

Retrotransposons and the evolution of mammalian gene expression

D. M. Robins¹ & L. C. Samuelson²

¹ Department of Human Genetics, ² Department of Physiology,
University of Michigan Medical School, Ann Arbor, MI 48109, USA

Received and accepted 28 March 1992

Key words: Retrotransposons, retroviral-like elements, provirus, androgen regulation, salivary specificity

Abstract

Transposable elements, and retroviral-like elements in particular, are a rich potential source of genetic variation within a host's genome. Many mutations of endogenous genes in phylogenetically diverse organisms are due to insertion of elements that affect gene expression by altering the normal pattern of regulation. While few such associations are known to have been maintained over time, two recently elucidated examples suggest transposable elements may have a significant impact in evolution of gene expression. The first example, concerning the mouse sex-limited protein (*Slp*), clearly establishes that ancient retroviral enhancer sequences now confer hormonal dependence on the adjacent gene. The second example shows that within the human amylase gene family, salivary specific expression has arisen due to inserted sequences, deriving perhaps from a conjunction of two retrotransposable elements.

Transposable elements and evolution

The ubiquity of transposable elements has made them objects of interest for their own natural history as well as for effects they may exert on their host genomes. Transposons can act as insertional mutagens and can cause chromosomal rearrangements, but the extent to which this affects populations rather than individuals has been studied in only a few instances. It is not obvious that transposable elements could increase the overall fitness of a population (Charlesworth, 1985). However, it has been shown in bacteria that the increased mutation rate of a strain harboring a *Tn10* element is of selective advantage in chemostat competition experiments (Chao *et al.*, 1983). Further, in *Drosophila*, P elements cause hybrid dysgenesis that may result in speciation due to reproductive isolation (Bingham *et al.*, 1982). In a wild mouse population, a defective endogenous provirus benefits its host by expressing envelope protein, which protects against further MuLV infection (Gardner *et al.*, 1986). Nevertheless, most characterized transposon insertions do not have an obvious evolutionary value,

but rather cause gene inactivation (Rubin *et al.*, 1982; Copeland *et al.*, 1983) or, in vertebrate systems, activation of cellular oncogenes (Hayward *et al.*, 1981).

In some cases, transposable elements have been shown to affect regulation of neighboring cellular genes, as documented in detail by Barbara McClintock (1956). The alteration of gene expression during development by corn controlling elements led to conjecture that transposition events also might change gene regulation during evolution (Fincham & Sastry, 1974). This could be of particular importance because morphological evolution may be influenced more by regulatory changes in gene expression than by mutations in structural genes (King & Wilson, 1975). However, there are surprisingly few examples of genes whose expression has been stably altered in evolution due to transposon insertion.

We will briefly review one type of transposon, retrovirus-like elements, and then discuss two examples of altered host gene regulation resulting from insertion events for mouse sex-limited protein (*Slp*) and human salivary amylase genes. The origin

of specialized regulatory elements of these genes suggests that transposable elements may be significant evolutionary agents.

Retroviral-like transposable elements

One class of transposable elements that has received much attention, stemming largely from their role in cancer and disease, are the retroviruses (Weiss *et al.*, 1985). Study of retroviral replication has led to the realization that several kinds of endogenous genomic elements transpose in a manner analogous to a modified retroviral life cycle using reverse transcription of RNA intermediates (Boeke *et al.*, 1985; Varmus, 1988). These elements are structurally diverse but can be divided into two distinct groups: those with and those without direct long terminal repeats (LTRs), or 'retrotransposons' and 'retroposons', respectively (Finnegan, 1989).

To become part of the host's genome, retroviruses must infect germ cells and integrate into a chromosome. In mice, many of the resulting endogenous proviruses represent evolutionarily recent infections because their number and location vary between strains (Stoye & Coffin, 1988). This is in contrast to the majority of proviral-like elements that have had time to accumulate numerous mutations since integration. While some proviral insertions are deleterious, most are presumably benign or these elements would not be so numerous in the genome.

Retroviral-like elements include not only infectious agents of oncogenesis, such as murine leukemia and mammary tumor viruses (MuLV, MMTV), but genetically well-characterized transposons, such as yeast *Ty* and *Drosophila copia*. The first stringent proof that transposition to new sites in the genome could occur via an RNA intermediate and a reverse transcriptase (RT) reaction was found for the yeast transposable element *Ty* (Boeke *et al.*, 1985). There are several types of mammalian retroviral-like elements, such as IAPs (intracisternal A particles) and VL30s (virus-like 30S RNA). Most retroviral-like elements in the genome are defective as retroviruses in that they only contain a subset of essential functions or sequences (Keshet *et al.*, 1990).

One of the first retroviral-like elements to be extensively characterized was *Drosophila copia*

(Rubin, 1983; Mount & Rubin, 1985), which is a major repeated sequence in the fly genome (Bingham & Zachar, 1989). Structurally, both *copia* and *Ty* differ from retroviruses in lacking *env* genes and containing integrase-like sequences before rather than after the RT domain (Mount & Rubin, 1985; Kingsman & Kingsman, 1988). This illustrates that elements from widely disparate species can be more similar to each other than to other elements within the same species. These classes of elements are either evolutionarily ancient, or horizontal transfer may have occurred between species; there is evidence for both of these alternatives (Mount, 1992).

Distinct families of mammalian retroviral-like elements have been characterized. In mice there are several dozen endogenous proviruses that can be activated to C-type (such as MuLV) or B-type (MMTV) retroviruses (Stoye & Coffin, 1985). The vast majority of provirus-like elements, however, are defective due to mutation and/or recombination (Ellis *et al.*, 1981). There are six characterized murine retroviral-like families, ranging widely in number of members (reviewed in Keshet *et al.*, 1990). The families are grouped largely by LTR similarities and by the tRNA primer used, which is a useful idiosyncrasy for classification of new elements.

Murine intracisternal A particles (IAPs) were the first family of endogenous virus-like particles recognized. They are transcribed actively but are deleted for much of the *env* gene, which is probably why particles accumulate but cannot exit the cell. The VL30 family (virus-like element encoding 30S RNA) was discovered due to its propensity to co-package with MuLV. This may be an efficient mechanism for dispersal since VL30s, similarly to *Ty* and *copia*, lack any *env*-like sequence. The VL30s are highly recombinogenic and show intriguing transcriptional responses to hormones and growth factors (Schiff *et al.*, 1991). MuRRS (murine retroviral-related sequences) and GLNs (for the characteristic tRNA^{Gln} primer site) are also highly recombinogenic and were first noted as insertions within MuLV sequences (Schmidt *et al.*, 1985; Keshet *et al.*, 1990). Both of these families have a few dozen complete members in the genome but over a thousand solo LTRs. MuRRS are deleted in both *pol* and *env* domains. Two additional proviral-like families have interesting specificities: MuRVYs (murine repeated virus on the Y chromosome) exist

in about 500 copies on the male sex chromosome and ETNs (early transposons) express in undifferentiated but not differentiated cells. Only within IAPs have open reading frames long enough to encode functional proteins been detected, indicating that most of these elements have accumulated significant mutations. They are therefore incapable of producing proteins for their own propagation; however, these functions may be provided in trans from other elements in the cell.

