The Remarkable Evolutionary History of the Human Amylase Genes

Miriam H. Meisler and Chao-Nan Ting
Department of Human Genetics, University of Michigan, Ann Arbor, MI 48109-0618

ABSTRACT: Analysis of the structures of the human amylase genes has demonstrated that this multigene family contains at least five tandem gene copies, closely related in sequence but with distinct tissue specific expression. The structures of the genes demonstrate that the human salivary amylase gene was derived from a preexisting pancreatic amylase gene. Insertion of a retrovirus upstream of the amylase gene is responsible for the alteration in tissue specificity. A parotid specific enhancer has been identified within the retrovirus by expression studies in transgenic mice. The independent origin of salivary amylase in rodents and primates suggests that there has been strong evolutionary selection for amylase in saliva. The amylase genes demonstrate a novel mechanism for evolution of new patterns of tissue specific gene expression.

KEY WORDS: salivary amylase, evolution, retrovirus, amylase genes, enhancer.

INTRODUCTION

Alpha-amylase (α-1,4-glucan-4-glucanohydrolase, EC 3.2.1.1) is an enzyme that plays a major role in the digestion of dietary starch and glycogen. Amylase hydrolyzes α-1,4-glycosidic bonds in large polymers to produce the disaccharide maltose, which is cleaved to two molecules of glucose by the enzyme maltase. Amylase thus performs the first step in a crucial process by which cells are provided with glucose, their principal energy and carbon source.

In humans, the major tissues that produce amylase are the pancreas and salivary glands (Merritt et al., 1973). The ease of detection of salivary and pancreatic amylases in serum contributed to an extensive clinical literature, relating the levels of amylase isozymes to various disease states. Based on differences in electrophoretic mobility and antigenicity, it was postulated that pancreatic and salivary amylases were the product of different genes (Merritt and Karn, 1977). During the 1980s, amylase cDNAs were cloned from pancreas and parotid gland and the differences between their sequences confirmed the earlier proposal. Isolation of the amylase genes and comparison of their structures revealed an unusual series of events by which recent duplication of an ancestral pancreatic amylase gene was followed by acquisition of salivary specificity. Human salivary amylase thus represents a recent product of evolution. In this review, we describe the landmarks in the molecular analysis of the amylase gene family, our functional analysis of the salivary amylase regulatory elements in transgenic mice, and the implications for the evolution of other salivary proteins.

CLONING OF HUMAN AMYLASE cDNAs FROM PANCREAS AND PAROTID GLAND

Amylase mRNA is an abundant transcript in pancreas and in salivary glands. The isolation of the human amylase cDNAs was accomplished by in vitro translation of size-selected poly(A') RNA from these tissues followed by precipitation of the protein product with specific antibodies (Nakamura et al., 1984; Wise et al., 1984; Nishide et al., 1986a). The full-length salivary amylase cDNA contains 215 bp of 5'-untranslated region, 1536 bp of coding sequences, and 33 bp of 3'-untranslated region. The coding region of the salivary amylase cDNA differs by only 2% from the pancreatic amylase cDNA, but the 5'-untranslated region of the pancreatic amylase cDNA is considerably shorter. The predicted protein product of the salivary amylase gene contains 511 amino acids and differs by only 3% from the pancreatic amylase protein (Nishide et al., 1986a).

CLONING OF THE HUMAN AMYLASE GENES

The human amylase genes were cloned independently by three groups of investigators. In our labora-
tory, a mouse amylase cDNA was used to screen a cosmid library of human genomic DNA (Gumucio et al., 1988). Overlapping cosmids spanning 230 kb were ordered on the basis of their restriction maps. The 230-kb region was found to contain five complete amylase genes and one truncated pseudogene. The tissue specificity of each gene was determined by comparison with the cDNA sequences. The tissue expression of each gene was also tested experimentally by ribonuclease protection assay of RNA from pancreas and parotid gland (Samuelson et al., 1988). The ribonuclease protection assays confirmed the expectation of strict tissue specificity. The salivary amylase genes are expressed exclusively in salivary gland, whereas the two pancreatic amylase genes are expressed at nearly equal levels in pancreas but are not expressed in salivary gland.

