UCSF
UC San Francisco Electronic Theses and Dissertations

Title
Ribosomal frameshifting in retroviral gene expression

Permalink
https://escholarship.org/uc/item/16n161mf

Author
Jacks, Tyler Edwards

Publication Date
1988

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

Ribosomal Frameshifting in Retroviral Gene Expression

by

Tyler Edwards Jacks

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Biochemistry

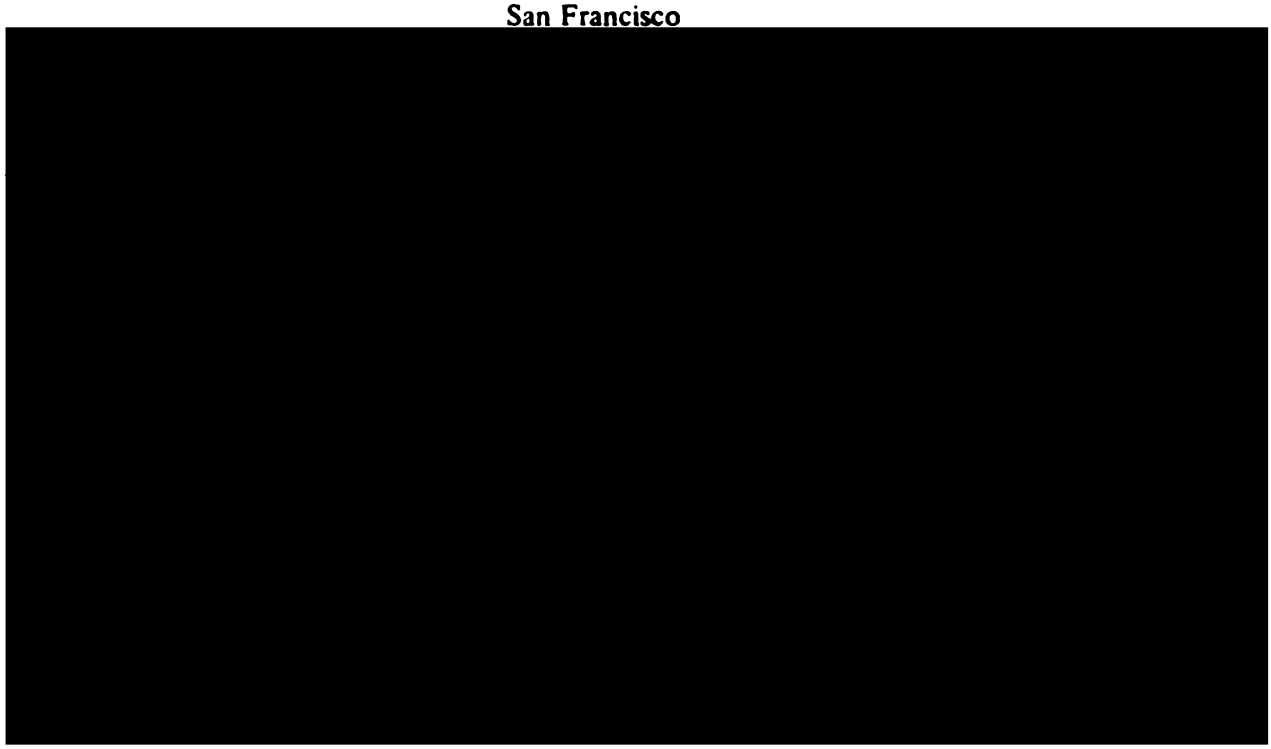
in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA

San Francisco



For my father

ACKNOWLEDGEMENTS

This thesis represents the end of my graduate years at UCSF. There are many people I wish to thank for getting me to this point. First, my collaborators who have contributed their time and energy to this project. Foremost among these is my advisor, **Harold Varmus**, who has provided numerous thoughtful ideas and always appreciated enthusiasm. I have learned a great deal from him. **Hiten Madhani** was a tireless worker during his year and a half in the lab. Hiten contributed some of the experiments described in Chapter 5, not to mention uncountable discussions that helped shape our conception of frameshifting. **Robin Colgrove**, the fourth floor's computer whiz, spent a great deal of time writing and running frameshift-detecting programs. I appreciate all of his efforts and advice. I refer to some of Robin's work in the Conclusions. Other valuable members of the "Frameshift Club", who donated either materially or intellectually to the cause are: **Peter Pryciak, Michael Glotzer, Greg Enders, Virginia Rath, Frank Masiarz, Mike Power, John Majors, Kim Townsley, Phil Barr,** and **Paul Luciw**. I am indebted to them all. I have benefitted a great deal from my friendship with three shifty characters from Salt Lake City: **John Atkins, Bob Weiss,** and **Ray Gesteland**. Other members of the Varmus lab and the UCSF community deserve particular thanks for their assistance in the completion of this project. **Titia de Lange** has been invaluable as my critical sounding board. She is also a model scientist, a terrific roommate, and a dear friend. **Bruce Bowerman**, along with being a classmate, labmate, and soulmate, made two key suggestions that speeded the progress of this project. Peter Walter and Pablo Garcia were

instrumental in teaching me various tricks of the protein trade. I also thank **Don Ganem** and **Sandy Johnson** for many helpful discussions and for their support. **Janine "Aqua" Marinos** typed the bulk of the manuscripts that are now Chapters 1-5; it was always a challenge and a great deal of fun. **Sue Adams** prevented me from donating all of my money to the U.C. Regents. All of her help over the years is greatly appreciated. Finally, I would like to recognize **Christine Guthrie**, who convinced me to come to UCSF and has been an ardent supporter, thoughtful advisor, and good friend. The first three chapters of this thesis have been published: Chapter 1 (*Science* 230, 1237-1242); Chapter 2 (*Proceedings of the National Academy of Science* USA 84, 4298-4304); Chapter 3 (*Nature* 331, 280-283). Chapters 4 and 5 have been submitted to *Cell*.

Beyond the science, there are many people who have helped make my years in San Francisco happy ones. Although the list of individuals is too long to include everyone's name, I would like to expressly thank: **Daryl Butler**, **Ann Tsukamoto**, **Jan Tuttleman**, **Caroline Goute**, **Suzanne Ortiz**, **Tony Brown**, **John Young**, **Cristoph Seeger**, **Kevin Mossie**, **Paul Bates**, and **David Morgan** for their friendship. Life would have been much less interesting without my good friends and fellow members of the Breakfast and South of Market Clubs: Bruce, **Trey Simmons**, and **Doug Kellogg**, **Susan DeGroat**, and **Paul Cusick**. All of that squash would not have been possible without my partners: Sandy, Harold, Don, Doug, **Peter Huang**, **Byron Hann**, **Steve Doxsey**, and the Dog. I thank my parents, **Barbara** and **Stanley Jacks**, and my sister **Alyson Jacks** for their love and valient efforts to understand what the hell I was up to. And last, and most important, I thank **Sally Gibbons**, for all her love and support, and for always understanding why I had to get out of bed at two in the morning to load another gel.

Ribosomal Frameshifting in Retroviral Gene Expression

Tyler Edwards Jacks

ABSTRACT

In all known retroviruses, the *pol* gene (which encodes the reverse transcriptase and integrase functions) lies downstream of the *gag* gene, which codes for the structural proteins of the virus core. In most retroviruses, *gag* and *pol* lie in different translational reading frames, either directly overlapping or else with a third gene (*pro*, encoding the protease) intervening them and overlapping both. Despite these apparent blocks to continuous translation, all retroviruses express the *pol* gene by first forming a *gag-pol* or *gag-pro-pol* fusion protein.

We have used *in vitro* transcription and translation techniques to help elucidate the mechanism of retroviral *pol* gene expression. When used to program rabbit reticulocyte lysate translation reactions, *in vitro*-synthesized mRNA corresponding to the *gag-pol* domains of either Rous sarcoma virus (RSV) or human immunodeficiency virus (HIV) yields both *gag* and *gag-pol* proteins, indicating that ribosomal frameshifting occurs in the *gag-pol* overlaps. The frequency of these frameshifting events *in vitro* (5-10% of translating ribosomes change frame) is high enough to account for *gag-pol* expression *in vivo*. A similar experiment using cloned DNA from mouse mammary tumor virus (MMTV), one of the viruses with an intervening *pro* gene, demonstrates that two highly efficient frameshifting events occur during expression of this virus's *gag-pro-pol* protein.

Amino acid sequencing and deletion and site-directed mutagenesis, principally with RSV, have suggested that frameshifting in retroviral gene expression requires two features of the mRNA: a site and a structure. The site is the position on the mRNA where the frameshift occurs. It is composed of two short homopolymeric sequences that appear to allow adjacent tRNAs to slip into the alternate (-1) reading frame. The second important feature, a downstream RNA structure, may act by impeding ribosome movement through the frameshift site and thereby allowing increased time for tRNA slippage. While the majority of these experiments were performed on the RSV *gag-pol* region, similar potential frameshift sites and downstream stem-loop structures exist in other retroviral genes that have been shown to or are believed to permit efficient ribosomal frameshifting.

