACUUAGCUCCUAAGGGAUUUGGAGGGGUGCAGGUCUCUCCACCCAAUGAAAACGUUGUAGUUCAUAACCCAUCAAGACCUUGGUGGG
77
85
360
AAAGAUACCAACCAAUCAGCUAUAAAAUCUGCACAAGGUCUGGAAAUGAAGAUGAAUUCAGAGACAUGGUGACAAGGUGCAACAAUGUUGGUGUCCGUAUUUAUGUGGAUGCUGUCAUUA
120
141
480
ACCACAUGUGUGGCGCAGGCAAUCCUGCAGGAACAAGCAGUACCUGUGGAAGUUACCUCAAUCCAAAUAACAGGGAAUUCCCAGCAGUUCCAUACUCUGCUUGGGACUUUAACGAUAAUA
161
167
174/75
188
600
AAUGUAAUGGAGAAAUUGAUAACUACAAUGAUGCUUAUCAGGUCAGAAAUUGUCGUCUGACUGGCCUUCUGGAUCUUGCACUUGAGAAAGAUUAUGUUCGUACCAAGGUGGCUGACUAUA
226
233
720
UGAACCAUCUCAUUGACAUUGGAGUAGCAGGGUUCAGACUUGAUGCUGCUAAGCACAUGUGGCCUAGAGACAUAAAGGCAGUUUUGGACAAACUGCAUAAUCUCAAUACAAAAUGGUUCU
UGA
254
270
272
840
CCCAAGGAAGCAGACCUUUCAUUUUCCAAGAGGUCAUUGAUCUGGGUGGUGAGGCAAUUAAAGGUAGUGAGUACUUUGGAAAUGGCCGUGUGACAGAAUUCAAGUAUGGUGCAAAACUUG
298
960
GCACAGUUAUCCGCAAGUGGAAUGGCGAGAAGAUGUCCUAUUUAAAGAACUGGGGAGAAGGUUGGGGUUUGGUGCCUUCUGACAGAGCCCUUGUGUUUGUGGACAACCAUGAUAAUCAGC
321/22
337
354
1080
GAGGACAUGGUGCUGGAGGAUCAUCCAUCCUGACAUUCUGGGAUGCUAGAAUGUAUAAAAUGGCUGUCGGAUUUAUGUUGGCUCAUCCUUAUGGAUUCACAAGAGUAAUGUCAAGUUACC
1200
GUUGGAAUAGAAAUUUCCAGAAUGGAAAAGAUCAGAAUGACUGGAUUGGACCACCCAAUAACAAUGGAGUAACAAAAGAAGUGACCAUUAAUGCAGACACUACUUGUGGCAAUGACUGGG
398
410
1320
UCUGUGAACACAGAUGGCGUCAAAUCAGGAACAUGGUUGCCUUCAGGAAUGUGGUCAAUGGUCAGCCUUUUUCAAACUGGUGGGAUAAUAACAGCAACCAAGUAGCUUUUAGCAGAGGAA
450
464
1440
ACAGAGGAUUCAUUGUCUUUAACAAUGAUGACUGGGCUUUGUCAGCCACUUUACAGACUGGUCUUCCUGCUGGCACAUACUGUGAUGUCAUCUCUGGAGAUAAGGUCGAUGGCAAUUGCA
UUGAGAAGACUUUUCUAUAAUGCCUACUUGAUUGACUACAACAAAAUGAUUGAAG

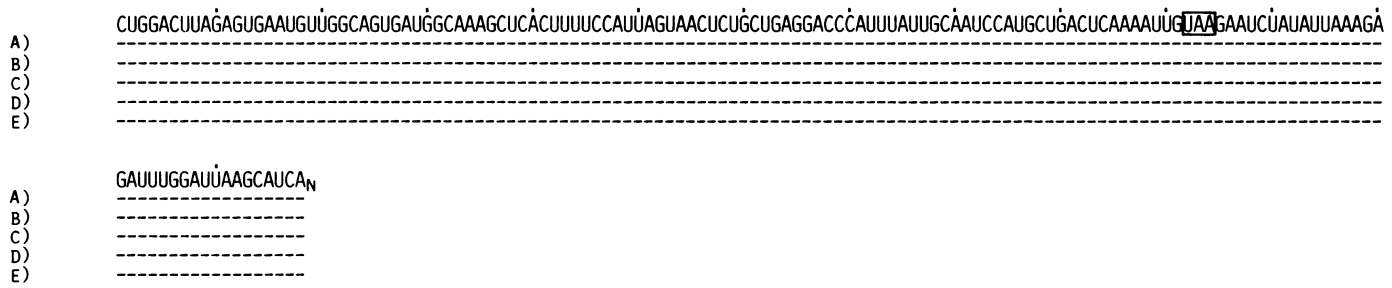


Fig. 2. Nucleotide sequence of various alpha-amylase mRNAs from CE/J pancreas. The CE/J sequences are compared with the mRNA sequence found in the pancreas of A/J mice (Hagenbüchle *et al.*, 1980). The nucleotide substitutions in CE/J mRNAs are shown below the A/J sequence. Dashed lines indicate identical sequences for CE/J and A/J mRNA. Translation initiation and termination signals for alpha-amylase are boxed. Nucleotide changes which confer amino acid changes are also boxed. Altered codons are indicated by a bracket. The number above the bracket designates the position of the amino acid within the protein sequence of Figure 3. X in **(A)** indicates that the 5′-terminal two nucleotides may be missing in this cDNA. The G residue shown in parentheses for cDNAs **B**–**D** is either the cap site or part of the oligo(dG) primer used for second-strand synthesis. Note the presence of a 55 nucleotide insert in **(E)** between residues 1355 and 1356 of the reference sequence. **(A)** pCEPa4; **(B)** pCEPa5; **(C)** pCEPa12; **(D)** pCEPa15; **(E)** pCEPa88.

etic groups (A and B) parallels the behaviour of the same proteins in electrofocusing gels (Wanner *et al.*, 1982) and therefore reflects a difference in net charge between these two groups.