Several thousand retroviral-like elements, identified by sequence similarity to known retroviruses, also inhabit the human genome (see Larsson *et al.*, 1989, for review). Several distinct families of endogenous proviruses have been characterized and fall into two broad classes, one related to type C retroviruses and a mosaic class related to types A, B, and D retroviruses (Carlson & Larsson, 1988). In contrast to mice, all of the human endogenous proviruses characterized thus far are defective, although retroviral-like functions, such as RT activity and budding of retroviral-like particles, may exist in normal tissues (Nelson *et al.*, 1978; Dirksen & Levy, 1977). Transcripts have been detected originating from several endogenous elements, but functional proteins have not been identified. Many of the endogenous retroviruses appear to have integrated long ago during primate evolution and are now stable within the genome.

Effects of retroviral-like elements on host gene expression

Since retrotransposons are so ubiquitous and abundant in nature, it is not surprising that their insertion has in several instances led to notable mutation of host genes. This is particularly evident in *Drosophila*, since transposons account for the majority of known spontaneous mutations. Null alleles can be created by simple insertion within or near a gene, as exemplified by several of the *white* alleles of *Drosophila* (Bingham & Judd, 1981) and by the *dilute* and *hairless* mutations of mice (Jenkins *et al.*, 1981; Stoye *et al.*, 1988a).

Subtler effects on host gene expression have been documented most extensively in *Drosophila* due to the genetic accessibility. At least some effects are determined by specific transposon sequences, since different elements inserted in the

same locale can produce strikingly distinct phenotypes (Kidd & Young, 1986). Functions that can be altered by transposon insertion, or donated from inserted sequences, include transcription start sites, RNA processing sites (splicing, polyadenylation), or transcription regulatory signals. An example of gene regulation altered by a retrotransposon is the *yellow* allele y^2 of *Drosophila* (Corces & Geyer, 1991); insertion of a *copla*-like gypsy element upstream of the *yellow* promoter distances cellular enhancer elements so that only a subset of tissues express *yellow* correctly.

In addition to removing promoter and enhancer functions, retrotransposons can provide regulatory signals that alter gene expression. A clear example of a retrotransposon donating a promoter which still responds to normal tissue-specific controls is the solo IAP LTR promoter of the rat oncomodulin gene (Banville & Boie, 1989; see article by D. Banville, M. Rotaru and Y. Boie in this volume). Altered regulation is exemplified by the *c-mos* oncogene in a murine myeloma cell line that is constitutively expressed from the LTR of an inserted IAP (Horowitz *et al.*, 1984). Similarly, an IAP upstream of the interleukin-3 gene in a leukemia cell line has rendered that gene constitutive (Ymer *et al.*, 1985). An intriguing variation in an ape cell line is interleukin-2 production due to the use of a polyadenylation signal from a leukemia virus inserted in the 3' untranslated region of the IL-2 gene (Chen *et al.*, 1985). An IAP insertion in the 3' flank of a duplicated mouse renin gene may be associated with high levels of submaxillary gland expression (Burt *et al.*, 1984). In these examples, it may be difficult to establish causality or to determine without further analysis whether the inserted element has brought in a new function, disrupted a cellular control, or is a secondary mutation.

The best-studied examples of altered cellular gene regulation imposed by transposable element insertion are the yeast ROAM mutations (regulated overproducing alleles responding to mating type) (Roeder & Fink, 1983). These activating mutations have been observed for a multitude of genes when *Ty* elements transpose next to them (usually oppositely oriented in the upstream flank). *Ty* is sensitive to yeast mating type status due to cis-acting regulatory elements that enhance transcription in haploid cells. In ROAM mutations, these elements now place neighboring genes under mating type control.

Transposable elements thus alter expression of genes in a variety of ways, including preventing expression, producing altered products, and providing new regulatory signals. However, few wild-type genes have been found stably associated with elements that influence their regulation. Either deleterious effects (Hayward *et al.*, 1981) or continued transposition (McClintock, 1956) causes loss of the mutation from the population. We have found two examples of transposons that have become heritably associated with mammalian gene regulation. In one case this has led to dependence on steroid hormone for expression and in the other case to altered tissue specificity.

LTR sequences of an ancient provirus impose androgen regulation on the *Slp* gene

The mouse *C4* and *Slp* genes, which are neighbors in the major histocompatibility complex, encode two very similar proteins that are distinct in their function and regulation. *C4*, the fourth complement component, is pivotal in the humoral immune response, whereas the variant *Slp* is not active in the complement pathway and is expressed only in mature male mice (hence sex-limited) (Shreffler, 1982). Molecular analysis confirms that *Slp* arose from a duplicated *C4* gene, with the duplication extending over 55 kb including the adjacent 21-hydroxylase gene (Chaplin *et al.*, 1983) (Fig. 1A). Divergence resulted in *Slp*'s loss of lytic activity and acquisition of hormonal control, but greater than 95% sequence identity is maintained to *C4* in coding and flanking regions (Hemenway *et al.*, 1986; Nonaka *et al.*, 1986). This striking homology, yet distinct regulation, allows correlation of sequence differences with differences in expression (Stavenhagen *et al.*, 1987). The comparison is enhanced by alleles of *Slp* that differ in androgen regulation. In man and rat, the *C4* gene has also been duplicated, but sex-specific expression has not been seen. We have found that the major regulatory difference in mice between *C4* and *Slp* is due to an ancient retroviral-like insertion that has imposed androgen dependence on the downstream gene (Stavenhagen & Robins, 1988)

An androgen-responsive enhancer of *Slp* was first mapped 2 kb upstream of the gene by virtue of its DNaseI hypersensitivity in chromatin of ex-

pressing tissues (Hemenway & Robins, 1987) and its ability to confer androgen response on heterologous promoters in transfection experiments (Lorenz *et al.*, 1988). The enhancer resides just beyond a breakpoint in the *C4-Slp* homology and is repetitive within the mouse genome. A second copy of the enhancer was found 6 kb further upstream of *Slp*, in the same orientation as the first, immediately beyond which homology to the *C4* flank recurred (Stavenhagen & Robins, 1988). The disruption of *Slp*'s 5' flank by a 6 kb insert with direct terminal repeats, which had enhancer activity, immediately suggested that a retroviral-like element was altering expression of the adjacent *Slp* gene via LTR regulatory sequences.