Table of Contents

ii Dedication

iii Acknowledgements

v Abstract

vii Table of Contents

1 Introduction

44 Chapter 1

Expression of the Rous Sarcoma Virus *pol* Gene by Ribosomal Frameshifting

51 Chapter 2

Two Efficient Ribosomal Frameshift Events Are Required for Synthesis of Mouse Mammary Tumor Virus *gag*-related Polyproteins

57 Chapter 3

Characterization of Ribosomal Frameshifting in HIV-1 *gag-pol* Expression

62 Chapter 4

A Detailed Analysis of the Ribosomal Frameshift Site in the Rous Sarcoma Virus *gag-pol* gene

107 Chapter 5

Ribosomal Frameshifting in the Rous Sarcoma Virus *gag-pol* gene Requires Downstream Secondary Structure.

144 Conclusions

Introduction

Introductory remarks. Historically, the study of viruses, especially those of higher eukaryotes, has had dual motivation. On the one hand, it is hoped that a basic understanding of viruses will help prevent or ameliorate the many diseases caused by them. A second motivation is the oft-supported claim that viruses are effective probes into the inner workings of the cells they parasitize. The small size, relative simplicity, and rapid expression of their genomes render viruses attractive objects of experimentation. As viruses rely on their hosts for part or all of the expression of the information required for their propagation, the understanding of virus replication can hold profound lessons for related processes in uninfected cells. Thus, the study of viruses and phage has lead to important findings regarding the structure and expression of genes, the synthesis and processing of proteins, and cellular import and export.

This thesis concerns the discovery of another eukaryotic cellular process through the study of viruses, here, retroviruses. The examination of this process, the efficient changing of translational reading frame by ribosomes (or ribosomal frameshifting), has not not only clarified an enigma in retrovirology and thereby explained the mechanism of expression of important viral genes, it has furthered our understanding of more basic issues in protein synthesis. In many cases in the past, once novel viral mechanisms were later found to be utilized by the host cell. While no frameshift-controlled eukaryotic cellular genes have yet surfaced, the discovery of the use of frameshifting by retroviruses has heightened the awareness of this mechanism as a viable means of gene control. The elucidation of the rules governing

frameshifting in retroviral genes should hasten the discovery of similarly controlled cellular genes in higher cells. Finally, although medical applicability was not a prime motivator in this research, the fact that human immunodeficiency virus (HIV), the causitive agent of AIDS, utilizes ribosomal frameshifting in the production of critical proteins leaves open the possibility for effective antiviral therapy directed against this process.

The INTRODUCTION will be divided into two parts: in the first part, I will review the relevant data from retrovirology that has defined the problem addressed in the body of the thesis; the second part will focus on the solution to the problem, ribosomal frameshifting, and more generally on translational accuracy. The INTRODUCTION will be followed by five CHAPTERS that consist of published (or submitted) papers concerning the discovery (chapters 1-3) and mechanism (chapters 4 and 5) of frameshifting in retroviral gene expression. In the CONCLUSIONS section, I will summarize our current conception of frameshifting in retroviruses and describe preliminary and planned experiments that are designed to further our understanding of this process.

THE gag-pol STORY

Beginning in the early 1970s, retrovirology experienced a boom. Initiated by the discovery of reverse transcriptase (1,2), an enzyme with the heretical ability to convert information stored in RNA into DNA, and spurred on by the identification of cellular homologues to the oncogenes carried by acutely transforming retroviruses (3), the interest and effort in the study of retroviruses increased rapidly. Among the many questions asked and answered beginning in those years were: is the RNA genome intact or segmented?; is the genetic information in a diploid or haploid state?; what are the details of the synthesis of a DNA provirus from an RNA genome?; how is that DNA provirus integrated into the host's chromosomes?; what is the nature of genes encoding the various structural and enzymatic functions of the virus?; how many genes are there and how are they expressed?. What follows is a small collection of moments from this burgeoning field, a fast-forward slide show of one particular problem. The study of the *pol* gene, the gene that encodes reverse transcriptase, had its roots in the earliest efforts to understand the mechanics of retroviruses. The details of the expression of this gene are still being addressed today.

The retroviral pol gene: function, location, and expression. The *pol* gene encodes the proteins reverse transcriptase (a polymerase responsible for synthesizing a DNA provirus from genomic RNA), integrase (implicated in the integration of proviral DNA into the host chromosome), and sometimes protease (the enzyme that cleaves viral

polyproteins into their mature forms). *pol* is the central of three replication genes carried by all replication-competent retroviruses. This assignment was first suggested by early mapping studies of genomic RNA (4-6) and by the analysis of viral mRNAs using specific radiolabelled probes (7,8) and *in vitro* translation (9-15). It has been confirmed for many retroviruses through sequencing of proviral DNA (see below).

The *pol* gene is bordered on its 5′ side by the *gag* gene (for group-specific antigens) which encodes a polyprotein composed of the structural proteins of the virus core. The *env* gene lies 3′ to *pol*; it encodes a polyprotein that is cleaved to produce the two envelope glycoproteins of the virus. Many retroviruses carry additional genes, usually located toward the 3′ end of the genome. These additional genes may encode oncoproteins or transactivators of one sort or another. They will not be considered further.

After the early steps of attachment, penetration, reverse transcription, and integration, the retroviral life cycle co-exists with that of the host cell. The integrated provirus gives rise to two major transcripts: a genome-length mRNA, which is indentical to the RNA packaged into virions, containing all three replication genes (*gag*, *pol*, and *env*) (7,8) and a subgenomic mRNA carrying solely the *env* gene (7,8). Following the simple rule that eukaryotic ribosomes generally initiate translation only at the 5′ proximal open reading frame in an mRNA, these two messages should yield the *gag* and *env* polyproteins, respectively. Indeed, *in vitro* translation of these mRNAs gives rise to *gag* and *env* proteins that are indistinguishable from the primary translation

products of these genes in infected cells (9-15,17,18).

At a first glance, then, the expression of *pol* is an enigma, since there is no mRNA species in which *pol* is the 5'-most gene. The first clue to the solution to this problem came from the analysis of *pol*-specific proteins produced in infected cells using antibodies directed against reverse transcriptase. This work showed that the primary translation product of the *pol* genes of Rous sarcoma virus (RSV) and murine leukemia virus (MLV) was a large (180-200 kD) polyprotein that also contained sequences encoded by the upstream *gag* gene (19,20). Pulse-chase experiments demonstrated that the *gag-pol* fusion protein was the precursor of the reverse transcriptase protein and that it was not the precursor of the *gag* polyprotein, which is present in infected cells approximately twenty times more abundantly (19,20). For RSV a comparison of the tryptic peptides generated from the two proteins indicated that most if not all of the sequences present in the *gag* protein were also present in the *gag-pol* protein (19,21). Several workers also showed that in addition to the *gag* protein, a high molecular weight *gag-pol* fusion protein was produced upon *in vitro* translation of genome-length mRNA (or virion RNA (vRNA)) of RSV and MLV (12,13,15). The ratio of the two proteins produced *in vitro* was similar to that observed *in vivo*, approximately 20:1.

The gag-pol problem. That the initial protein product of retroviral *pol* genes was a *gag-pol* fusion protein answered the question of how *pol* could be expressed without appearing at the 5' end of any mRNA species. However, this solution raised another, perhaps more interesting question: How could the same genome-length mRNA give rise

to both the *gag* and *gag-pol* proteins? What might have been termed the "*pol* problem" had been transformed into the "*gag-pol* problem." The various groups that first defined the *gag-pol* problem suggested two possible mechanisms to solve it:

(i) The suppression hypothesis. According to this model, the genome- length mRNA is the template for synthesis of both the *gag* and *gag-pol* proteins. Normally, translation of this mRNA generates the *gag* protein. At some low efficiency, however, the signal or signals that usually terminate *gag* translation are read-through, allowing continued translation into *pol* and the synthesis of the *gag-pol* fusion protein. By this scheme the ratio of the two proteins would be determined by the efficiency of the translational suppression event. This model was bolstered by the abundant evidence for termination suppression in prokaryotes. Also, it was known that QB phage utilized stop codon suppression to synthesize two versions of its coat protein (20). The suppression hypothesis is compatible with the observation that *in vitro* translation of vRNA yields both *gag* and *gag-pol* proteins; it also predicts the observed representation of *gag* sequences in the *gag-pol* fusion protein.

(ii) The splicing hypothesis. The second hypothesis advanced to account for *gag-pol* expression called for the production of a low-abundance mRNA specific to the *gag-pol* protein. Here, an inefficient processing event was proposed to convert a fraction of the genome-length mRNA into a message in which *gag* and *pol* were fused into one large open reading frame. According to this model the ratio of *gag* to *gag-pol* would be a function of the efficiency of the necessary processing event.

The supporters of this model were buoyed by the recent discovery of mRNA splicing. In fact, at approximately this time the retroviral *env* mRNA was proposed to be the product of a splicing event. The ability to produce *gag-pol* protein upon translation of vRNA was explained away by postulating the occasional packaging of the very similar *gag-pol* mRNA into virions in place of the unspliced, genome-length message. The abundance of *gag* sequences in the *gag-pol* protein suggested that the 5′ splice junction in the processed mRNA should be near the end of the *gag* gene.