At least five distinct Amy-2 type genes are expressed in CE/J mice

It is not possible to correlate directly the number of isozymes observed and the number of structural genes involved in their synthesis. In order to estimate the number of *Amy*-2 genes that specify the pancreatic isozymes of CE/J mice, we have cloned alpha-amylase mRNAs in the form of their cDNAs. Five distinct CE/J cDNA clones were obtained. Thus, at least five different *Amy*-2 genes must be expressed in CE/J mice since we have no evidence for exon shuffling. Figure 2 compares the mRNA sequences deduced from one partial and four full-length cDNA inserts with the sequence that was determined for the unique pancreatic alpha-amylase mRNA species of A/J mice (Hagenbüchle *et al.*, 1980).

Two cDNAs (D and E) contain unusual sequence elements. A stop codon (residues 714–716) interrupts the coding frame of D; and 55 nucleotides are inserted at residues 1355/56 in the coding region of E. A detailed examination of these unusual features reveals that they are apparently not characteristic for authentic abundant CE/J mRNA. The presence of a translational stop codon in cDNA D raises the possibility that a truncated form of alpha-amylase exists in the pancreas of CE/J mice. However, two lines of evidence argue against the existence of such a protein. First, *in vitro* translation of total pancreatic CE/J alpha-amylase mRNA, which was selected by hybridization to filter-bound cloned cDNA, did not reveal the presence of such a polypeptide under conditions which should have allowed the detection of a product from a moderately abundant mRNA species (data not shown). Second, the nucleotide substitution which generates the stop codon generates an additional *Hinf*I restriction site in this cDNA. However, *Hinf*I digests of double-stranded cDNA synthesized from total pancreatic poly(A)$^+$ RNA, either by oligo(dT) priming or by extension of a radiolabeled primer near the putative site, did not reveal the existence of a D type mRNA species (data not shown). Therefore, this stop codon is present either on a rare RNA species or is an artefact of the cDNA cloning procedure. The latter interpretation is supported by a comparison of the nucleotide sequence of cDNA D with that of a particular genomic alpha-amylase DNA clone isolated from a CE/J DNA-containing cosmid library (Figure 3). The T residue in exon b corresponding to position 272 of the cDNA sequence and the presence of an *Eco*RI site in exon c unambiguously assign this genomic clone to cDNA D. All the remaining nucleotides specified by the exons b, d and e are identical to the corresponding sequence of cDNA D except for the residue that is responsible for generating the stop codon.

The 55 nucleotide insertion at residues 1355/56 of cDNA E was suspected to be an intronic sequence since whole-cell rather than cytoplasmic poly(A)$^+$ RNA was used to construct our pancreatic cDNA libraries. Furthermore, the location of this extra sequence coincides with the location of the last intron in the *Amy*-2^a gene of A/J as determined by electron microscopy (Schibler *et al.*, 1982). To establish the origin of these sequences we determined their cellular location using the RNA fractions and the hybridization strategy described in the legend to Figure 4. The specificity of hybridization was determined in control experiments using M13 cloned single-stranded DNAs carrying or lacking the 55 nucleotide insertion (Figure 4, lanes 6 and 7). Only the nuclear fraction of pancreatic RNA contains sequences homologous to the insertion (Figure 5, lane 5). The sequences at the 3′ end of the insertion suggest the presence of a normal splice acceptor site. In contrast, the sequence elements defining a splice donor site are missing at the other end of the insertion (see Figure 2). Therefore, the 55 nucleotides of cDNA E insertion may be interpreted as the remnant of an aberrant splicing event that has occurred in the nuclear precursor of a *bona fide* mRNA. Alternatively, cDNA E could be the copy of a non-functional RNA product originating from an *Amy*-2-like pseudogene. Although we cannot formally rule out the possibility that mRNAs D and E are products of pseudogenes, we shall refer to them as mRNAs throughout this paper to facilitate further description.

Comparison of the protein coding regions of pancreatic alpha-amylase mRNAs from CE/J mice

The non-coding regions of the various alpha-amylase mRNAs are completely conserved in sequence with the exception of the 5′-terminal one or two nucleotides of the 5′ non-coding portion of mRNA A (see Figure 2).

The coding regions of the various CE/J alpha-amylase mRNAs differ between each other, and from the coding region of A/J mRNA, by ~1% of their nucleotides (see