A similar chromosomal organization was observed by investigators at Vrije University in Amsterdam (Groot et al., 1989, 1991), who generated overlapping cosmid clones by screening a human genomic library with a human pancreatic amylase cDNA. A composite view of the amylase gene cluster, based on results from both groups, is presented in Figure 1. The basic array contains two pancreatic amylase genes (AMY2A and AMY2B) and three salivary amylase genes (AMY1A, AMY1B and AMY1C). The three salivary genes appear to be the product of very recent duplication because their restriction maps are identical over a 27-kb region and no sequence difference was observed within a 1-kb region (Gumucio et al., 1988; Samuelson et al., 1988).

The intron/exon structure of the human amylase genes was analyzed by Matsubara and colleagues at Osaka University. They reported that the human salivary amylase gene is approximately 10 kb in length with 11 exons and 10 introns (Nishide et al., 1986b). The size of the exons varies from 100 to 231 bp. The structures of the salivary and pancreatic amylase genes are nearly identical; the major difference is the presence of an extra nontranslated exon at the 5' end of the salivary amylase gene (Horii et al., 1987).

**GENETIC VARIATION IN THE HUMAN POPULATION**

The amylase genes were assigned to human chromosome band 1p21 by in situ hybridization (Zabel et al., 1983) and somatic cell genetics (Münke et al., 1984; Tricoli and Shows, 1984). Molecular techniques have revealed two types of polymorphism at this locus.

A repeated CA microsatellite is located approximately 1-kb upstream of the amylase gene cluster (Gumucio, 1986). A population survey revealed six common alleles in which the number of dinucleotide repeats varies from 16 to 21 (Dracopoli and Meisler, 1990). Genotyping of the CEPH families for this polymorphic marker enabled us to position the amylase cluster on the genetic map of chromosome 1. The microsatellite alleles are detectable by polymerase chain reaction and provide a convenient marker for segregation of the amylase gene cluster in human families.

Variation in gene copy number was detected by Southern blot analysis of human genomic DNA hybridized with amylase cDNA probes (Groot et al., 1989, 1991). The varying densities of restriction fragments in different individuals were consistent with 11 haplotypes generated by unequal, homologous inter- and intrachromosomal crossovers. This variation in copy number would be expected to result in quantitative variation in amylase production.

**THE PROMOTER REGION OF THE SALIVARY AMYLASE GENE IS DERIVED FROM TWO INSERTED ELEMENTS**

Sequencing the upstream regions of the amylase genes and comparison with the sequences in Genbank...
lead to the recognition of two foreign elements in the proximal promoter regions of these genes. A complete γ-actin processed pseudogene is located upstream of AMY2B (Samuelson et al., 1988). The 2-kb pseudogene shares 89% sequence identity with the human γ-actin mRNA, which is consistent with a divergence time of approximately 40 million years (Samuelson et al., 1990). The same junction with γ-actin sequences is present upstream of all of the amylase genes, indicating that they are derived from the single ancestral gene copy in which the insertion occurred (Emi et al., 1988; Samuelson et al., 1988, 1990) (Figure 2).

In all of the genes except AMY2B, the γ-actin pseudogene is itself interrupted by an endogenous retroviral-like element (Emi et al., 1988; Samuelson et al., 1988). Each salivary gene is associated with a complete copy of the endogenous retroviral-like element (Figure 2). Of the pancreatic genes, in contrast, AMY2A contains only a residual LTR left by excision of the retrovirus and AMY2B contains no retroviral sequence (Figure 2). Thus, the presence of a complete endogenous retroviral-like element is correlated with amylase expression in the salivary glands, suggesting that retroviral sequences may contribute salivary-specific regulatory sequences.

Transcription of the pancreatic amylase genes is initiated at exon a, whereas transcription of the salivary amylase genes are initiated from a nontranslated exon (NTE) within the γ-actin pseudogene (Nishide et al., 1986b; Emi et al., 1988; Samuelson et al., 1988). Insertion of the retroviral-like element thus appeared to activate a cryptic promoter within the γ-actin sequences that serves as the initiation site for

Gene | Expression
--- | ---
Ancestral | Pancreas
AMY2B | Pancreas
AMY1A | Parotid gland
AMY1B | Parotid gland
AMY1C | Parotid gland
AMY2A | Pancreas

FIGURE 2. Evolutionary relationships among the human amylase genes. Insertions of the γ-actin pseudogene (solid bar) and the retrovirus ERVA1 occurred approximately 40 million years ago (Samuelson et al., 1990). Exon a and the nontranslated exon (NTE) are represented by open boxes. The major start site for transcription of each amylase gene is indicated by an arrow.
transcription of the salivary amylase gene. Excision of the retrovirus from *AMY2A* is associated with reversion to pancreas-specific expression (Figure 2).