The same experiment: different viruses, different results. The first attempt to distinguish between the suppression and splicing hypotheses was performed by Phillipson *et al.* in 1978 (23). As mentioned above, *in vitro* translation of MLV or RSV RNA gives rise to a characteristic ratio of gag to *gag-pol* proteins. These authors reasoned that if the mechanism of expression of *gag-pol* were the simple suppression of the *gag* terminator, then the addition of nonsense suppressor tRNAs to such an *in vitro* translation should augment synthesis of the *gag-pol* protein. In fact, when yeast amber suppressor tRNA was added to a rabbit reticulocyte lysate translation of MLV vRNA, the synthesis of the *gag-pol* fusion protein was increased and the level of the *gag* protein diminished. This experiment suggested that in MLV *gag* and *pol* were in the same translational reading frame and separated by a single amber stop codon. This configuration is at least compatible with the suppression hypothesis.

DNA sequencing of an MLV provirus later confirmed this presumed genetic structure (24). However, the nucleotide sequence at the end of

gag did not offer any indication of how the proposed read-through of the *gag* terminator might occur. In fact, Shinnick *et al.* (24) noted a potential stable stem-loop structure involving nucleotides at the MLV *gag-pol* junction and suggested that this structure might be mistakenly recognized by a pre-tRNA processing enzyme and removed, creating a separate *gag-pol* mRNA.

Shortly after the report that an MLV *gag-pol* protein could be synthesized *in vitro* with the addition of nonsense suppressor tRNAs, a similar experiment performed with RSV vRNA produced a contrary result. As had been found with MLV, Weiss *et al.* (25) observed that *in vitro* translation of RSV vRNA in the presence of yeast amber suppressor tRNA reduced the yield of *gag* protein. However, rather than producing a corresponding increase in the amount of *gag-pol* protein, this treatment resulted in the appearance of an extended *gag* protein and no additional *gag-pol* protein. The conclusion from this experiment was that the RSV *gag* gene is also terminated by an amber stop codon, but this stop codon is followed by a second one shortly downstream. Further, since the efficiency of two successive stop codon suppression events was thought to be insufficient to allow synthesis of the observed levels of *gag-pol* protein, and given that other types of translational suppression (for example, frameshift suppression) had not been described in eukaryotes, Weiss *et al.* (25) concluded that the most likely mode of *gag-pol* expression for RSV was the production of a spliced *gag-pol* mRNA.

The RSV sequence and the beginning of the end for the suppression hypothesis. The nucleotide sequence of RSV reported by Schwartz, Tizard and Gilbert in 1983 (26) clarified the genetic structure of the RSV *gag-*

pol domain. The authors' consideration of the sequence was instrumental in obscuring the ultimate solution to the *gag-pol* problem. Consistent with the *in vitro* translation result, they found that the RSV *gag* terminates with an amber stop codon and that this stop codon is followed in 111 nt by a second amber terminator in the *gag* reading frame. The *pol* open reading frame (identified by its position relative to *gag* and the presence of a coding region whose predicted amino acid sequence matched the known N-terminal sequence of RSV reverse transcriptase) is in a different translation reading frame than *gag*. The 5' end of the *pol* open reading frame overlaps the 3' end of *gag* by 58 nucleotides in the -1 direction. As defined by Schwartz *et al.* (26), the *pol* "gene" began with the portion known to be translated into reverse transcriptase, located some 20 nucleotides downstream of the *gag* terminator. With an incidental nod to the possibility for ribosomes to shift reading frame during translation of the 58 nucleotide overlap between *gag* and *pol*, these authors firmly concluded that the only reasonable way to synthesize a *gag-pol* protein in RSV would be from an RNA, derived from the genome-length mRNA by splicing, that had *gag* and *pol* fused in frame.

This view, first formed with respect to RSV, soon dominated the field of retrovirology generally. This is indicated most obviously in the treatment of the subject in the comprehensive review *RNA Tumor Viruses* (27). Largely on the evidence presented above, the authors make several references to the near necessity of a spliced *gag-pol* RNA for RSV. At one point it is claimed that such a species "must" exist (p. 581). Since the replication strategies of different retroviruses are

similar, it was generally believed that MLV, for which stop codon suppression was a structural possibility, and other retroviruses also expressed the *gag-pol* protein from a separate, spliced mRNA.

Indisputable evidence: a gag-pol intron. The suppression hypothesis fell deeper into obscurity with the first nucleotide sequence of human T-cell leukemia virus type 1 (HTLV-1) (28). This sequence included a 300 nucleotide "intergenic" region between the *gag* and *pol* open reading frames that was closed in all three reading frames by multiple stop codons. As such, translation from *gag* into *pol* would require multiple termination and/or frameshift suppression events. Although physical evidence for a spliced *gag-pol* mRNA for any retrovirus was still lacking, this presumed intron in HTLV-1 left no reasonable alternative.

An editorial comment. One expects that given the dominant position of the day, the discovery of an intron between HTLV-1 *gag* and *pol* was more than unsurprising; it was a welcomed relief. If *gag* and *pol* were joined by splicing, there should be viruses with untranslatable sequences between these genes. In retrospect, we can even speculate that such a preconception might have actually influenced the care given in obtaining or scrutinizing the DNA sequence in the relevant region, or the attention afforded the HTLV-1 sequence given that the sequenced clone was not known to be infectious. This reflection is warranted because it is now known that the original sequenced clone of HTLV-1 is not infectious and that the sequence reported by Seiki *et al.* (28) has numerous errors in the region between *gag* and *pol*. Sequences from other HTLV-1 isolates (29) have shown that this virus, like several others

(see below), has an additional open reading frame located between *gag* and *pol* that overlaps them both and is believed to encode the viral protease. Thus, the death knell for the suppression hypothesis was prematurely sounded.

A startled reawakening. The reinvigoration of the *gag-pol* problem required the results of an experiment that was not designed to solve it. Yoshinaka *et al.* (30), in a continuing effort to understand retroviral proteins at the amino acid sequence level, purified and sequenced the viral protease found in MLV virions. This protein is expressed as part of the *gag-pol* fusion protein; it is responsible for cleaving itself and other mature viral proteins from their precursors. It was known from crude mapping and sequence comparisons that the MLV protease was encoded upstream at the reverse transcriptase domain, near the 5′ end of *pol* (31). The N-terminal amino acid sequence determined by Yoshinaka *et al.* (30) showed, however, that the protein actually spans the *gag-pol* junction. The first four amino acids of the protease are encoded by the last four codons of *gag*. The fifth amino acid is glutamine, and the reminder of the protein is encoded by *pol*, beginning with the codon that immediately follows the *gag* terminator. From this amino acid sequence it was simple to deduce the mechanism of MLV *gag-pol* expression: suppression of the *gag* amber terminator by a glutamine-charged tRNA. Since all of the nucleotides at the *gag-pol* junction were required to encode the protease, a spliced *gag-pol* mRNA was definitively excluded.

This result rapidly changed the conception of the *gag-pol* problem. No longer was it a tired search for the elusive spliced *gag-pol* mRNA that would lay to rest any possible objections to the splicing

hypothesis. With the knowledge that MLV utilized one form of translational suppression and the continued belief that retroviruses utilize like replication strategies, the possibility arose that RSV used what would be a novel expression mechanism in higher eukaryotes: translational frameshifting. Moreover, the emerging nucleotide sequences of several retrotransposons (32-35) (transposons that utilize reverse transcription during transposition) showed overlapping *gag*- and *pol*-like genes, suggesting that this mechanism might be at work in a diverse set of organisms from yeast to man. In fact, the yeast transposable element TY-1 was shown to encode a fusion protein from its *tya* (*gag*-like) and *tyb* (*pol*-like) genes, despite the fact that *tyb* overlapped *tya* in the +1 reading frame and no spliced *tya-tyb* mRNA could be detected (32,33).

Additional retrovirus sequences appeared at this time and further heightened interest in the possibility of ribosomal frameshifting. Two viruses, HTLV-2 (36) and bovine leukemia virus (BLV) (37,38), were shown to carry a separate open reading frame for the protease that lay between *gag* and *pol* and overlapped them both. These viruses would require two separate highly efficient frameshifting events to produce appreciable amounts of a *gag-pro-pol* fusion protein. Finally, the nucleotide sequences of the various HIV-1 isolates showed that, as in RSV, *gag* and *pol* overlapped out of frame (39-41). If ribosomal frameshifting were responsible for HIV-1 *gag-pol* expression, perhaps this mechanism could be targeted for antiviral therapy.

A final source of excitement over this possible solution to the *gag-pol* problem was less concerned with retroviruses than with their

cellular hosts. Unlike splicing, ribosomal frameshifting had not previously been described in higher eukaryotic cells. The discovery of the mechanism in retroviral gene expression might portent similarly controlled cellular genes and thereby expand our knowledge of the vast repertoire of cellular mechanisms of gene expression. If no cellular genes were found to require ribosomal frameshifting, understanding the mechanism of frameshifting in retroviral gene expression would at least help explain how such errors are normally avoided.

FRAMESHIFTING

The effective translation of any code requires knowing where to start, where to finish, and how to correctly proceed from the beginning to the end. These three elements of effective translation are addressed by a cell's protein synthesizing machinery during the three phases of translation of the genetic code: initiation, elongation, and termination. To accurately begin translation, specific signals in the form of a specialized codon or codons and optimize nucleotide context are recognized by the ribosome and initiation factors. Similarly, translational termination is achieved in a rather simple way. Three of the 64 code words are the genetic equivalent of grammatical periods, and when encountered signal the protein synthesizing machinery to stop and release its product.