Sequence analysis confirmed the retroviral nature of the insertion, which was nick-named 'imposon' to connote its effect on *Slp* regulation (but not to imply a novel class of element, as it is simply a degenerate C-type retrovirus). The provirus is divergently oriented to *Slp* (Fig. 1B). Comparison to *C4* sequence showed that the insertion is flanked by a 4 bp target site duplication. The binding site for a tRNA^{Pro} shows additional homology to the MuLV primer binding site. A cryptic start site detected in transfection corresponds to the LTR promoter, but endogenous proviral transcripts have not been detected in mice.

Perhaps most intriguing for this element is the 5% sequence divergence of the LTRs. As one LTR is replicated from the other, they should be identical upon integration; thus this sequence difference implies residence in the genome long enough to accumulate numerous mutations. Most characterized proviruses share greater identity of their LTRs (Hodgson *et al.*, 1983; Schmidt *et al.*, 1985). Five million years may approximate the time of insertion based on the rate of synonymous substitutions in mouse coding DNA (0.5%/myr, divided by two for divergence from a common sequence) (Wu & Li, 1985). However, some recent IAP insertions have non-identical LTRs (Keshet *et al.*, 1990).

Further evidence for the age of the element is seen in the lack of any open reading frame longer than 200 codons, despite homology over the entire 6 kb to retroviral sequences (Fig. 1C). Relative to MuLV structure, there is a large deletion of the RT portion of the *pol* gene and a deletion from the end of *pol* through the first two-thirds of the *env* gene. The *env* deletion is similar to MuRRS structure and

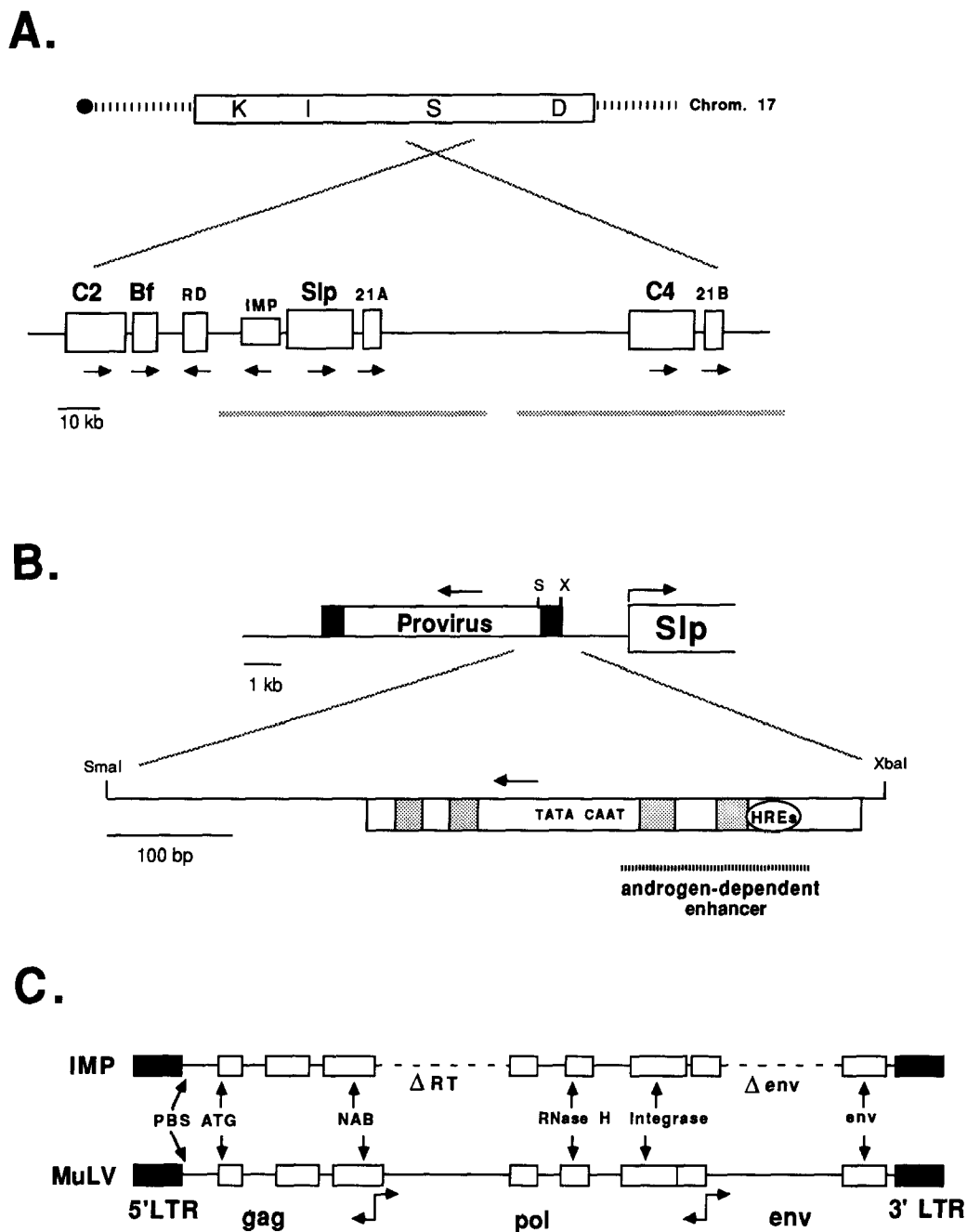


Fig. 1. An ancient provirus imposes androgen regulation on the mouse *Slp* gene. A). The S locus of the major histocompatibility complex is shown, inverted with respect to the centromere, below a diagram of its relative position on chromosome 17. S locus genes include complement C2, C4 and factor B (Bf) (Chaplin *et al.*, 1983). The C4 duplication, indicated by stippled lines, included 21-hydroxylase genes A and B. Arrows indicate 5' to 3' orientation. The provirus is called IMP, short for impon. B). The provirus is oppositely oriented 2 kb upstream of *Slp*. The hormone-responsive SmaI-XbaI fragment is expanded below, with the 5' LTR indicated as the large rectangle. TATA and CAAT promoter elements are marked; shaded boxes represent nuclear protein binding sites in the region of the male-specific DNase I hypersensitive sites. The androgen-dependent enhancer defined in transfection is indicated by a stippled line below the LTR and includes a region of consensus hormone response elements (HREs). C). The provirus is compared schematically to Moloney murine leukemia virus, with functional domains marked (PBS - primer binding site; ATG - translation start; NAB - nucleic acid binding domain). Boxed regions have more than 50 % nucleotide homology between the two sequences; dashed lines indicate two large deletions in IMP. Displaced arrows (at NAB and integrase) indicate smaller insertions or deletions in IMP. Divergent arrows mark the boundaries of the MuLV *gag*, *pol* and *env* genes. Figure adapted from Stavenhagen & Robins, 1988, and Adler *et al.*, 1991.

the LTRs show significant homology to MuRRS LTRs (Schmidt *et al.*, 1985). Further sequence comparison to other retrotransposons may reveal more clearly the phylogenetic origins of this element. Further examination of this locus in feral mouse species may help to pinpoint the time of element insertion.