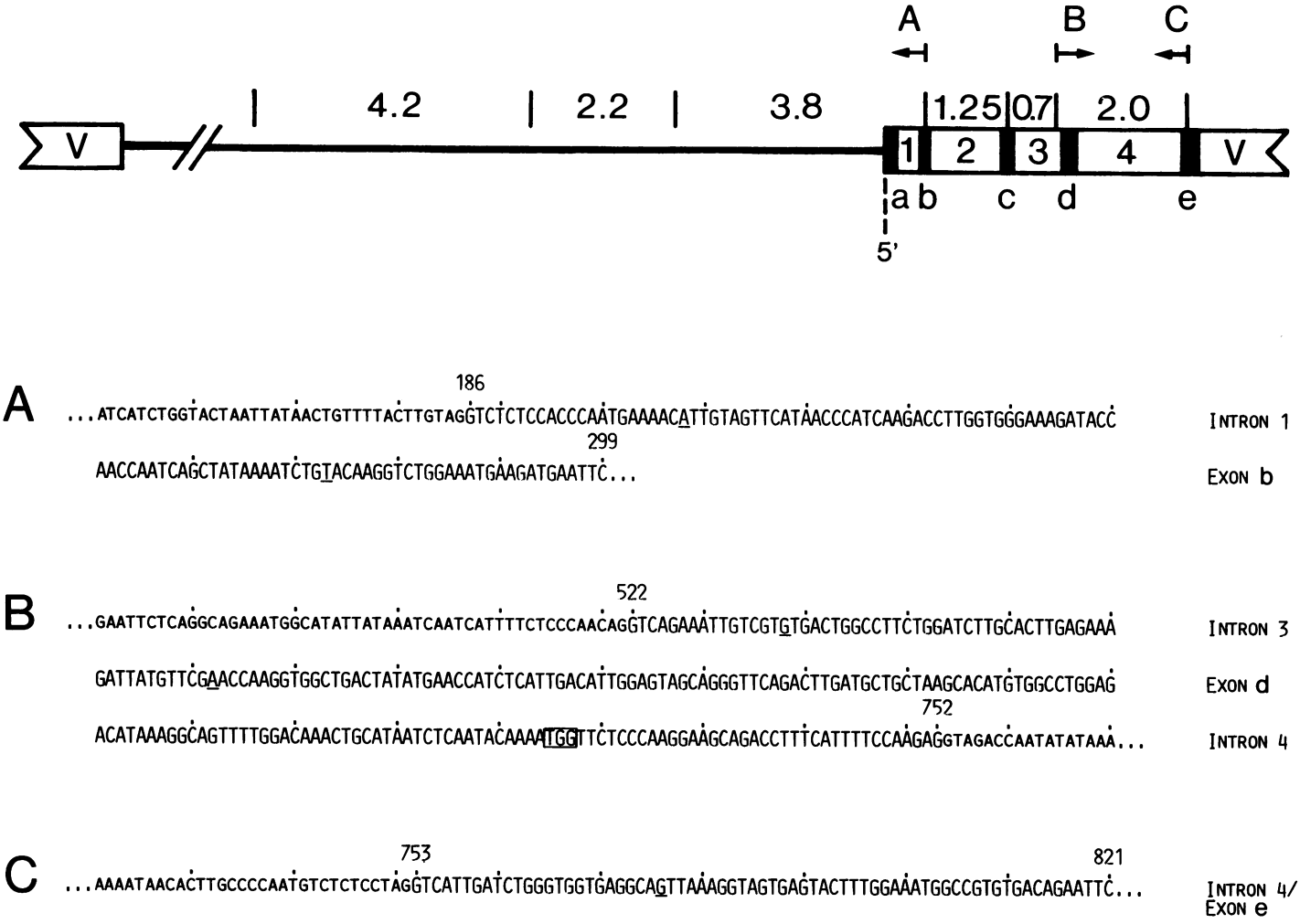


Fig. 3. Exon sequences of the *Amy*-2 gene that specifies the mRNA sequence of pCEPa15. The map on top shows the origin of sequences A, B and C. Introns are represented as open bars and are numbered. Exons are shown as black bars and are designated by lower-case letters. The size of *Eco*RI restriction fragments is given in kb above the map. Cosmid vector sequences are denoted by V. Exonic and intronic sequences in **A, B** and **C** are represented by large and small capital letters, respectively. The nucleotide numbers above exon sequences correspond to those of cDNA sequences of Figure 2. Nucleotides which differ from the cDNA sequence of pMPa21 are underlined. The boxed trinucleotide TGG in exon d is at the position of the UGA chain terminator codon in pCEPa15 (see Figure 2).

Figure 2). Of a total of 25 sites where nucleotide substitutions occur, 14 are silent while 11 determine amino acid replacements. The nucleotide changes which confer amino acid changes in the protein are expected to occur in protein regions with lower functional constraints. They are not distributed randomly but are clustered in two regions of the alpha-amylase protein (Figure 5).

The five distinct CE/J mRNAs encode only four different proteins. The mRNAs C and D specify the same protein. The protein sequences deduced from the nucleotide sequences of cDNAs B and C differ by only one amino acid residue (Glu 450 in B). The region between the amino acid positions 120 and 175, which is also a region of particularly high variability between pancreatic and parotid alpha-amylases (Hagenbüchle *et al.*, 1980), contains three of the 11 amino acid replacement sites. A second highly variable region is located between amino acid residues 254 and 298. Only two amino acid replacements are found in the region extending from residue 298 to the carboxy-terminal end. Most changes are neutral due to the properties of the amino acid side chains substituted in CE/J isoenzymes (Table I). However, two amino acid replacements altering the protein net charge occur at the following positions. At position 161, all CE/J proteins (with the possible exception of E) lack the charged side chain present in the A/J protein. At position 450, the protein encoded by mRNA B contains a glutamic acid residue while the other isozymes of CE/J and A/J contain an alanine residue. It is noteworthy that these two amino acid replacements are the only ones that are specified by a double nucleotide substitution.

A large number of Amy-2 related sequences exist in strain CE/J

Our observation that at least five distinct but closely related alpha-amylase RNAs are expressed in the pancreas of CE/J mice predicts the existence of at least five structural genes at the *Amy*-2 locus of CE/J mice. To compare the distribution of *Amy*-2 related copies in the CE/J and A/J genomes we carried out Southern blot analysis (Figure 6). Hybridization of gel-fractionated, filter-bound *Eco*RI digests of the two genomic DNAs to the radiolabeled alpha-amylase cDNA insert of a recombinant plasmid reveals a strain-specific pat-