The elongation phase, the orderly progression from codon to codon down the mRNA, is on the surface the most problematic. Two considerations make it so. First, each new codon brought into the

ribosomal A site must be accurately decoded. Because tRNA selection is principally determined by the base pair interaction between the tRNA anticodon and the A site codon, and since the difference in free energy between "correct" and "incorrect" base pairs can be as little as 2kcal/mole (42), insertion of the appropriate tRNA is a formidable task. The task seems more difficult when one considers that the correct tRNA is outnumbered by incorrect tRNAs approximately 20:1. As discussed below, the cell has devised a clever solution to the problem.

While accuracy of the decoding process is extremely important for efficient protein synthesis, another aspect of the elongation phase, the maintenance of the reading frame, is more important still. If it occurred in a critical amino acid position, a missense error could result in an inactive protein. But many missense errors would be expected to lead to at least partially, and perhaps fully, active products. Reading frame or "frameshift" errors, on the other hand, would nearly always be lethal for the resulting protein. After a frameshift error, translation proceeds in one of two alternate frames accumulating what amount to missense mistakes until the first termination codon in that frame is encountered. Thus, unless it occurred very near the 3′ end of the gene or was compensated by a second, complementary frameshift, an unplanned frameshift error would be catastrophic.

This section will concentrate on frameshift errors, how often they occur, what influences their frequency and by what mechanism, and, finally, some examples where such errors are not errors at all. Throughout I will refer to a common theme in the study of this and other

error processes: by understanding the error, we may better understand its normal counterpart. In the case of frameshifting we are attempting to clarify a very poorly understood and yet critically important process, the proper movement of the translational apparatus by precisely three message nucleotides in each cycle. The two solutions to this problem (stated in their extreme forms) are: 1) the ribosome measures the translational step regardless of the nature of the tRNA-mRNA interaction; and 2) the precise translocation of three nucleotides during normal translation is the result of the three base pair association between tRNA and mRNA. While the study of frameshifting has continually addressed this problem, much is yet to be learned.

An historical perspective. From the earliest days in the study of protein synthesis and the genetic code, frameshift mutations, small insertions or deletions of nucleotides that lead to a change in the reading frame, were thought to be non-leaky. In part for the theoretical considerations discussed above, but mainly from empirical observations, this type of mutation was seen as completely inactivating. In fact, frameshift mutations (defined in part by their non-leaky character) of both polarities (+ and -) were critical to Crick and his co-workers in their pioneering work that accurately predicted the nature of the genetic code (43).

This conception of frameshift mutations as non-leaky has had a profound impact on the study of frameshifting as a process and, with reference to the above discussion, on the possibility of frameshift-controlled genes. If the translation process never compensates for a physical disruption of the proper reading frame (in the form of a

frameshift mutation), it could be safely assumed that translational frameshifts are extremely infrequent on wild-type mRNAs.

Error frequencies. There are very few estimates of the spontaneous frequency of translational frameshifting. This is due in part to the inherent difficulty of the experiment. Error frequencies in decoding can be deduced from the appearance of novel radioactive amino acids, electrophoretic mobilities, or chemical or protein cleavage sites in otherwise normal proteins or peptides. Conversely, in order to determine frameshifting frequency one must begin with a frameshift mutant and attempt to detect the small amount of normal product amidst a sea of defective protein.

The only available estimate for backround frameshifting frequency derives from the study of B-galactosidase activity of various frameshift alleles of *lac*Z in *E. coli*. The activity values obtained by Atkins *et al.* (44) were converted into a frameshifting frequency of $3x10^{-5}$ by Kurland (45). There are several reasons why this figure might be inaccurate. First, it is based on frameshifting events at a limited number of codons, all of which are present in a single gene. Second, the deduced frequency requires the assumption that the alternate open reading frame surrounding the original frameshift mutation is 20 codons in length. And, finally, the choice of frameshift mutants used as starting materials was influenced by the preconception that these mutants should be non-leaky. Other acridine-dye-induced *lac*Z mutations (probably also frameshift mutations) that showed relatively high B-galactosidase activity were not tested (44).

With all of the caveats aside, the figure of $3x10^{-5}$ frameshift

errors/codon does not seem unreasonable. As discussed above, this type of error should be less common than missense errors. The various estimates for missense error frequency are between 10^{-3} and 10^{-4}/codon (reviewed in ref. 68).

Factors that influence frameshifting. Numerous factors can affect how often ribosomes change reading frame. Increases in frameshifting frequency have been observed with mutant ribosomes, elongation factors, and tRNAs, addition of the aminoglycoside antibiotic streptomycin, and alterations in the ratio of tRNA species both *in vivo* and *in vitro*. Other ribosomal mutations, one in a protein, another in an rRNA, lower the frameshifting frequency in certain settings. Not surprisingly, the sequence in and around a codon greatly influences the level of frameshifting there. And it has recently been suggested that the pattern of codons in an mRNA could affect how well ribosomes maintain the proper reading frame on it. All of these factors will be discussed individually below.

Frameshift suppressors. As mentioned above, frameshift frequency can be assayed by comparing the activity of a frameshift mutant with wild-type. Similarly, mutations that enhance frameshifting efficiency can be scored by the ability of mutant cells harboring a frameshift allele of a selectable gene to grow under selective conditions. In this manner numerous extragenic frameshift suppressors have been isolated, principally from *Salmonella* and yeast.

Extragenic suppressors of +1 frameshift mutations are nearly always mutant tRNA species (46-54). +1 frameshift suppressor tRNAs have extended anticodon loops, usually by virtue of a base insertion adjacent

to the anticodon (46-49,53). Insertions elsewhere in the anticodon loop and disruption of the top base pair in the anticodon stem can also give rise to frameshift suppressor tRNAs (51,54).

The mechanism (or mechanisms) of action of these suppressor tRNAs remains unclear. It was initially postulated that the additional nucleotide in the anticodon loop allowed suppressor tRNAs to bind and translocate an additional message nucleotide (48). The consequence of this four-nucleotide translocation would be the appearance of the next +1 frame codon in the ribosomal A site and the correction to the original reading frame. (The +1 frame downstream of the +1 insertion is the 0 frame of wild-type.) The action of these suppressors first suggested that the anticodon-codon interaction might determine the step size of translocation. The model was supported by the observation that the first frameshift suppressor tRNAs acted only at a subset of sites. Suitable sites seemed to provide the opportunity for the fourth base pair between anticodon and codon.

Contrary to the simple four-base-pair-four-nucleotide-translocation model, Bossi and Roth (50) discovered a +1 frameshift suppressor tRNA that appeared not to require fourth base pair. The wild type version of this suppressor is an ACA-reading threonine tRNA; the suppressor carries an additional anticodon loop nucleotide between wild-type positions 31 and 32. The suppressor will "read" the four nucleotide codons, ACAA, ACAC, and ACAU. Bossi and Smith (51) have proposed that this suppressor (as well as a similar one from yeast (54)) might function by obscuring one nucleotide from the incoming A site tRNA. Thus, a normal three nucleotide anticodon:codon interaction is envisioned, followed by a

normal three nucleotide translocation. However, because of the extra-large anticodon loop, the now P-site suppressor tRNA obscures the next message nucleotide. Thus, the next triplet available for decoding is in the +1 reading frame.

Bossi and Smith (51) have argued that the difference between +1 suppressors that require a fourth base pair and those that do not lies in the conformations of the anticodon loops. As shown in Figure 1A, if the eight-nucleotide anticodon loop is arranged such that the anticodon sits at the top of a five-nucleotide stack on the 3′ side of the loop (a so-called "3-5" conformation), a three nucleotide translocation will result and the frameshift will follow from steric hindrance of the next 0-frame tRNA. Bringing an additional nucleotide into the 5′ stack (the 2-6 conformation) effectively creates a four-nucleotide anticodon (Fig. 1A). If a matching four-nucleotide codon is encountered, a four-nucleotide translocation and consequent +1 frameshift will ensue.

Curran and Yarus (57) have refined this concept of extended anticodon loops and alternative conformations further, and in so doing have relieved the tRNA of its role as determinant of the translocation step-size altogether. These authors combine the effects of all known +1 suppressors (as well as some missense suppressors and unusual wild-type tRNAs) with a systematic experimental analysis of three- and four-nucleotide translation by various constructed +1 suppressor tRNAs. They conclude that four nucleotide translocation never occurs. Instead, four-nucleotide translation is accomplished by tRNAs with extended anticodon loops in a manner akin to that suggested by Bossi and Smith (51): after three-nucleotide translocation, a fourth message nucleotide

is occupied by the P-site suppressor tRNA (see Fig. 1A). The effect of potential base-pairing in this fourth nucleotide position for some suppressors is explained by the stabilization of the anticodon loop conformation that allows the occupation of the fourth message nucleotide in the P site (the 2-6 conformation shown in Fig. 1A).

Ironically, the model of Curran and Yarus (57) is based on the same or similar data that first suggested an active role of the tRNA in determining translocation step size. However, the current model states that three-nucleotide translocation is unrelated to codon:anticodon base pairs, and suggests instead that precise three-nucleotide progression down the message is achieved by some fundamental process of the ribosome. Translocation is viewed here as the fixed movement of the tRNA from one point in the ribosome (the A site) to another (the P site) and the concomitant displacement of three message nucleotides. This conception is similar to that proposed by Watson in 1964 (58). I will return to still other models of translocation and their implications for frameshifting below.