In vivo and *in vitro* data are all consistent with the notion that this proviral LTR is the hormone responsive enhancer of the *Slp* gene. The responsiveness of these sequences to androgen is not un-

usual as transposons in general, and retroviruses in particular, have adapted ways to take advantage of cell growth conditions, as exemplified by yeast ROAM mutations (Varmus, 1982; Roeder & Fink, 1983). Mouse VL30s are induced by several agents, such as epidermal growth factor (Foster *et al.*, 1982; Schiff *et al.*, 1991). In fact, the classic experimental model of steroid hormone action is mouse mammary tumor virus; LTR sequences of MMTV respond to glucocorticoids, progesterone, and androgens (Yamamoto, 1985). Sequences within the

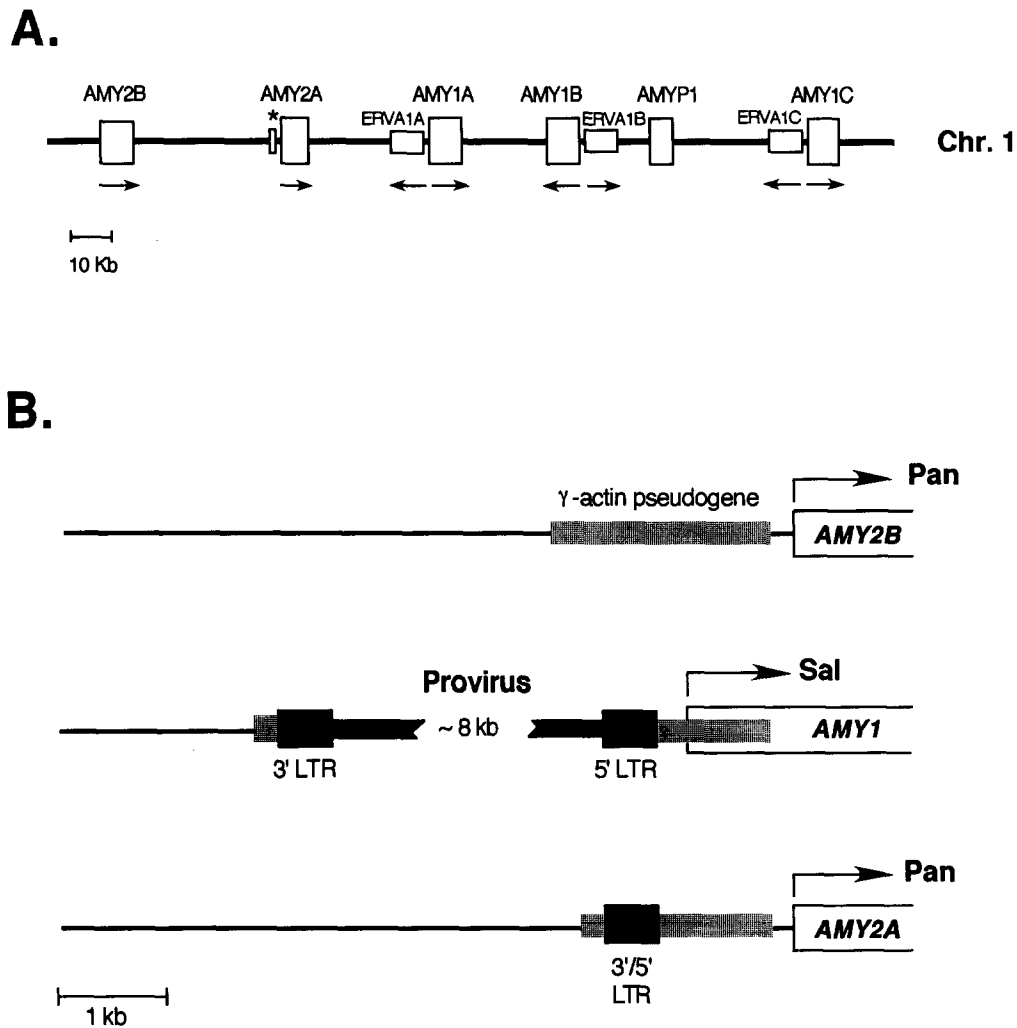


Fig. 2. Retrotransposable elements regulate human salivary amylase expression. A). Diagram of the human amylase gene complex on chromosome 1p21, including the two pancreatic genes, *AMY2A* and *AMY2B*, the three salivary amylase genes, *AMY1A*, *AMY1B* and *AMY1C*, and a truncated pseudogene *AMYP1*. The endogenous retroviral-like elements upstream of the salivary genes are *ERVA1A*, *ERVA1B* and *ERVA1C*. Arrows indicate 5' to 3' orientation. *, solo LTR. B). Two elements inserted into the promoter regions of the human amylase genes. The major start sites for transcription of amylase in pancreas (Pan) and parotid gland (Sal) are indicated. *AMY1* transcripts initiate within the γ -actin pseudogene, 235 bp downstream from the 5' LTR of the provirus. The three *AMY1* genes have identical structures. Relative to amylase, the γ -actin pseudogene is in the same orientation and the provirus in the opposite orientation. Open box, amylase sequences; shaded box, γ -actin sequences; filled box, retroviral-like sequences. Figure adapted from Samuelson *et al.*, 1990.

LTR that regulate *Slp* may prove to be a useful model for how hormone specificity is achieved, as these sequences respond selectively *in vivo* to androgen. This specificity appears to depend on accessory factors that bind within the LTR and cooperate with bound hormone receptors (Adler *et al.*, 1991). Ongoing mechanistic studies on hormone action substantiate this element as primarily responsible for *Slp*'s androgen dependence.