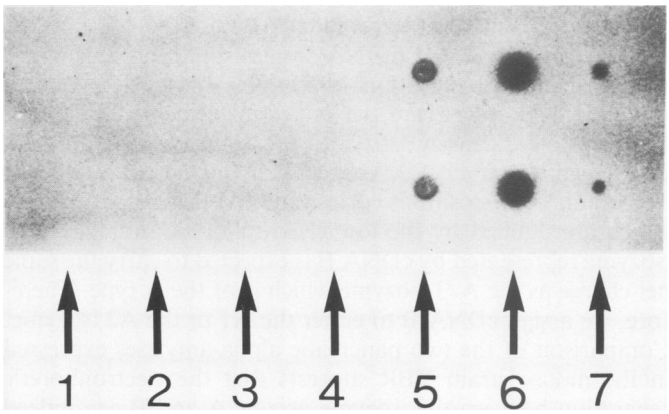


Fig. 4. Dot blot analysis to establish the cellular location of the 55-bp insertion element in pCEPa88 cDNA. RNA and DNA samples were spotted in duplicate on a 'gene screen' NEN membrane. The filter was pre-hybridized for 24 h in 50% formamide solution with 10 μg/ml of cold, gel purified DNA from a M13 recombinant phage containing the coding strand of the 377-bp *Ava*II cDNA fragment of pMPa21 (Schibler *et al.*, 1980; see Figure 7). The filter was then hybridized with the 432-bp *Ava*II DNA fragment of pCEPa88 which was ^{32}P-labeled by nick-translation. The M13 cloned non-coding strands of the 432-bp (6) and 377-bp (7) *Ava*II fragments were used as a control to monitor the specificity of hybridization. **Lane 1,** 5 μg of *E. coli* RNA; **lane 2,** 1 μg of cytoplasmic poly(A)$^+$ RNA of A/J kidney; **lane 3,** 10 μg of total nuclear RNA of A/J liver; **lane 4,** 1 μg of cytoplasmic poly(A)$^+$ RNA of CE/J pancreas; **lane 5,** 10 μg of total nuclear RNA of CE/J pancreas; **lane 6,** 200 ng non-coding strand of pCEPa88 432-bp AvaII fragment; **lane 7,** 200 ng non-coding strand of pMPa21 377-bp *Ava*II fragment.

tern (Figure 6A). Many *Eco*RI fragments differ between the two DNA samples, indicating that considerable polymorphism exists between *Amy*-2 sequences of the two mouse strains. The number of *Amy*-2 related sequences in the CE/J and A/J genomes was estimated by using a hybridization probe derived from the 5′ end of a cloned *Amy*-2^a gene. This genomic probe contains 158 residues of first exon and 130 nucleotides of 5′-flanking sequences. It has little homology to the alpha-amylase gene *Amy*-1 and is therefore diagnostic for *Eco*RI fragments containing the cap site of *Amy*-2. The abundance of *Amy*-2 5′ ends in CE/J and A/J DNA was determined against DNA fragments of cloned alpha-amylase genes of A/J which were electrophoresed on the same gel and in amounts corresponding to one or several gene equivalents (Figure 6B, lanes a–d). The strong hybridization band at 9.6 kb in the genomic A/J DNA (lane e) corresponds to three *Amy*-2 copies. The weak band at 7.5 kb, which is a single copy sequence, contains the cap site of yet another *Amy*-2^a gene (Hagenbüchle *et al.*, 1984). The single copy 4.2-kb *Eco*RI fragment bears the 5′ end of the pseudogene *Amy*-X (Hagenbüchle *et al.*, 1984). Thus, the *Amy*-2 locus of A/J mice harbors a total of four *Amy*-2 copies. In contrast, gene counting reveals >10 *Amy*-2 related sequences per haploid genome of strain CE/J mice. The 5′ end sequences of the majority of CE/J *Amy*-2 copies reside in 4.2-kb and 3.8-kb *Eco*RI fragments (lane f). It should be stressed, however, that this approach, while providing a good estimate for the multiplicity of *Amy*-2 5′ ends, is not necessarily a reliable measure for assessing the number of complete *Amy*-2 genes.

Our data suggest that during evolution numerous and independent gene duplication events have occurred in mice from different inbred strains thus generating variable numbers of *Amy*-2 copies which are expressed in a strain-specific fashion.