Other types of extragenic suppressors. +1 frameshift suppressor tRNAs are the most common of the extragenic frameshift suppressors. However, other types of mutations will also enhance frameshifting frequency. -1 frameshift mutations in the *Salmonella trp*E gene can be suppressed by mutations in a glycine tRNA (59). A base substitution at position 34 (U to C) or other changes in the body of the $tRNA^{Gly}$ appear to confer the ability for doublet decoding. Interestingly, since this tRNA species is the only one capable of decoding GGA codons in *Salmonella*, the suppressor tRNAs must retain the ability to decode

triplets as well. Although not yet fully characterized, a frameshift suppressing form of elongation factor Tu has been isolated in a similar selection in *Salmonella* (J. Atkins, personal communication). The mechanism of action of this suppressor is completely unknown (although see below). Also, suppressors at *his4* frameshift alleles in yeast include some non-tRNA mutations (60). These recessive mutations may affect ribosomal proteins or translation factors.

As suggested by Watson (58) and recently reaffirmed by Curran and Yarus (57), translocation may be a physical property of the ribosome. If so, one would expect that selections for increased frameshifting would turn up mutations in ribosomal proteins involved in the process. The recessive suppressors in yeast mentioned above (60) are candidates for this type of mutation. Interestingly, though, there are no reported frameshift-stimulating mutations in elongation factor G (EF-G or EF-2 in eukaryotes), the one factor known to be required for the translocation step.

A link between decoding accuracy and frame maintenance. While no selected frameshift suppressors have been shown to encode ribosomal proteins, certain ribosomal protein mutations do enhance frameshifting frequency. The *str*A and *ram* mutations, selected by Gorini and co-workers (61,62), are principally studied for their effects on decoding accuracy. These two mutations reside in the genes for the *E. coli* small ribosomal subunit proteins S12 and S4, respectively, and were the first of a number of ribosomal protein mutations isolated that increase or decrease the level of suppression of missense and nonsense lesions (reviewed in ref. 63). Atkins *et al* (44) showed that *str*A and *ram*

ribosomes also display altered frameshifting frequencies. As is true with missense and nonsense errors, *str*A causes a decrease in frameshifting errors on frameshift mutant *lac*Z genes; the *ram* mutation has the opposite effect. Also, the general error-promoting aminoglycoside streptomycin causes increased frameshifting in this setting (44).

Before considering how the various accuracy modulating mutations might influence frame maintenance, I will first discuss how they act to influence the fidelity of decoding. In essence, these mutations seem to act through alterations of the elementary rates in the pathway leading from initial tRNA binding to peptide bond formation. This pathway is bifurcated with a decision step, occurring after initial binding and GTP hydrolysis but before peptidyl transfer, in which the tRNA can either go on to donate its amino acid to the growing peptide or dissociate from the ribosome. This process is termed kinetic proofreading and was suggested nearly simultaneously by Hopfield (64) and Ninio (65). Proofreading affords the ribosome a second chance to evaluate the correctness of a tRNA anticodon-codon interaction. By using its inherent ability to distinguish correct from incorrect in two separate, sequential steps, the ribosome can effectively square the accuracy attainable with a single selection step. There is now ample experimental support for proofreading during protein synthesis (reviewed in ref. 63).

Accuracy-modulating mutant ribosomal proteins and antibiotics appear to affect how quickly tRNAs flow through the initial and proofreading selection steps (66). A mutation that slows a forward

reaction rate or increases a reverse rate would be expected to enhance fidelity by allowing the ribosome increased time to evaluate the correctness of the anticodon-codon interaction. Conversely, speeding up the process of translation by adjusting these elementary rates should result in increased errors. In fact, the error-suppressing *str*A mutation in S12 causes an overall slowing of the reaction leading to peptide bond formation (66). The *ram* mutation, which leads to increased error frequencies, speeds up translation by increasing this same reaction rate (66).

Note that this relationship between kinetics and accuracy would suggest that translation could proceed with fewer errors so long as enough time was allotted to it. The fact that the cell has chosen to accept a certain error frequency is assumed to reflect the need for efficient protein synthesis (67). I will return to the relationship between error frequencies and translation rates below.

It is not clear from the above discussion why increased decoding errors should lead to frameshift errors, especially given the view that translocation is a physical property of the ribosome. Any tRNA that survives to the peptidyl bond stage should, by this view, be automatically transferred to the P site along with three message nucleotides. The coupling of incorrectly paired tRNAs and translocation errors would suggest that translocation is sensitive to the anticodon-codon interaction. This so-called "error coupling" hypothesis (68) would be strengthened if it could be shown that the increased frameshifts caused by the *ram* mutation, steptomycin, or the *Salmonella* EF-Tu mutant resulted directly from improperly paired tRNAs. Such a

demonstration might be impossible, however. These agents would be expected to promote frameshifting at multiple sites and would consequently produce a heterogeneous population of protein molecules resulting from various frameshifting events. As such, the nature of the individual events would difficult to ascertain.

Error coupling and pool bias. A direct link between improper decoding and frameshifting can be established by a series of experiments that fall under the general heading of pool bias. The purpose of these experiments is to try to force decoding errors by adjusting the tRNA pools in order to then see whether frameshifting frequency is also raised. Pool bias experiments can be performed in vivo in *E. coli* cells that have their starvation response (ppGpp) system mutationally inactivated (69) or in cell-free extracts (70).

Whether the tRNA pools are altered by in vivo starvation for a particular amino acid in an auxotroph, inactivation of a temperature-sensitive aminoacyl-tRNA synthetase, or the poisoning of a biosynthetic pathway (69), along with the expected increase in decoding errors, frameshift errors do also increase. Similarly, addition of purified tRNA species to an *E. coli* cell-free translation system can dramatically increase frameshifting frequency (70). Thus, as was first suggested in explaining the effects of *ram* ribosomes, selection of an improper tRNA into the ribosomal A site can lead to an inappropriate translocation to the P site.

An interesting extension of the pool bias experiment is the readjustment of error-promoting pools in order to determine which tRNA species cause these errors and where they act. In the experiments

performed in vivo, the identity of the error-promoting tRNA species (or "culprit tRNA") can be deduced by systematically lowering the concentration of other charged tRNA species through starvation for other amino acids (69). The amino acid whose absence originally brought on the high error rate indicates at what codons the errors are occurring (or "shifty codons"). In the *in vitro* experiments, the lowering of the increased error frequency by addition of a second purified tRNA species identifies the shifty codon; the original error-promoting tRNA is the culprit (70).

Weiss (71) has proposed a model that accounts for the action of all known culprit tRNAs. He observed that the second and third anticodon positions (nucleotides 34 and 35) of all of the identified culprit tRNAs could pair with the first and second positions of their cognate shifty codons. This arrangement would place the uracil residue present in the 33rd position of all known tRNAs in apposition with the third codon position. This type of anticodon-codon pairing was termed "U33 grapple" pairing (71; Fig. 1B).

The proposed mechanism through which U33 grapple pairing leads to translocation errors, errors in both the + and - direction, depends on an earlier model of translocation postulated by Woese in 1970 (72). The reciprocating ratchet model of Woese is the epitome of the tRNA-centric translocation models. Stated simply, it calls for the successive interconversion of two conformational states of the anticodon loop to physically drag the mRNA through space (Fig. 1B). Woese extended an earlier observation of Fuller and Hodgson (73) who noted two apparently equally stable conformations of the normal, seven-nucleotide tRNA

anticodon loop. The two conformations differ only in which five nucleotides (numbers one to five or three to seven) are stacked along one side of the loop. Woese appreciated (72) that if tRNAs always arrived at the decoding site on the ribosome in one conformation and then flipped conformations after peptide bond formation, the mRNA would be systematically and precisely displaced by three nucleotides in each cycle (Fig. 1B).

Woese's model has an inherent attractiveness. Translocation is reduced to a physical property of the tRNA without need for protein factors in the form of measuring sticks or arm's reaches. A tRNA-centric view of translocation is also compatible with the belief that the earliest protein synthesizing systems consisted solely of RNA moieties. But despite all of its beauty, the reciprocating ratchet model has no experimental support. Estimates for the A- and P-site locations of tRNAs (74) are not consistent with Woese's view. Curran and Yarus (75) have also argued against the reciprocating ratchet based on the lack of an increase in translocation errors made by mutant tRNAs with subtle changes in their anticodon stems and loops.

In fact, it has taken Weiss's model of frameshifting by mispaired tRNAs to reinvigorate the reciprocating ratchet. Unfortunately, this model has also failed its only experimental test. According to Weiss (71), a tRNA bound to a codon via U33 grapple pairing would undergo one of two types of anticodon loop isomerizations (Fig. 1B). Depending on the favorability of stacking interactions, the isomerizations would cause two or four message nucleotides to be translocated, and would result in a -1 or +1 shift in the reading frame. The model predicts

that the U33 position is critical in establishing the initial offset anticodon-codon arrangement. Bruce *et al.* (76) have mutated the U residue at position 33 in one culprit tRNA. They find that -1 frameshifts (resulting from two-nucleotide translocations) do not require a base pair between the mRNA and position 33 of the tRNA.