It is difficult to argue that the altered regulation of *Slp* is of selective value to the mouse, even though the association of this element with *Slp* appears to be quite old. In fact, while all *mus* strains examined so far have an *Slp* gene and appear to have the provirus upstream (Stavenhagen & Robins, 1988), most mice do not express *Slp*; *Slp* is apparently continuing its progression to pseudogene status. Whether *Slp* is no longer under selection due to the coding defect that renders it inactive as a complement component (Hemenway *et al.*, 1986) or due to the regulatory change to male-specificity may be determined by dating these two events in different mouse species. Regardless, the maintained association of *Slp* and its retrotransposon is curious and makes it unlikely that this imposition of novel regulation on a host gene is a unique event in the mammalian genome.

Retroposons regulate human salivary amylase expression

Human amylase is encoded by two sets of genes that differ in tissue specificity; the two *AMY2* genes are expressed in pancreas and the three *AMY1* genes in the parotid salivary gland (Gumucio *et al.*, 1988; Samuelson *et al.*, 1988; Groot *et al.*, 1989) (Fig. 2A). Amylase is secreted from the acinar cells of these tissues and functions in carbohydrate digestion. The 98% sequence homology between the pancreatic and salivary amylase cDNAs (Nishide *et al.*, 1986) suggests recent evolutionary derivation. There is also 92% identity over a 750 bp region around the promoters (Horii *et al.*, 1987). However, the promoters of the pancreatic and salivary amylase genes are distinct. The salivary promoter is associated with a nontranslated exon 0.5 kb upstream from the region corresponding to the pancreatic promoter (Horii *et al.*, 1987). Furthermore, the genes are expressed in a strictly

tissue specific manner (Samuelson *et al.*, 1988).

Analysis of the human amylase genes revealed two elements inserted at the promoter regions, a γ -actin pseudogene and an endogenous retroviral-like element (Emi *et al.*, 1988; Samuelson *et al.*, 1988) (Fig. 2B). The actin-like sequences begin 0.2 kb upstream of the first protein coding exon in all five amylase genes. A complete pseudogene, 89% identical to γ -actin mRNA, is associated with *AMY2B*; the other four genes show truncated versions (Samuelson *et al.*, 1988; Samuelson *et al.*, 1990). The insert typifies processed pseudogenes, having an A-rich 3' end and no introns, implying an origin from reverse transcribed mRNA. Of about fifteen γ -actin pseudogenes in the human genome (Ponte *et al.*, 1983), five are in the amylase cluster.

The *AMY1* genes have a retroviral-like insertion in the γ -actin pseudogene (Emi *et al.*, 1988; Samuelson *et al.*, 1988; Samuelson *et al.*, 1990). The element is a member of the 4-1 family of C-type proviruses (Steele *et al.*, 1984; Repaske *et al.*, 1985). This family has 50-100 members, distinguished by a tRNA^{Glu} primer binding site. All three *AMY1* genes contain complete proviral inserts (*ERVA1A*, *ERVA1B* and *ERVA1C*). *ERVA1* is only 235 bp 5' to the *AMY1* transcription start site, in the opposite orientation. The pancreatic *AMY2A* gene has only a solo LTR due to retroviral excision.

Since the promoter and first exon of *AMY1* are derived from γ -actin pseudogene sequences, it was tempting to speculate that integration of the retrovirus activated a cryptic promoter in the pseudogene. Further, *ERVA1* appears to play a role in tissue specificity, since only the amylase genes with a complete provirus express in salivary glands. The retrotransposon's role was tested in transgenic mice, where a cosmid clone with an *AMY1C* gene containing the complete provirus expressed in parotid (personal communication of M. H. Meisler, C. N. Ting, and L.C. Samuelson). Even a 1 kb fragment derived entirely from inserted elements was sufficient to direct parotid expression of a heterologous gene. This demonstrates that one or a combination of both inserts contain DNA sequences sufficient for parotid specificity.

Amylase is synthesized in the pancreas of all vertebrates, but not all produce salivary amylase. Within orders expressing salivary amylase, i.e. primates, rodents and lagomorphs, expressing species include human, mouse, rat and rabbit. Species that

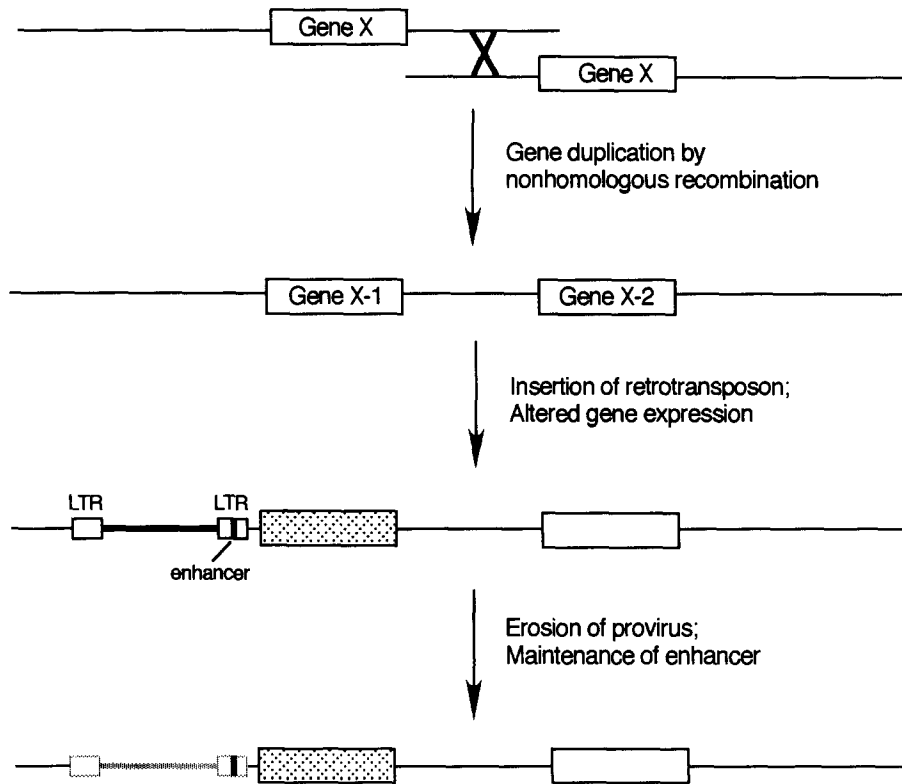


Fig. 3. Gene expression may be altered in evolution by transposable element insertion. Prior gene duplication, which occurs commonly by nonhomologous recombination, may be a necessary first event so that changes in gene structure or regulation do not cause loss of an essential preexisting function; gene duplications occurred first for both Slp and amylase. Divergence may occur slowly over time for the duplicated genes (X-1, X-2) by gradual base change, or there may be a sudden gain of function by insertion of sequences resident in a transposable element. These sequences may be retroviral-like LTR enhancers or may involve more complex juxtapositions of proviral and endogenous gene sequences that generate novel expression patterns. While the association between the inserted regulatory elements and the host gene may be maintained, due to active or neutral selection, other sequences of the transposable element may mutate so that eventually only the regulatory sequences remain.