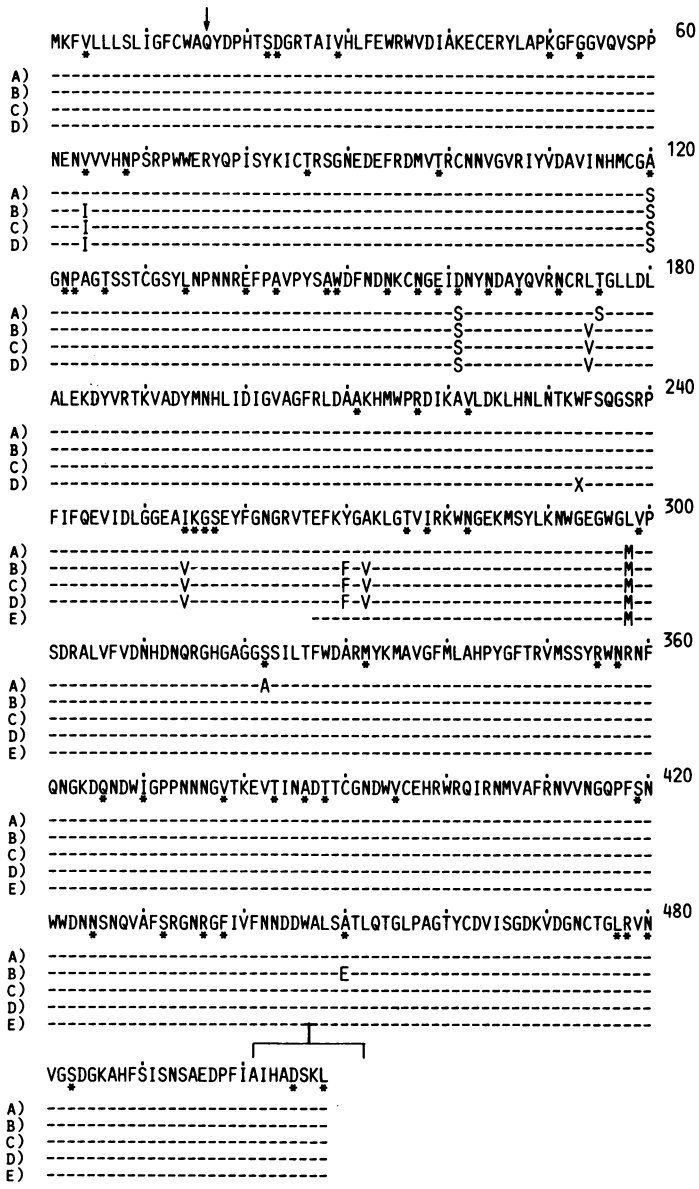


Fig. 5. Comparison of the amino acid sequence of pre-alpha-amylases from CE/J and A/J pancreas as predicted by the cDNA sequences. The A/J protein sequence is shown at the top. Amino acid substitutions in CE/J proteins are indicated. Dashed lines represent identical amino acids. The sites at which A/J pancreatic pre-alpha-amylase differs from its parotid gland and liver counterparts (Hagenbüchle *et al.*, 1981) are designated by an asterisk. The arrow denotes the amino terminus of the mature enzyme (Karn *et al.*, 1981). X indicates the position of the internal UGA chain terminator in **(D)**; and the bracket designates the site of the 55 nucleotide insert in **E**. For identification of **(A)**–**(E)** refer to Figure 2.

Discussion

The data presented here establish the basis for a detailed analysis of the expression of non-allelic *Amy*-2 genes in different inbred strains of mouse. The amino acid sequences of four distinct pancreatic alpha-amylase isozymes of CE/J mice have been deduced from the nucleotide sequences of their mRNAs. We have attempted to correlate these protein sequences with the electrophoretic variants observed on native acrylamide gels. Two amino acid replacements are found to be significant with respect to the electrophoretic separation of CE/J isozymes into A and B forms. At position 161, CE/J mRNAs A–D encode an uncharged residue that replaces the aspartic acid found in the pancreatic alpha-

Table I. Comparison of the amino acid (aa) replacements between pancreatic alpha-amylase isozymes of CE/J and pancreatic and salivary gland alpha-amylase of A/J

	Strain CE/J (locus *Amy*-2)		Strain A/J[a] (locus *Amy*-2)	(locus *Amy*-1)
aa position	aa residue[b]	cDNA[c]	aa residue	aa residue
64	Ile	B,C,D	Val	Ile
120	Ser (P)	A,B,C,D	Ala (NP)	Val (NP)
161	Ser (P)	A,B,C,D	Asp (C)	Glu (C)
174	Val	B,C,D	Leu	Leu
175	Ser	A	Thr	Ser
254	Val	B,C,D	Ile	Val
270	Phe	B,C,D	Tyr	Tyr
272	Val	B,C,D	Ala	Ala
298	Met	A,B,C,D,E	Leu	Leu
322	Ala (NP)	A	Ser (P)	Ala
450	Glu (C)	B	Ala (NP)	Glu (C)

[a]From Hagenbüchle *et al.*, 1980.
[b]P, polar; NP, non-polar; C, charged.
[c]The proteins specified by cDNAs D and E are tentative because of the presence of a stop codon and an insertion, respectively, in these cDNA sequences (see text).

amylase of strain A/J (Hagenbüchle *et al.*, 1980). At position 450, all CE/J and A/J mRNA species specify an uncharged residue with the exception of cDNA B which encodes a glutamic acid. These are the only replacements of charged residues found among isozymes of strains CE/J and A/J. This suggests that all cDNAs specifying uncharged residues at both amino acid positions correspond to the B type isozymes, which are defined by the low electrophoretic mobility. The isozyme represented by cDNA B is expected to carry the same net charge as the A/J isozyme which is of the A type. Therefore, we assign cDNA B to either the A1 or the A2 isozyme. Comparison of the two pancreatic alpha-amylases expressed in the mouse strain YBR suggests that the electrophoretic separation between the isozyme groups A and B can indeed be interpreted on the basis of the replacement of a single charged amino acid. The two YBR isozymes can also be separated into an A and a B form on non-denaturing acrylamide gels (Meisler *et al.*, 1983). The amino acid sequence from residues 407 to 455 has been determined for the rapidly migrating A form of YBR alpha-amylase by direct protein sequencing methods (Strahler and Meisler, 1982) and found to be identical to the sequence specified by CE/J cDNA B within that region. The separation of CE/J isozymes into A1,