It would be much more interesting to test Weiss's model with a culprit tRNA that causes +1 frameshifts. These events are predicted to require U33 pairing not only during initial recognition but throughout the proposed isomerization from one anticodon loop conformation to the other (Fig. 1B). In the two-nucleotide translocation suggested for a -1 frameshift, the U33 base pair is broken during isomerization (Fig. 1B).

The +1- frameshift-causing culprit tRNAs are also more intriguing for lack of reasonable alternative explanations of their action. The fixed-translocation type models would adequately explain the -1 frameshifts caused by culprit tRNAs. After assuming the same out-of-phase pairing proposed by Weiss (71; Fig. 1B), one needs only postulate that when the tRNA is brought to the fixed point in the P site, only the two message nucleotides complexed with the anticodon positions 34 and 35 are brought with it. The same type of argument applied to the +1 frameshift-inducing culprits, however, would require that they be paired out of phase by one nucleotide in the +1 direction. In this conformation, four message nucleotides would be predicted to travel with the tRNA to the P site. However, the nucleotide sequence of the anticodons of +1-frameshift-inducing culprit tRNAs and the sequence of the sites at which they act are not consistent with pairing by these tRNAs in the +1 frame.

Nucleotide context. Frameshifting mediated by mutant and normal tRNAs can be influenced by the nucleotide sequence neighboring the codon at which the frameshift occurrs (50,53,54,69,70) So-called context effects are not unique to frameshift events. Both nonsense suppression and missense suppression are also affected by nucleotide context. For nonsense suppression, the nucleotide or nucleotides immediately 3′ to the suppressed stop codon are most important (77-79), while in the few cases where context effects have been observed for missense suppression, the 5′ nucleotides are implicated (80). The nature of the context effect in the examples of frameshifting discussed thus far is unclear, save for the obvious importance of the 3′ neighboring nucleotide for certain +1 frameshift suppressor tRNAs (48,49,52,55,56).

A potential nucleotide context effect on a grand scale has recently been suggested by Trifonov (81). By comparing the codon composition of several genes from many organisms, Trifonov observed a periodicity of the sequence GNX (where N can be A, C, or U and X can be any nucleotide) in the 0 frame of coding sequences. This periodicity is not present in the alternative reading frames or in other parts of the gene. Moreover, it breaks down upstream of several known frameshift sites. Trifonov (81) argues that this periodicity may indicate a mechanism by which the ribosome, presumably through the rRNAs, could keep track of the proper reading frame.

Slippery stops. The most dramatic example of a context effect on frameshifting frequency was recently discovered by Weiss and his colleagues (82). These workers studied the level of frameshifting on hundreds of synthetic constructs in *E. coli*. All of these "frameshift

windows" are inserted upstream of a *lac*Z gene, and frameshifting frequency deduced from the level of B-galactosidase activity compared to an in-frame *lac*Z construct. N-terminal amino acid sequence analysis of the full-length B-galactosidase molecules indicates the site and nature of the frameshift event.

This tour de force study has revealed three facts about frameshifting. First, nucleotide runs--short homopolymeric sequences--are sites of frequent frameshifting. These runs of nucleotides allow tRNAs to re-align with the message in an alternative reading frame (82). Re-alignments in the -2, -1, +1, and +2 reading frames were observed with efficiencies approaching 0.1% (i.e., one in every thousand translating ribosomes entered the alternative reading frame). Along with these tRNA "slips", tRNA "hops" were also detected. A hop differs from a slip by having non-base paired nucleotides between the pre- and post-frameshift tRNA binding sites (82).

The frequency of tRNA slippage along homopolymeric runs is greatly enhanced by the presence of a 3' neighboring stop codon (82). Addition of a stop codon adjacent to a nucleotide run can increase frameshifting frequency twenty fold to approximately 2%. Precisely why terminators augment frameshifting is not known, but they may act to pause translation in order for the tRNA in the P site to re-align with the mRNA. The suggestion that increased time increases the error frequency runs counter to the aforementioned concept that translational accuracy is limited by the need for efficient translation. Yarus and Thompson (63) have pointed out, however, that an error mechanism requiring a kinetically "slow step" would be enhanced by a general slowing of the

translation rate.

Frameshifting by the mechanism of tRNA slippage has been suggested previously. Fox and Weiss-Brummer (83) have proposed slippage along runs of U residues to account for the unusual leakiness of +1 and -1 frameshift mutations in the yeast mitochondrial gene oxy-1. Slippage along U residues has also been implicated in a leaky -1 frameshift allele of the *Salmonella trp*E gene (84).

Frameshifting by tRNA slippage, especially if it occurs in the P site (as suggested by Weiss *et al.* (82) for slippery stops in *E. coli*), is uninformative with regard to the mechanism of translocation. This mechanism is relevant to normal frame maintenance, however. If homopolymeric runs are problematic for frame maintenance, they should be underrepresented in mRNAs. Given the complicity of stop codons in slippage events, the sequence upstream of a terminator would be expected to be devoid of possible alternative base pairing opportunities. Finally, the claim that stop codons make runs more slippery by allowing increased time for tRNAs to sample the alternative reading frames brings with it the notion that translation speed might be selected for in order to avoid reading frame ambiguity.

Physiological frameshifts. Up to this point I have reviewed frameshifting as it occurs on mutant genes or synthetic constructs. This type of study can be informative as to the mechanisms that prevent these errors. I will end with a discussion of frameshift events that do more than compensate for an earlier frameshift mutation or inform us about normal translocation. These frameshift events are required for the expression of wild-type genetic information.

Two physiological frameshift events were known before the investigation of frameshifting in retroviral gene expression. One occurs near the end of gene *10* in bacteriophage T7 (85). The nucleotide sequence of T7 by Dunn and Studier showed an extended open reading frame overlapping the 3' end of gene *10*. The knowledge that T7 virions contain two versions of the gene *10* product led these workers to analyze gene *10* RNA transcripts by *in vitro* translation (85). Along with the expected gene *10* product, this translation yielded a less abundant, slightly larger product presumed to result from frameshifting near the end of gene *10*. The ratio of the normal sized and extended gene *10* proteins indicates that the efficiency of this frameshift is approximately 5% (85). As with several other frameshift events discussed above, tRNA slippage along runs of U residues is the suspected mechanism of frameshifting. No protein sequence or mutational data is available to confirm this prediction.

The second ribosomal frameshift required for wild-type gene expression occurs in the gene encoding the *E. coli* release factor II (RFII), one of two translation factors in *E.coli* that recognizes stop codons and initiates release of the nascent peptide from the ribosome. This frameshift is the most efficient yet studied, with nearly one in two ribosomes changing frame under certain conditions. The necessity for frameshifting was reported by Craigen and co-workers (86) through comparison of the nucleotide sequence of the *RFII* gene and the N-terminal amino acid sequence of the protein. The first twenty-five amino acids of RFII are encoded in the 0 frame beginning with the initiator methionine, but the next 0-frame codon is a UGA terminator,

and the coding frame switches by one nucleotide in the 3′ direction at this point.

Craigen *et al.* (86) noted that UGA is one of the stop codons recognized by RFII and hypothesized that the necessity for shifting into the +1 reading frame may be a form of autoregulation for *RFII*. After translating the first 25 codons on the *RFII* RNA, ribosomes would encounter the UGA terminator. If cellular RFII concentrations were low, the UGA codon would remain in the A site unrecognized. The corresponding delay might then all the ribosome to shift into the +1 frame and continue synthesis of RFII. With abundant RFII, the UGA codon would be efficiently recognized and translation of the *RFII* mRNA terminated. Craigen and Caskey (87) proved this hypothesis by fusing the 5′ region of the RFII gene to *lac*Z and performing *in vitro* translation in the presence of different concentrations of purified RFII. They found that the frameshifting efficiency (measured by B-galactosidase activity) was inversely proportional to the to the RFII concentration.

Weiss *et al.* (88) have analyzed the RFII frameshift event in detail. They have learned that it combines some familiar features with one totally unexpected one. The RFII frameshift is a slippery stop. $tRNA^{Leu}$ slips from the 0-frame CUU codon that precedes the UGA terminator to the +1-frame UUU codon. Replacement of this slippery sequence with a different one maintains high-level frameshifting. The importance of the 3′ neighboring terminator codon is confirmed by the ten-fold drop in frameshifting frequency when the UGA is converted to UGG.

To their surprise, Weiss *et al.* (88) discovered an important sequence located four to nine nucleotides 5′ to the frameshift site. This sequence bears striking resemblance to the Shine-Dalgarno (S-D) sequence recognized by the 3′ end of 16S mRNA during translational initiation in prokaryotes (89). Any of several mutations in the S-D-like sequence in *RFII* mRNA greatly inhibits frameshifting efficiency (88). Also, a nucleotide insertion between the S-D-like sequence and frameshift site reduces frameshifting by greater than ten fold. Weiss *et al.* (88) confirmed that this S-D-like sequence was interacting with 16S rRNA to promote frameshifting by suppressing the effects of a message mutation in the S-D-like sequence with a complementary mutation in the relevant region of 16S rRNA. The 16S rRNA mutation by itself also lowers frameshifting efficiency on wild-type *RFII* mRNA. It is not known whether the interaction between the mRNA and rRNA helps to force ribosomes into the +1 frame or is just another means to slow translation in order to let the $tRNA^{Leu}$ slip naturally. It is interesting that this rRNA sequence is involved in any capacity, however, since it has long been believed to function only in initiation.