do not express salivary amylase within these orders include Chinese hamster and squirrel monkey (Dawson & Huang, 1981; McGeachin & Akin, 1982). The complex pattern of salivary amylase expression suggests that tissue specificity has been acquired independently in different lineages (Meisler & Gumucio, 1986). The human amylase genes are thought to have derived from a single precursor as all have the γ -actin pseudogene. This insertion occurred approximately 40 million years ago, well after the separation of mammalian lineages (Samuelson *et al.*, 1990). The proposed derivation of human salivary and pancreatic amylase genes from a single precursor implies a change in site of expression from pancreas to salivary gland during

primate evolution. There are no γ -actin or retroviral inserts in the mouse amylase cluster, which also has salivary specific genes. Although the mouse and human pancreatic promoters are markedly conserved (Gumucio *et al.*, 1988), the salivary promoters are widely divergent in position and sequence. Sequences necessary for expression of amylase in mouse parotid have not been mapped but are far from the promoter (Jones *et al.*, 1989). However, since the human salivary amylase gene is expressed in mouse parotid, the mechanism of salivary specific expression must be similar. Further analysis of primate amylase gene structure and expression may provide insight into the switch in tissue specificity that occurred upon (or after) the γ -actin and retro-

viral insertions. Squirrel monkeys should be particularly informative as they do not produce salivary amylase and are thought to have diverged from the primate tree near the time of the pseudogene insertion.

Conclusions

These two cases of specific gene regulation altered by transposable elements have some intriguing similarities as well as differences that may be informative for unveiling additional examples of this phenomenon. The hormonal dependence of *Slp* is relatively simply imposed by (former) proviral regulatory sequences, similar to the manner in which *Ty* confers mating type control on yeast genes. Origin of the switch in amylase tissue specificity may be more complex and may involve interaction of multiple elements.

For both *amylase* and *Slp*, the retroviral-like insertions have occurred subsequent to a gene duplication event (Fig. 3). (Interestingly, both insertions have also occurred within other transposed elements - a B1 *Alu*-like repeat for the *Slp* provirus and an actin pseudogene for amylase.) Because of gene duplication, the alterations in regulation have not caused loss of pre-existing specificities or functions. At least within multigene families, therefore, altered gene regulation by transposon-induced mutation may be tolerated. In instances where retroviral insertions donate a new transcription start site but do not alter specific regulation, as for rat oncomodulin, prior gene duplication may not be essential as function is not lost.

Whether the association of these elements with endogenous genes is actively maintained, or whether mammalian chromosome evolution simply is conservative, is another question. An evolutionary advantage of salivary amylase is somewhat apparent, as it may allow expanded breadth of dietary substrates. As *Slp* is being progressively silenced in mouse evolution, it is difficult to argue that its male-specificity is of great selective value. Perhaps other gene targets of similar proviruses, however, affect mouse behavior, which is influenced by the diversity of secreted proteins in urine. From another viewpoint, the hormonal and tissue specificities may relate more to selfish goals of the proviral elements for their own dispersal. That is, for viruses

to spread vertically they must integrate in the germline; androgen responsive LTRs may allow for high expression in testis leading to germ cell infection. For horizontal transmission, expression in parotid would allow passage between individuals by secretion into saliva.

Effects of neighboring genes would come under increasing evolutionary scrutiny once retroviral elements are inactivated by mutation. Certainly in the *Drosophila* genome, elements are generally selected against, perhaps due to pressure to keep the genome small; this is supported by the finding that few if any solo LTRs remain to mark former integration sites. In contrast, mammalian genomes are littered with remnants of ancient (and not so ancient) proviruses. Maybe these elements have had some role in speciation, by changing foraging or reproductive behavior of individuals due to altered gene regulation. It seems that in mammals, the integrated elements are easier to ignore than to remove, so long as they are not in themselves deleterious.

Given the large number of proviruses in the mammalian genome that have accumulated over millions of years, it is perhaps surprising that so few examples of their functional and stable liaison with endogenous genes have been found. A great source of regulatory variation resides in these sequences. Likely candidates are the VL30s, which have recently been shown to express at high levels in steroidogenic tissues and to have LTRs that respond to trophic hormones (Schiff *et al.*, 1991), which may account for their efficiency in germline integration and reveal a capacity for conferring novel patterns of regulation. The ability of transposable elements to diversify host gene expression, while unlikely to be commonly used, provides a plausible mode of rapid and dramatic molecular evolution.

Acknowledgements

D.M.R. would like to thank Steve Mount for numerous enjoyable and informative discussions. L.C.S. would like to thank Miriam Meisler for her support and encouragement throughout the course of the amylase work. The research of the authors was supported by grants from the N.I.H.