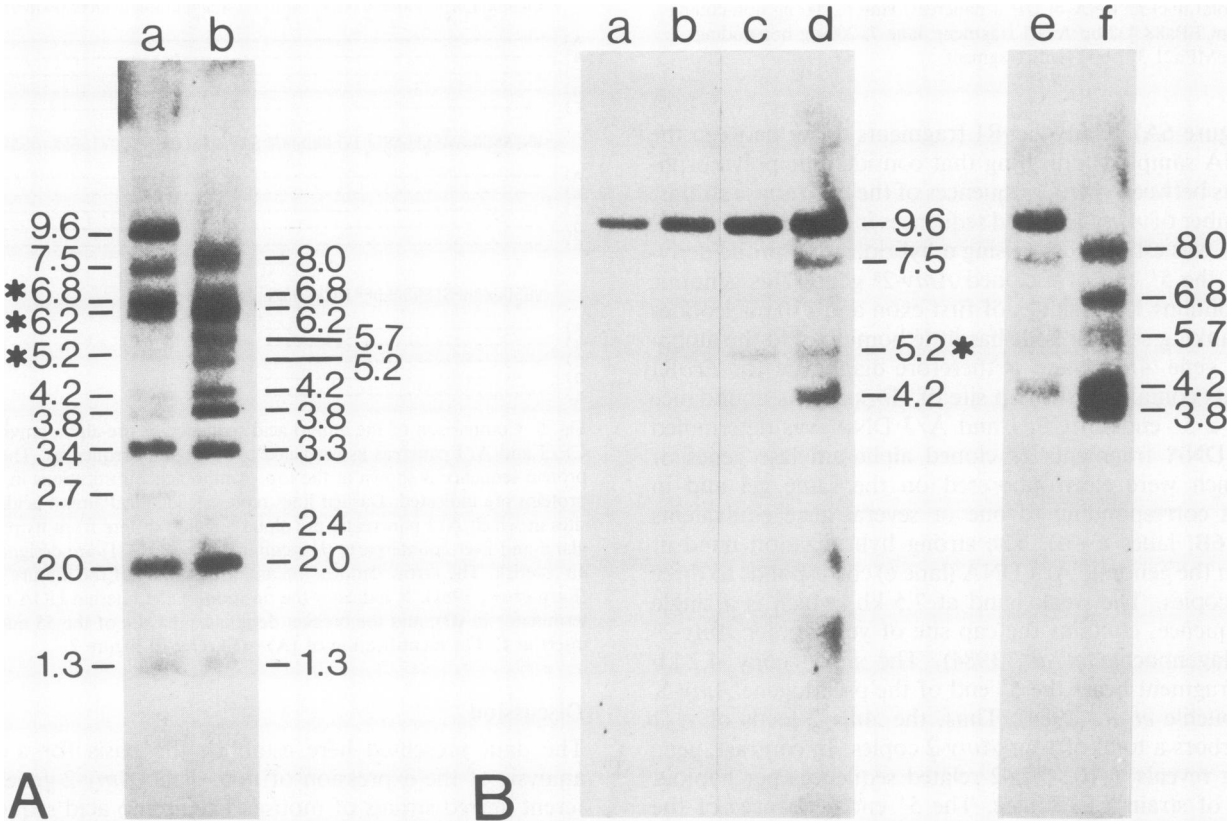


Fig. 6. Southern blot analysis of *Amy*-2 related DNA sequences in CE/J and A/J mice. The DNAs were digested by *Eco*RI restriction endonuclease, fragments were separated by electrophoresis on a 0.8% agarose gel, transferred to nitrocellulose and hybridized to ^{32}P-labeled DNA probes. In **A**, 7.5 μg each of A/J **(lane a)** and CE/J **(lane b)** genomic DNA were hybridized to the cDNA insert of pMPa21 (Schibler *et al.*, 1980). In **B**, 7.5 μg each of A/J **(lane e)** and CE/J **(lane f)** DNA were hybridized to the 287-bp *Alu*I fragment of the genomic clone λMchrαa2^a/1 which contains 5'-terminal sequences of the *Amy*-2^a gene of A/J (Schibler *et al.*, 1982). **Lanes a–d** contain various amounts of an equimolar mixture of cloned *Amy*-2^a and *Amy*-1^a gene sequence of A/J used to titrate amylase DNA sequences in mouse genomic DNA. **Lanes a, b** and **c** contain 1, 2.5 and 4 gene equivalents, respectively. **Lane d** contains 7.5 μg of A/J DNA to which 1.5 gene equivalents of cloned sequence have been added. The size of *Eco*RI restriction fragments is given in kb. Their copy number was determined by densitometric scanning of the autoradiograph. DNA fragments labeled by an asterisk originate from the *Amy*-1^a gene of A/J (Schibler *et al.*, 1982).

A2 and B1, B2 subspecies is probably due to more subtle differences in protein conformation and/or to a different environment of the charged residues. Our cDNA analysis identifies only one of the A type isozymes of CE/J mice and predicts an additional mRNA encoding a charged residue at either position 161 or 450. This mRNA could correspond to the partial cDNA clone E or may be one that has been missed in our search.

The high degree of sequence conservation among exon sequences suggests that the *Amy*-2 genes have originated by multiple duplications from a common ancestor. The fact that 13 out of the 25 bp substitutions are shared among CE/J mRNAs suggests that the genes have undergone rather recent expansion and/or homogenization events. The sequence comparison of pancreatic alpha-amylase cDNAs discloses only five active *Amy*-2 genes while gene titration reveals the presence of more than twice that number of *Amy*-2 related sequences in the CE/J genome. Most of these copies are expected to be clustered in the *Amy*-2 locus on chromosome 3 since Southern blots of genomic DNA from CE/J and from the congenic line C3H.Amy^{ce} yield indistinguishable patterns when hybridized with the entire alpha-amylase cDNA insert probe (our unpublished observations).

The genome of CE/J mice contains a several fold higher number of *Amy*-2 sequences than that of A/J which, like BALB/c, C57BL/6 and C3H, produces a single form of pancreatic isozyme (Meisler *et al.*, 1983). Whether such extensive gene duplication correlates with an exceptionally high number of expressed genes or merely reflects the presence of a larger number of pseudogenes in CE/J is not known. It is conceivable that the four very similar *Amy*-2 copies found in A/J mice are all expressed to yield a single electrophoretic form of alpha-amylase.

The relative rates of alpha-amylase biosynthesis, which are similar for most inbred strains tested (Bloor *et al.*, 1981), may result from a combination of gene dosage effects and differential promoter efficiencies associated with individual gene copies. Several independent lines of evidence suggest that both of these mechanisms play a role in pancreatic alpha-amylase production. Among the five distinct alpha-amylase mRNAs of CE/J mice, those represented by cDNA clones C and D encode the same protein. If a *bona fide* mRNA D is indeed expressed (i.e., if the stop codon in the corresponding cDNA D results from an error during cDNA synthesis), this would indicate that gene dosage is involved in determining the relative abundance of some pancreatic alpha-amylase isozymes. However, in addition to gene dosage effects, more refined patterns of regulation may underly the quantitative differences in the expression of non-allelic *Amy*-2 genes. This is suggested by the observation that two congenic lines (C3H.Amy^{ce} and C3H.Amy^{w3}), which give identical patterns of *Amy*-2 sequences in Southern blots, produce different proportions of the same four electrophoretic variants (Bloor *et al.*, 1981; Nielsen, 1982; and our unpublished data). It is unlikely that major variations between alpha-amylase haplotypes are due to differential mRNA stability and/or translation efficiency since the various mRNAs of CE/J and A/J mice have virtually identical non-coding regions. Therefore, different transcriptional activity of individual *Amy*-2 copies appears to be an additional factor in determining the quantitative variation of pancreatic alpha-amylase isozymes. Although we have not directly demonstrated that individual *Amy*-2 genes are independently regulated at the transcriptional level, total alpha-amylase synthesis in the salivary gland, the liver and the pancreas of A/J mice is modulated by promoters of different strengths (Schibler *et al.*, 1983; and unpublished data). That *Amy*-2 genes can indeed differ in their regulatory elements is also suggested by studies on diabetic mice. The diabetic condition has strong, differential effects on the relative rates of synthesis of CE/J isozymes (Meisler *et al.*, 1983).