These structural comparisons suggest that the human amylase gene family was generated by expansion of one ancestral gene copy during primate evolution (Samuelson *et al.*, 1990) and that the inserted retrovirus plays a role in regulation of tissue-specificity.

**FUNCTIONAL ANALYSIS OF THE RETROVIRAL-LIKE ELEMENT IN TRANSGENIC MICE**

To test experimentally the role of the retroviral-like element in salivary amylase expression, we transferred a cosmid containing the *AMY1C* gene into the genome of transgenic mice (Ting *et al.*, 1992). We observed tissue-specific expression of the human gene in parotid glands of the transgenic mice. A series of deletion constructs and fusion genes were then tested in 22 independent transgenic lines (Ting *et al.*, 1992). A 1-kb fragment from *AMY1C* (-1003 to +2) was found to be sufficient for parotid-specific expression of a human growth hormone reporter gene. Deletion of the retroviral-derived sequences from -1003 to -826 resulted in reduced levels of transgene expression and loss of tissue specificity. We were able to transfer parotid-specificity to the thymidine kinase promoter with the fragment -1003 to -327, which is derived entirely from the retroviral-like element. The structure of the minimal parotid-specific promoter is presented in Figure 3. Subsequent work has shown that the LTR sequences are not essential, and that the important sequences are derived from the untranslated region of the retrovirus. This retroviral sequence contains three short elements that are present in the 5'-flanking region of other salivary-specific genes (Ting *et al.*, 1992). The analysis of expression in transgenic mice has provided experimental verification of the hypothesis that was originally based on comparison of the structures of the amylase genes.

**PRACTICAL APPLICATIONS OF TRANSGENE EXPRESSION IN SALIVARY GLANDS**

The human salivary amylase enhancer can be used to direct production of foreign proteins into salivary gland of mice and, possibly, other species. This development has potential applications in basic and applied research. Directed expression of oncogenes in parotid gland can be used to develop cultured cell lines that retain some of the differentiated characteristics of salivary acinar cells. Such lines are not currently available but would be quite useful in salivary research. Proteins considered to be protective against oral disease could be directed toward parotid expression to test their effectiveness. For example, induction of caries could be studied in transgenic rats expressing putative protective proteins of mammalian or bacterial origin. Finally, isolation of valuable proteins from saliva of transgenic animals is a future possibility.

![Diagram of *AMY1C* parotid-specific enhancer/promoter and *AMY1C* parotid enhancer](image-url)

**FIGURE 3.** Retroviral origin of sequences sufficient for parotid-specific expression of the human *AMY1C* salivary amylase gene. The 1-kb fragment (-1003 to +2) directed parotid-specific expression of the human growth hormone reporter gene in transgenic mice. The 677-bp fragment (-1003 to -327) was sufficient to transfer parotid-specificity to the heterologous thymidine kinase promoter. Nucleotides are numbered from +1, the start site for transcription. Striped box: retroviral sequences; solid bar: actin-derived sequences; the arrow indicates the direction of transcription of the *AMY1C* amylase gene. (Adapted from Ting *et al.:* Genes Dev. 6: 1457–1465 (1992). With permission.)
**EXPRESSION OF Amy1 IN VARIOUS SALIVARY GLANDS AND OTHER TISSUES**

Among the human salivary glands, the major gland that produces salivary amylase is the parotid gland, but a high level of expression is also observed in submandibular and sublingual gland (Whitten et al., 1988; Korsrud and Brandzaeg, 1982). Ectopic expression of human salivary amylase is also observed in some bronchiogenic lung tumors and ovarian tumors (Zakowski et al., 1984; Hayashi et al., 1986; Tomita et al., 1988) as well as in normal thyroid gland and thyroid adenomas (Doi et al., 1991). In mice and voles, the parotid gland is the only salivary gland that produces amylase (Hjorth et al., 1979), whereas in rat, both parotid and submandibular gland produce amylase (Shear et al., 1973; Bloom et al., 1975). In transgenic mice carrying the human *AMY1C* gene, expression was observed in parotid gland but not in submandibular gland and a very low level of transcripts was also detected in lung and ovary. The transgenic expression mimics the normal expression of human amylase, except for the lack of expression in submandibular gland. These observations suggest that mouse submandibular gland lacks a regulatory factor that is required for amylase expression.