References

- Adler, A.J., A. Scheller, Y. Hoffman & D. M. Robins, 1991. Multiple components of a complex androgen-dependent enhancer. *Mol. Endocrinol.* 5: 1587-1596.
- Banville, D. & Y. Boie, 1989. Retroviral long terminal repeat is the promoter of the gene encoding the tumor-associated calcium-binding protein oncomodulin in the rat. *J. Mol. Biol.* 207: 481-490.
- Bingham, P. M. & B. H. Judd, 1981. A copy of the copia transposable element is very tightly linked to the *w^a* allele at the white locus of *D. melanogaster*. *Cell* 25: 705-711.
- Bingham, P. M., M. G. Kidwell & G. M. Rubin, 1982. The molecular basis of P-M hybrid dysgenesis: the role of the P element, a P-strain-specific transposon family. *Cell* 29: 995-1004.
- Bingham, P. M. & Z. Zachar, 1989. Retrotransposons and the FB transposon from *Drosophila melanogaster*, pp. 485-502 in *Mobile DNA*, edited by D. E. Berg & M. M. Howe. American Society for Microbiology, Washington, D.C.
- Boeke, J. D., D. J. Garfinkel, C. A. Styles & G. R. Fink, 1985. Ty elements transpose through an RNA intermediate. *Cell* 40: 491-500.
- Burt, D. W., A. D. Reith & W. J. Brammar, 1984. A retroviral provirus closely associated with the Ren-2 gene of DBA/2 mice. *Nucl. Acids Res.* 12: 8579-8593.
- Carlson, M. & E. Larsson, 1988. Human endogenous retroviruses. *Bioessays* 9: 191-196.
- Chao, L., C. Vargas, B. B. Spear & E. C. Cox, 1983. Transposable elements as mutator genes in evolution. *Nature* 303: 633-635.
- Chaplin, D. D., D. E. Woods, A. S. Whitehead, G. Goldberger, H. R. Colten & J. G. Seidman, 1983. Molecular map of the murine S region. *Proc. Natl. Acad. Sci. USA* 80: 6947-6951.
- Charlesworth, B., 1985. The population genetics of transposable elements, pp. 213-232 in *Population Genetics and Molecular Evolution*, edited by T. Ohta & K. Aoki. Springer Verlag, Berlin, Ger.
- Chen, S. J., N. J. Holbrook, K. F. Mitchell, C. A. Vallone, J. S. Greengard, G. R. Crabtree & Y. Lin, 1985. A viral long terminal repeat in the interleukin 2 gene of a cell line that constitutively produces interleukin 2. *Proc. Natl. Acad. Sci. USA* 82: 7284-7288.
- Copeland, N. G., K. W. Hutchison & N. A. Jenkins, 1983. Excision of the DBA ecotropic provirus in dilute coat-color revertants of mice occurs by homologous recombination involving the viral LTRs. *Cell* 33: 379-387.
- Corces, V. G. & P. K. Geyer, 1991. Interactions of retrotransposons with the host genome: the case of the gypsy element of *Drosophila*. *Trends in Genet.* 7: 86-90.
- Dawson, W. D. & L. L. Huang, 1981. Comparative genetics of hamster amylases. *Biochem. Genet.* 19: 623-633.
- Dirksen, E. & J. Levy, 1977. Virus-like particles in placentas from normal individuals and patients with systemic lupus erythematosus. *J. Natl. Cancer Inst.* 59: 1187-1192.
- Ellis, R. W., D. DeFeo, T. Y. Shih, M. A. Gonda, H. A. Young, N. Tsuchida, D. R. Lowy & E. M. Scolnick, 1981. The p21 src genes of Harvey and Kirsten sarcoma viruses originate from divergent members of a family of normal vertebrate genes. *Nature* 292: 506-511.
- Emi, M., A. Horii, N. Tomita, T. Nishide, M. Ogawa, T. Mori & K. Matsubara, 1988. Overlapping two genes in human DNA: a salivary amylase gene overlaps with a gamma-actin pseudogene that carries an integrated human endogenous retroviral DNA. *Gene* 62: 229-235.
- Fincham, J. R. S. & G. R. K. Sastry, 1974. Controlling elements in maize. *Annu. Rev. Genet.* 8: 15-50.
- Finnegan, D. J., 1989. Eukaryotic transposable elements and genome evolution. *Trends in Genet.* 5: 103-107.
- Foster, D. N., L. J. Schmidt, C. P. Hodgson, H. L. Moses & M. J. Getz, 1982. Polyadenylated RNA complementary to a mouse retrovirus-like multigene family is rapidly and specifically induced by epidermal growth factor stimulation of quiescent cells. *Proc. Natl. Acad. Sci. USA* 79: 7317-7321.
- Gardner, M., S. Dandekar & R. Cardiff, 1986. Molecular mechanism of an ecotropic MuLV restriction gene Akvr-1/FV-4 in California wild mice. *Current Topics in Micro. & Immunol.* 127: 338-345.
- Groot, P. C., M. J. Bleeker, J. C. Pronk, F. Arwert, W. H. Mager, R. J. Planta, A. W. Eriksson & R. R. Frants, 1989. The human α -amylase multigene family consists of haplotypes with variable numbers of genes. *Genomics* 5: 29-42.
- Gumucio, D. L., K. Wiebauer, R. M. Caldwell, L. C. Samuelson & M. H. Meisler, 1988. Concerted evolution of human amylase genes. *Mol. Cell. Biol.* 8: 1197-1205.
- Hayward, W. G., B. E. Neel & S. M. Astrin, 1981. Activation of a cellular *onc* gene by promoter insertion in ALV-induced lymphoid leukemia. *Nature* 21: 475-480.
- Hemenway, C., M. Kalff, J. Stavenhagen, D. Walthall & D. M. Robins, 1986. Sequence comparison of the fourth component of complement (C4) and sex-limited protein (Slp). *Nucl. Acids Res.* 14: 2539-2554.
- Hemenway, C. & D. M. Robins, 1987. DNase I-hypersensitive sites associated with expression and hormonal regulation of mouse C4 and Slp genes. *Proc. Natl. Acad. Sci. USA* 84: 4816-4820.
- Hodgson, C. P., P. K. Elder, T. Ono, D. N. Foster & M. J. Getz, 1983. Structure and expression of mouse VL30 genes. *Mol. Cell. Biol.* 3: 2221-2231.
- Horii, A., M. Emi, N. Tomita, T. Nishide, M. Ogawa, T. Mori & M. Matsubara, 1987. Primary structure of human pancreatic α -amylase gene: Its comparison with human salivary α -amylase gene. *Gene* 60: 57-64.
- Horowitz, M., S. Luria, G. Rechavi & D. Givol, 1984. Mechanism of activation of the mouse c-mos oncogene by the LTR of an intracisternal A-particle gene. *EMBO J.* 3: 2937-2941.
- Jenkins, N. A., N. G. Copeland, B. A. Taylor & B. K. Lee, 1981. Dilute (d) coat colour mutation of DBA/2J mice is associated with the site of integration of an ecotropic MuLV genome. *Nature* 293: 370-374.
- Jones, J. M., S. A. Keller, L. C. Samuelson, L. Osborn, M. P. Rosenberg & M. H. Meisler, 1989. A salivary transgene is efficiently expressed in liver but not in parotid gland of transgenic mice. *Nucl. Acids Res.* 17: 6613-6623.
- Keshet, E., R. Schiff & A. Itin, 1990. Mouse retrotransposons: A cellular reservoir of repeat (LTR) elements with diverse transcriptional specificities. *Adv. Cancer Res.* 56: 215-251.
- Kidd, S. & M. W. Young, 1986. Transposon-dependent mutant phenotypes at the Notch locus of *Drosophila*. *Nature* 323: 89-91.