Materials and methods

Animals

Three to four months old mice from the inbred strains A/J and CE/J were used for most experiments. The congenic line C3H.Amy CE (Hjorth, 1979) was used to prepare pancreatic proteins.

Extraction of RNA and DNA

Whole-cell RNA from pancreas was prepared according to Harding *et al.* (1978) with the modification described by Schibler *et al.* (1980). Nuclear RNA was prepared as described by Perry *et al.* (1979). Isolation of kidney cytoplasmic RNA using citric acid and purification of polyadenylated RNA on oligo(dT) columns were done as previously reported (Schibler *et al.*, 1980). Mouse DNA was prepared from nuclei according to Wellauer and Dawid (1977).

Gel electrophoresis of DNA

DNA restriction fragments were separated on neutral agarose gels according to McDonnel *et al.* (1977).

Nucleic acid blotting and hybridization

DNA fragments were transferred from agarose gels to nitrocellulose filters (Schleicher and Schuell) or to 'gene screen' membranes (NEN) by the technique of Southern (1975). cDNA and genomic DNA fragments used as hybridization probes were radiolabeled by nick-translation (Maniatis *et al.*, 1975). Hybridization of DNA immobilized on filters was done according to Groner and Hynes (1980).

Molecular cloning

First strand cDNA synthesis was performed on 50 μg of whole-cell polyadenlyated RNA according to Schibler *et al.* (1980). Following filtration on a Biogel A 1.5 m column in 50 mM Tris-HCl (pH 7.4), 1 mM EDTA, the excluded fractions were ethanol-precipitated. 4 μg of single-stranded cDNA were tailed with dC residues; and second-strand synthesis was performed with oligo(dG)$_{12-18}$ as a primer according to the protocol of Land *et al.* (1981). Double-stranded cDNA (2.5 μg) was fractionated on a 1.2% low melting agarose gel and the band containing alpha-amylase cDNA was extracted. 100 ng of this cDNA were tailed with dC residues using 13 units of terminal transferase (TdT) in a final volume of 50 μl as described by Land *et al.* (1981). The *Pst*I-cleaved vector pBR322 was gel purified and tailed with dG residues in the $CoCl_2$-containing buffer according to Deng and Wu (1981) with the following modifications: 4 μg of DNA were tailed for 30 min at 30°C in a final volume of 50 μl using 30 units of TdT and a 35-fold excess of dG over the concentration of DNA ends. Annealing of cDNA to pBR322 and transformation of *Escherichia coli* 5 K (Hubacek and Glover, 1970) were carried out as described by Land *et al.* (1981). The efficiency of transformation was ~3 x 10^5 tetracycline resistant colonies/μg of annealed vector DNA. Approximately 10^3 colonies were replicated onto Millipore filters using sterile velvets and hybridized to the nick-translated 1.5-kb *Pst*I/*Eco*RI fragment of the genomic alpha-amylase clone λMchrαa2^a/1. This fragment contains 295 bp of 5′ exonic sequences from the *Amy*-2^a gene of A/J (Schibler *et al.*, 1982). Among the 120 positive clones obtained in this initial screening a total of 26 were selected by separate hybridizations with an extreme 5′ probe (the 287-bp *Alu*I fragment derived from the genomic fragment described above) and with a 3′ probe (the 2.7-kb *Eco*RI fragment of the genomic clone λMchrαa2^a/2; Schibler *et al.*, 1982). The recombinant plasmids containing apparently full-length cDNA inserts were then transferred to the *Eco*RI strain HB101.

Restriction enzyme digestion and DNA sequence analysis

The recombinant plasmids from 26 clones selected as described above with probes specific for the 5′ end and for the 3′ end of the alpha-amylase mRNA were digested separately with the restriction endonucleases *Pst*I and *Eco*RI. The *Pst*I digestions allowed a more rigorous identification of full length clones while the *Eco*RI digestions revealed two classes of cDNA inserts which differ at one *Eco*RI site. Seven full-length clones were compared by subcloning either the 381-bp *Eco*RI fragment or the partially homologous 522-bp *Eco*RI fragment into the M13mp7 vector (see Figure 7). Sequence analysis of these

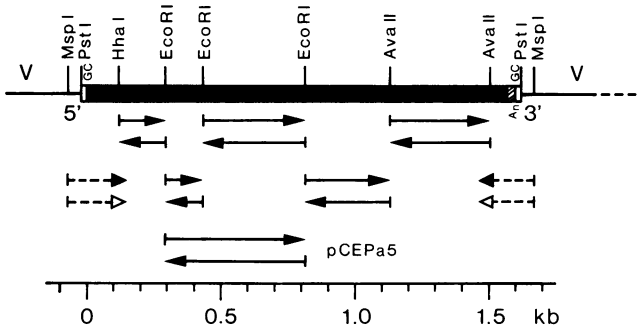


Fig. 7. The strategy used to determine the nucleotide sequences of various alpha-amylase cDNA clones of CE/J. The cDNA insert is shown as a thick bar and pBR322 vector (V) as a thin line. The GC-tails and the poly(A) residues are represented by open and dashed boxes, respectively. The restriction sites used for sequencing are indicated. The *Pst*I sites are those generated by cloning. The direction and extent of DNA sequences obtained by the enzymatic technique of Sanger *et al.* (1977) are designated by solid arrows. Sequences that were obtained by the chemical method of Maxam and Gilbert (1977) are represented by broken arrows. Solid and open heads of broken arrows specify restriction fragments labeled at their 5′ and 3′ ends, respectively. Note that clone pCEPa5 lacks an *Eco*RI site because of a point mutation at residue 440 (see Figure 2).