Summary. Throughout this section, I have referred to various models of frameshifting and their implications for normal frame maintenance. Despite the importance of this process and the numerous attempts to understand it, we actually know very little. For example, regarding the central issue of whether translocation is primarily a function of the tRNA-mRNA interaction or the ribosome, we have come full circle but are not yet on solid ground. The fixed translocation model of Watson (58) has recently regained prowess thanks to the efforts of Curran and Yarus

(57,75). Both sets of experiments by these authors argue against the notion that the tRNA, either through contortions of its anticodon loop or through the number of base pairs between it and the mRNA, determines the step-size of translocation. But while these authors deserve credit for conducting the only systematic analyses of translocation, their conclusions should be taken with some caution. All of the experiments of Curran and Yarus concentrate on translocation by nonsense suppressor tRNAs, which are not fully efficient at decoding and are continuously in competition with release factors at their sites of action. Also, the level of frameshifting observed in these studies is quite low; one worries that this may be an effect of the particular tRNA under study and not generally applicable. Before abandoning the tRNA-based models of translocation (such as the reciprocating ratchet (72)), it would be necessary to carry out similar experiments on normal tRNAs.

The model of frameshifting proposed by Weiss (71) is still provocative and deserves further experimentation. If substantiated, it would return translocation back into the realm of the tRNA.

Regarding physiological frameshifts, the message appears to be that tRNA slippage is the most efficient way to change reading frame. (This message will be reiterated in our studies of frameshifting in retroviruses.) Whether these systems will inform us about the normal mechanisms of frame maintenance must await characterization of the relevant tRNA species, the sites at which they act, and where on the ribosome these tRNAs slip on the mRNA.

1. Baltimore, D. (1970) *Nature* 226, 1209-1211.

2. Temin, H.M. and Mizutani, S. (1970) *Nature* 266, 1211-1218.

3. Stehelin,D., Varmus, H.E., Bishop, J.M. and Vogt, P.K. (1976) *Nature* 260, 170-173.

4. Joho, R.H., Stoll, E., Friis, R.R., Billeter, M.A. and Weissmann, C. (1976). In *Animal Virology* (ed. D. Baltimore et al.), vol. 4, pp. 127-145. Academic Press, New York.

5. Wang, L.-H, et al. (1976). *Proc. Natl. Acad. Sci.* USA 73, 3952-3956.

6. Aaronson, S. A. and Barbacid, M. (1980) *J. Exp. Med.* 151, 467-480.

7. Hayward, W.S. (1977) *J. Virol.* 24, 47-63.

8. Weiss, S.R., Varmus, H.E. and Bishop, J.M. (1977) *Cell* 12, 983-992.

9. von der Helm, K. and Duesberg, P.H. (1975) *Proc. Natl. Acad. Sci.* USA 72, 614-618.

10. Pawson, T., Martin, G.S. and Smith, A.E. (1976) *J. Virol.* 19, 950-967.

11. Pawson, T., Harvey, G.S. and Smith. A.E. (1977) *Nature* 268, 416-420.

12. Paterson, B.M., Marciani, D.J. and Papas, T.S. (1977) *Proc. Natl. Acad. Sci.* USA 24, 4951-4954.

13. Purchio, A.F., Erikson, E. and Erikson, R.L. (1977) *Proc. Natl. Acad. Sci.* USA 74, 4661-4655.

14. Kerr, I.M., Olshevsky, U., Lodish, H.F. and Baltimore, D. (1976) *J. Virol.* 18, 627-635.

15. Murphy, E.C., Kopchick, J.J., Watson, K.F., Arlingshause, R.B. (1978) *Cell* 13, 359-369.

16. Kozak, M. (1978) *Cell* 15, 1109-1123.

17. Stacey, D.W., Allfrey, V.G. and Hanafusa, H. (1977) *Proc. Natl. Acad Sci.* USA 74, 1614-1618.

18. Murphy, E.C., Campos, D. and Arlinghaus, R.B. (1979) *Virology* 93, 293-302.

19. Oppermann H., Bishop, J.M., Varmus, H.E. and Levintow, L. (1977) *Cell* 12, 993-1005.

20. Jamjoon, G.A., Naso, R.B. and Arlinghaus, R.B. (1977) *Virology* 78, 11-34.

21. Rettenmier, C.W., Karess, R.E., Anderson, S.M. and Hanafusa, H. (1979) *J. Virol.* 32, 102-113.

22. Weiner, A.M. and Weber, K. (1971) *Nature New Biol.* 234, 206-209.

23 Phillipson, L., Anderson, P., Olshevsky, U., Weinberg, R. and Baltimore, D. (1978) *Cell* 13, 189-199.

24. Shinnick, T.M., Lerner, R.A. and Sutcliffe, J.G. (1981) *Nature* 293, 543-548.

25. Weiss, S.R. et al. (1978) *Cell* 15, 607-614.

26. Schwartz, D.E., Tizard, R. and Gilbert, W. (1983) *Cell* 32, 853-869.

27. *RNA Tumor Viruses*, second edition,(Weiss, R., Teich, N., Varmus, H. and Coffin, J., eds.), Cold Spring Harbor Laboratory, Cold Spring Harbor, 1984.

28. Seiki, M. Hattori, S., Hirayama, Y. and Yoshima, M. (1983) *Proc. Natl. Acad. Sci.* USA 80, 3618-3622.

29. Hiramatsu, K., Nishida, J., Naito, A. and Yoshikura, H. (1987) *J. Gen. Virol.* 68, 213-218.

30. Yoshinaka, Y., Katoh, I., Copeland, T.D. and Oroszlan, S. (1985)

Proc. Natl. Acad. Sci. USA 82, 1618-1622.

31. Levin, J.G., Hu, S.C., Rein, A., Messer, L.I. and Gerwin, B. (1984) *J. Virol.* 513, 470-478.

32. Clare, J. and Farabaugh, P. (1985) *Proc. Natl. Acad. Sci.* USA 82, 2829-2833.

33. Mellor, J. et al. (1985), *Nature* 313, 243-246.

34. Saigo, K. et al. (1984), *Nature* 312, 659-661.

35. Marlor, R.L., Parkhurst, S.M. and Corces, V.G. (1986) *Mol. Cell. Biol.* 6, 1129-1134.

36. Shimotohno, K. et al. (1985) *Proc. Natl. Acad. Sci.* USA 82, 3101-3105.

37. Sagata, N. et al. (1985) *Proc. Natl. Acad. Sci.* USA 82, 677-681.

38. Rice, N.R., Stephens, R.M., Burny, A. and Gilden, R.V. (1985) *Virology* 142, 357-377.

39. Wain-Hobsen, S., Sonigo, P., Danos, O., Cole, S. and Alizon, M. (1985) *Cell* 40, 9-17.

40. Ratner, L. et al. (1985) *Nature* 313, 277-284.

41. Sanchez-Pescador, R. et al. (1985) *Science* 227, 484-492.

42. Borer, P.N., Dengler, B., Tinoco, I. and Uhlenbeck, O. (1974) *J. Mol. Biol.* 86, 843-848.

43. Crick, F.H.C., Leslie-Barnett, F.R.S., Brenner, S. and Watts-Tobin, R.J.(1961) *Nature* 192, 1227-1232.

44. Atkins, J.F., Elseviers, D. and Gorini, L. (1972) *Proc. Natl. Acad. Sci.* USA 69, 1192-1195.

45. Kurland, C.G. (1979). In *Nonsense Mutations and tRNA suppressors* (Celis, J.E. and Smith, J.D., eds.), pp. 98-108. Academic Press,

New York.

46. Riddle, D.L. and Roth, J.R. (1970) *J. Mol. Biol.* 54, 131-144.

47. Riddle, D. and Carbon, J. (1973) *Nature New Biol.* 242, 230-234.

48. Roth, J.R. (1974) *Ann. Rev. Genet.* 8, 319-346.

49. Kohno, T. and Roth, J.R. (1978) *J. Mol. Biol.* 126, 37-52.

50. Bossi, L. and Roth, J.R. (1981) *Cell* 25, 489-496.

51. Bossi, L. and Smith, D.M. (1984) *Proc. Natl. Acad. Sci.* USA 81, 6105- 6109.

52. Culbertson, M.R., Charnas, L., Johnson, M.T. and Fink, G.R. (1977) *Genetics* 86, 745-764.

53. Gaber, R.F. and Calbertson, M.R. (1984) *Mol. Cell. Biol.* 4, 2052-2061.

54. Cummins, C.M., Culbertson, M.R. and Knapp, G. (1985) *Mol. Cell. Biol.* 5, 1760-1771.

55. Yourno, J. and Kohno, T. (1972) *Science* 174, 650-652.

56. Yourno, J. (1972) *Nature New Biol.* 239, 219-221.

57. Curran, J.F. and Yarus, M. (1987) *Science* 238, 1545-1550.

58. Watson, J.D. (1964) *Bull. Soc. Chim. Biol.* 46, 1399-1425.

59. O'Mahony, D.J., Mims, B.H., Thompson, S., Murgola, E.J. and Atkins, J.F. (1988) submitted.

60. Culbertson, M.R., Gabar, R.F. and Cummins, C.M. (1982) *Genetics* 102, 361-378.

61. Stingini,P. and Gorini, L. (1970) *J. Mol. Biol.* 47, 517-530.

62. Rosset, R. and Gorini, L. (1969) *J. Mol. Biol.* 39, 95-112.

63. Yarus, M. and Thompson, R.C. (1983). In *Gene Function in Prokaryotes* (Beckwith, J.R., Davies, J.E. and Gallant, J.A., eds.)

pp. 23-63. Cold Spring Harbor Laboratory, Cold Spring Harbor.