- King, M. C. & A. C. Wilson, 1975. Evolution at two levels in humans and chimpanzees. *Science* 188: 107-116.
- Kingsman, A. J. & S. M. Kingsman, 1988. Ty: a retroelement moving forward. *Cell* 53: 333-335.
- Larsson, E., N. Kato & M. Cato, 1989. Human endogenous proviruses. *Curr. Top. Microbiol. Immunol.* 148: 115-132.
- Loreni, F., J. Stavenhagen, M. Kalff & D. M. Robins, 1988. A complex androgen-responsive enhancer resides 2 kilobases upstream of the mouse Slp gene. *Mol. Cell. Biol.* 8: 2350-2360.
- McClintock, B., 1956. Controlling elements and the gene. *Cold Spring Harbor Symp. Quant. Biol.* 21: 197-216.
- McGeachin, R. L. & J. R. Akin, 1982. Amylase levels in the tissues and body fluids of several primate species. *Comp. Biochem. Physiol.* 72: 267-269.
- Meisler, M. H. & D. L. Gumucio, 1986. Salivary amylase: evolution and tissue-specific expression, pp. 457-466 in *Molecular and Cellular Basis of Digestion*, edited by P. Desnuell, H. Sipstrom & O. Noren. Elsevier/North Holland Publishing Co., Amsterdam.
- Mount, S. M., 1992. Diversity among Drosophila transposable elements and in their effects on gene expression, in *Mechanisms of Eukaryotic DNA Recombination*, edited by M. E. Gottesman & H. J. Vogel. Academic Press, Orlando, FL (in press).
- Mount, S. M. & G. M. Rubin, 1985. Complete nucleotide sequence of the Drosophila transposable element copia: homology between copia and retroviral proteins. *Mol. Cell. Biol.* 5: 1630-1638.
- Nelson, J., J.-A. Leong & J. A. Levy, 1978. Normal human placentas contain RNA-directed DNA polymerase activity like that in viruses. *Proc. Natl. Acad. Sci. USA* 75: 6263-6267.
- Nishide, T., M. Emi, Y. Nakamura & K. Matsubara, 1986. Corrected sequences of cDNAs for human salivary and pancreatic α -amylases. *Gene* 50: 371-372.
- Nonaka, M., H. Kimura, Y. D. Yeul, S. Pattanakitsakul & M. Takahashi, 1986. Identification of the 5'-flanking regulatory region responsible for the difference in transcriptional control between mouse complement C4 and Slp genes. *Proc. Natl. Acad. Sci. USA* 83: 7883-7887.
- Ponte, P., P. Gunning, H. Blau & L. Kedes, 1983. Human actin genes are single copy for α -skeletal and α -cardiac actin but multicopy for β - and γ -cytoskeletal genes: 3' untranslated regions are isotype specific but are conserved in evolution. *Mol. Cell. Biol.* 3: 1783-1791.
- Repaske, R., P. E. Steele, R. R. O'Neill, A. B. Rabson & M. A. Martin, 1985. Nucleotide sequence of a full-length human endogenous retroviral segment. *J. Virol.* 54: 764-772.
- Roeder, G. S. & G. R. Fink, 1983. Transposable elements in yeast, pp. 299-328 in *Mobile Genetic Elements*, edited by J. A. Shapiro. New York: Academic Press.
- Rubin, G. M., 1983. Dispersed repetitive DNAs in Drosophila, pp. 329-361 in *Mobile Genetic Elements*, edited by J. A. Shapiro. New York: Academic Press.
- Rubin, G. M., M. G. Kidwell & P. M. Bingham, 1982. The molecular basis of P-M hybrid dysgenesis: the nature of induced mutations. *Cell* 29: 987-994.
- Samuelson, L. C., K. Wiebauer, D. L. Gumucio & M. H. Meisler, 1988. Expression of the human amylase genes: recent origin of a salivary amylase promoter from an actin pseudogene. *Nucl. Acids Res.* 16: 8261-8276.
- Samuelson, L. C., K. Wiebauer, C. M. Snow & M. H. Meisler, 1990. Retroviral and pseudogene insertion sites reveal the lineage of human salivary and pancreatic amylase genes from a single gene during primate evolution. *Mol. Cell. Biol.* 10: 2513-2520.
- Schiff, R., A. Itin & E. Keshet, 1991. Transcriptional activation of mouse retrotransposons in vivo: specific expression in steroidogenic cells in response to trophic hormones. *Genes & Dev.* 5: 521-532.
- Schmidt, M., T. Wirth, B. Kroger & I. Horak, 1985. Structure and genomic organization of a new family of murine retrovirus-related DNA sequences (MuRRS). *Nucl. Acids Res.* 13: 3461-3470.
- Shreffler, D. C., 1982. MHC-linked complement components, pp. 187-219 in *Histocompatibility Antigens: Structure and Function (Receptors and Recognition, series B, vol. 14)*, edited by P. Parham and J. Strominger. London: Chapman and Hall.
- Stavenhagen, J., F. Loreni, C. Hemenway, M. Kalff & D. M. Robins, 1987. Molecular genetics of androgen-dependent and -independent expression of mouse sex-limited protein. *Mol. Cell. Biol.* 7: 1716-1724.
- Stavenhagen, J. B. & D. M. Robins, 1988. An ancient provirus has imposed androgen regulation on the adjacent mouse sex-limited protein gene. *Cell* 55: 247-254.
- Steele, P. E., A. B. Rabson, T. Bryan & M. A. Martin, 1984. Distinctive termini characterize two families of human endogenous retroviral sequences. *Science* 225: 943-947.
- Stoye, J. P. & J. M. Coffin, 1985. Endogenous retroviruses, pp. 357-404 in *Molecular Biology of Tumor Viruses: RNA Tumor Viruses. Vol. 2*, edited by R. A. Weiss, N. Teich, H. Varmus & J. Coffin. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
- Stoye, J. P. & J. M. Coffin, 1988. Polymorphism of murine endogenous proviruses revealed by using virus class-specific oligonucleotide probes. *J. Virol.* 62: 168-175.
- Stoye, J. P. & J. M. Coffin, 1988a. Role of endogenous retroviruses as mutagens: the hairless mutation of mice. *Cell* 54: 383-391.
- Varmus, H. E., 1982. Form and function of retroviral proviruses. *Science* 216: 812-820.
- Varmus, H., 1988. Retroviruses. *Science* 240: 1427-1435.
- Weiss, R. A., N. Teich, H. Varmus & J. Coffin (eds.), 1985. *Molecular Biology of Tumor Viruses: RNA Tumor Viruses*, vols 1, 2. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.
- Wu, C.-I. & W.-H. Li, 1985. Evidence for higher rates of nucleotide substitution in rodents than in man. *Proc. Natl. Acad. Sci. USA* 82: 1741-1745.
- Yamamoto, K. R., 1985. Steroid receptor regulated transcription of specific genes and gene networks. *Annu. Rev. Genet.* 19: 209-252.
- Ymer, S., W. Q. J. Tucker, C. J. Sanderson, A. J. Hapel, H. D. Campbell & I. G. Young, 1985. Constitutive synthesis of interleukin-3 by leukemia cell line WEHI-3B is due to retroviral insertion near the gene. *Nature* 317: 255-258.