*Eco*RI fragments using the 'dideoxy method' of Sanger *et al.* (1977) revealed four groups. Representative clones of each group were then fully sequenced using the 'dideoxy method' or the method of Maxam and Gilbert (1977) as outlined in Figure 7. Their sequences were compared with the established sequence of the pancreatic alpha-amylase cDNA clone pMPa21 of A/J (Schibler *et al.*, 1980; Hagenbüchle *et al.*, 1980) which was always included as an internal standard in the sequencing gels.

Acknowledgements

We wish to thank Drs D.Cribbs and O.Hagenbüchle for critical reading of the manuscript, and S.Cherpillod for typing the manuscript. We are grateful to Dr U.Schibler for help and suggestions in the initial phase of this work and to Dr B.Hirt for his continuing support. S.Astolfi was supported by a short term fellowship from C.M.Pq (Brazil). This work was supported by a grant from the Swiss National Science Foundation.

References

Bloor,J.H., Meisler,M.H. and Nielsen,J.T. (1981) *J. Biol. Chem.*, **256**, 373-377.
Deng,G. and Wu,R. (1981) *Nucleic Acids Res.*, **9**, 4173-4188.
Eicher,E.M. and Lane,P.W. (1980) *J. Hered.*, **71**, 315-318.
Groner,B. and Hynes,N.E. (1980) *J. Virol.*, **33**, 1013-1025.
Hagenbüchle,O., Bovey,R. and Young,R.A. (1980) *Cell*, **21**, 179-187.
Hagenbüchle,O., Tosi,M., Schibler,U., Bovey,R., Wellauer,P.K. and Young,R.A. (1981) *Nature*, **289**, 643-646.
Hagenbüchle,O., Wellauer,P.K., Cribbs,D. and Schibler,U. (1984) *Cell*, in press.
Harding,J.D., Prizybyla,A.E., MacDonald,R.J., Pictet,R.L. and Rutter, W.J. (1978) *J. Biol. Chem.*, **253**, 7531-7537.
Hjorth,J.P. (1979) *Biochem. Genet.*, **17**, 665-682.
Hjorth,J.P. (1982) *Heredity*, **48**, 127-135.
Hjorth,J.P., Lusis,A.J. and Nielsen,J.T. (1980) *Biochem. Genet.*, **18**, 281-302.
Hubacek,J. and Glover,S.W. (1970) *J. Mol. Biol.*, **50**, 111-127.
Karn,R.C., Petersen,T.E., Hjorth,J.P., Nielsen,J.T. and Roepstorff,P. (1981) *FEBS Lett.*, **126**, 292-296.
Land,H., Grez,M., Hauser,H., Lindenmaier,W. and Schütz,G. (1981) *Nucleic Acids Res.*, **9**, 2251-2266.
Maniatis,T., Jeffrey,A. and Kleid,D.K. (1975) *Proc. Natl. Acad. Sci. USA*, **72**, 1184-1188.
Maxam,A.M. and Gilbert,W. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 560-564.
McDonnell,W.W., Simin,N. and Studier,F.W. (1977) *J. Mol. Biol.*, **110**, 119-146.
Meisler,M., Strahler,J., Wiebauer,K. and Thomsen,K.K. (1983) in Scandalios,J and Siliciano,M. (eds.), *Isozymes: Current Topics in Biological and Medical Research, Volume 7: Molecular Structure and Regulation*, Alan R. Liss, NY, pp 39-57.
Nielsen,J.T. (1982) *Genetics*, **102**, 571-582.
Nielsen,J.T. and Sick,K. (1975) *Hereditas*, **79**, 279-286.
Owerbach,D., Nielsen,J.T. and Rutter,W.J. (1981) *J. Biol. Chem.*, **256**, 6502-6506.
Perry,R.P., Kelley,D.E. and Schibler,U. (1979) *Proc. Natl. Acad. Sci. USA*, **76**, 3678-3682.
Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*. **74**, 5463-5467.
Schibler,U., Tosi,M., Pittet,A.C., Fabiani,L. and Wellauer,P.K. (1980) *J. Mol. Biol.*, **142**, 93-116.
Schibler,U., Pittet,A.C., Young,R.A., Hagenbüchle,O., Tosi,M., Gellman, S. and Wellauer,P.K. (1982) *J. Mol. Biol.*, **155**, 247-266.
Schibler,U., Hagenbüchle,O., Wellauer,P.K. and Pittet,A.C. (1983) *Cell*, **33**, 501-508.
Southern,E.M. (1975) *J. Mol. Biol.*, **98**, 503-517.
Strahler,J.R. and Meisler,M. (1982) *Genetics*, **101**, 91-102.
Tosi,M., Young,R.A., Hagenbüchle,O. and Schibler,U. (1981) *Nucleic Acids Res.*, **9**, 2313-2323.
Wanner,L.A., Neel,J.V. and Meisler,M.H. (1982) *Am. J. Hum. Genet.*, **34**, 209-215.
Wellauer,P.K. and Dawid,I.B. (1977) *Cell*, **10**, 193-212.
Young,R.A., Hagenbüchle,O. and Schibler,U. (1981) *Cell*, **23**, 451-458.

Received on 6 August 1984