**INDEPENDENT ORIGIN OF A SALIVARY AMYLASE GENE IN HUMAN AND MOUSE GENOMES**

Molecular analysis of the mouse amylase genes demonstrated the presence of distinct salivary and pancreatic amylase genes (Schibler et al., 1982; Wiebauer et al., 1985). A retroviral insertion into the human amylase cluster diverted one pancreatic gene to become a salivary gene (Ting et al., 1992). Because this retroviral insertion occurred *after* the separation of the primate and rodent orders (Samuelson et al., 1990), some of the elements required for salivary amylase gene regulation must have evolved independently in the mouse and human amylase genes (Figure 3). This model is consistent with the conserved intron/exon structure of the human and mouse genes and by comparisons of cDNA sequences, which demonstrate that the paralogous genes within each species are more similar than the orthologous genes (reviewed in Meisler and Gumucio, 1986). The fact that mouse salivary amylase is transcribed from an upstream NTE (Schibler et al., 1983) is also consistent with its derivation from a preexisting pancreatic amylase gene. Expression studies in transgenic mice suggested that the parotid determinant of mouse *AMY1* may be located more than 5-kb distant from the gene (Jones et al., 1989).
**DISCUSSION**

Independent convergent evolution of salivary amylase in human and rodents indicates that there has been strong positive selection for salivary amylase at some points during mammalian evolution. On the other hand, null alleles that lack amylase activity have been observed in primate and rodent populations unaccompanied by obvious deleterious effects (McGeachin and Akin, 1982; Hjorth et al., 1979). Furthermore, many mammalian species, including some species of rodents and primates, are completely lacking in salivary amylase (reviewed in Meisler and Gumucio, 1986). It would be of interest to determine the frequency of null alleles of human amylase by systematic population screening.

What selective advantage of salivary amylase might we propose to account for multiple independent evolutionary origins? Because the enzymatic activities of pancreatic and salivary amylases are quite similar, and all mammalian species produce pancreatic amylase, there is no obvious advantage to duplication of the digestive activity per se in the two organs. One interesting suggestion is that sweet-tasting sugars, produced in the mouth by the action of amylase on complex carbohydrates, might function as signals for identifying nutritious food sources. Alternatively, there may be nonenzymatic functions of salivary amylase that have not been fully appreciated. For example, a role for salivary amylase in oral microbial colonization, mediated by the active site of the enzyme, has recently been proposed (Scannapieco et al., 1990).

The parotid-specific enhancer of AMY1C contains sequences shared by other members of the HER4-1 retroviral family (Ting et al., 1992), raising the possibility that other human salivary genes may have been recruited by insertion of related retroviruses. This possibility could be tested directly by hybridization of genomic clones containing other salivary genes with retroviral-derived probes.

Finally, the origin of human salivary amylase tissue specificity by retroviral insertion may provide a model of a more general phenomenon. Whereas inactivation or overexpression of adjacent genes due to retroviral insertion is well documented, human salivary amylase provides one of the few examples of a qualitative alteration in gene expression that has been retained in the germ line as a species characteristic. Two similar examples have been described recently: insertion of an LTR-derived hormone-response element upstream of a mouse complement gene (Stavenhagen and Robins, 1988) and insertion of an LTR-derived extragonadal promoter upstream of the chicken aromatase gene (Matsumine et al., 1991). Additional examples may be expected to emerge with the continuing accumulation of information about the regulation of specific genes. It has been argued that changes in gene regulation may be more important to the process of speciation than the gradual accumulation of structural variation. Retroviral-like elements and other DNA inserts could be important vectors of rapid qualitative changes leading to complexity and diversity. The presence of thousands of copies of such elements in the mammalian genome, some of which are functional retroposons (Dombroski et al., 1991; Evans and Palmeter, 1991), is consistent with more widespread effects than are currently appreciated.

**ACKNOWLEDGMENTS**

Our work on the amylase genes has been supported by NIH grants GM24872 and DK36089.

**REFERENCES**


Horii, A., M. Emi, N. Tomita, T. Nishide, M. Ogawa, T. Mori, and M. Matsubara: Primary Structure of Human Pancreatic