64. Hopfield, J.J. (1974) *Proc. Natl. Acad. Sci.* USA 71, 4135-4139.

65. Ninio, J. (1975) *Biochemie* 57, 587-595.

66. Thompson, R.C. (1988) *Trends Biochem. Sci.*, in press.

67. Thompson, R.C. and Karim, A.M. (1982) *Proc. Natl. Acad. Sci.* USA 79, 4922-4926.

68. Kurland, C.G. and Gallant, J.A. (1986). In *Accuracy in Molecular Processes* (Galas, D., ed.), pp. 127-157. Marcel Decker, New York.

69. Weiss, R. and Gallant, J. (1983) *Nature* 302, 389-393.

70. Atkins, J.F., Gesteland, R.F., Ried, B.R. and Anderson, C.W. (1979) *Cell* 18, 1119-1131.

71. Weiss, R.B. (1984) *Proc. Natl. Acad. Sci.* USA 81, 5787-5801.

72. Woese, C. (1970) *Nature* 226, 817-820.

73. Fuller, W. and Hodgson, A. (1967) *Nature* 215, 817-820.

74. Johnson, A.E., Adkins, H.J., Matthews, E.A. and Cantor, C.R. (1982) *J. Mol. Biol.* 156, 113-140.

75. Curran, J. and Yarus, M. (1986) *Proc. Natl. Acad. Sci.* USA 83, 6538-6542.

76. Bruce, A., Atkins, J.F. and Gesteland, R.F. (1986) *ibid*, pp. 5062-5066.

77. Bossi, L. and Roth, J.R. (1980) *Nature* 286, 123-127.

78. Bossi, L. (1983) *J. Mol. Biol.* 164, 73-87.

79. Miller, J.H. and Albertini, A.M. (1983) *ibid*, pp. 59-71.

80. Murgola, E.J., Pagel, F.T. and Hijazi, K.A. (1984) *J. Mol. Biol.* 175, 19-28.

81. Trifonov, E.N. (1987) *J. Mol. Biol.* 194, 643-652.

82. Weiss, R.B., Dunn, D.M., Atkins, J.F. and Gesteland, R.F. (1988) *Cold Spring Harbor Symp. Quant. Biol.*, in press.

83. Fox, T. and Weiss-Brummer, B. (1980) *Nature* 288, 60-63.

84. Atkins, J.F., Nichols, B.P. and Thompson, S. (1983) *EMBO J.* 2, 1345-1350.

85. Dunn, J.J. and Studier, F.W. (1983) *J. Mol. Biol.* 166, 477-535.

86. Craigen, W.J., Cook, R.G., Tate, W.P. and Caskey, C.T. (1985) *Proc. Natl. Acad. Sci.* USA 82, 3616-3620.

87. Craigen, W.J. and Caskey, C.T. (1986) *Nature* 322, 273-275.

88. Weiss, R.B., Dunn, D.M., Dahlberg, A.E., Atkins, J.F. and Gesteland, R.F. (1988) *EMBO J.*, in press.

89. Shine, J. and Dalgarno, L. (1974) *Proc. Natl. Acad. Sci.* USA 72, 4734-4738.

Figure 1. Models of Frameshifting.

A. The two alternative conformations of 8-nucleotide anticodon loops of +1 frameshift suppressor tRNAs proposed by Bossi and Smith (51) are shown relative to wild-type tRNA with a 7-nucleotide anticodon loop. The wild-type tRNA is shown in the 2-5 conformation that has been observed in tRNA crystals (a). According to Bossi and Smith (51) insertion of a nucleotide into the anticodon loop results in adoption of either the 3-5 (b) or 2-6 (c) conformations shown. (Figure is taken from ref. 51.)

B. Normal and aberrant ratchets. (a) The anticodon loop isomerization proposed by Woese (72) for wild-type tRNA is shown with the tRNA in the so-called hf (5-2) conformation and paired to the codon (a,b,c) moving into the FH (2-5) conformation, bringing with it three message nucleotides. (b) The U33 grapple pairing of Weiss (71) distorts the anticodon loop into a 4-3 hf conformation in which the codon is paired by tRNA nucleotides 33, 34, and 35 (instead of 34, 35, 36). Depending on the strength of base stacking interactions (see ref. 71), the isomerization to the FH conformation will result in either 2 or 4 message nucleotides being translocated. (Figure taken from ref. 71.)

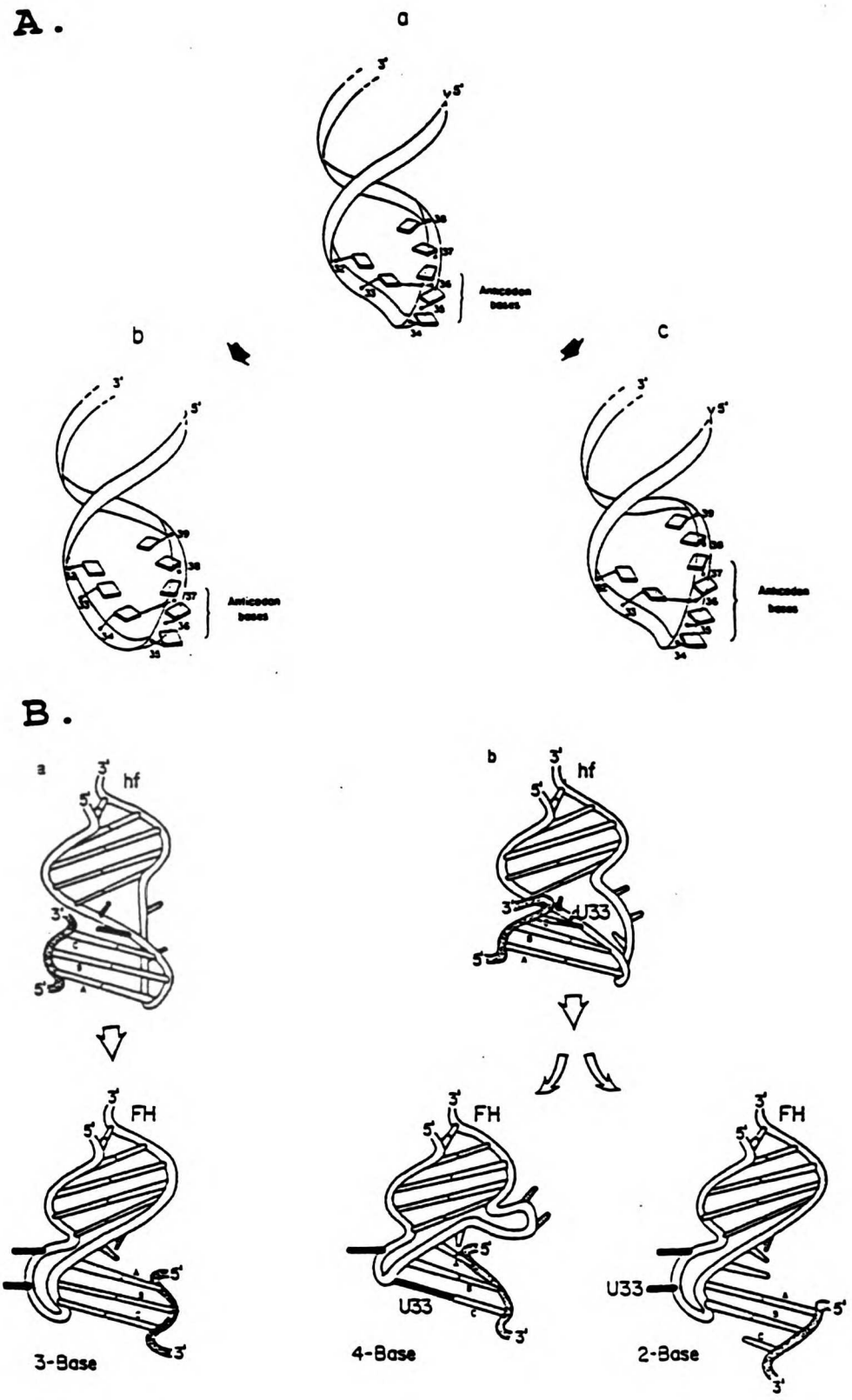


Figure 1.

Chapter 1

Expression of the Rous Sarcoma Virus *pol* Gene by Ribosomal Frameshifting

RESEARCH ARTICLE

Expression of the Rous Sarcoma Virus *pol* Gene by Ribosomal Frameshifting

Tyler Jacks and Harold E. Varmus

Abstract. *The* pol *gene of Rous sarcoma virus is positioned downstream of the* gag *gene in a different, briefly overlapping reading frame; nevertheless, the primary translation product of* pol *is a* gag-pol *fusion protein. Two mechanisms, ribosomal frameshifting and RNA splicing, have been considered to explain this phenomenon. The frameshifting model is supported by synthesis of both* gag *protein and* gag-pol *fusion protein in a cell-free mammalian translation system programmed by a single RNA species that was synthesized from cloned viral DNA with a bacteriophage RNA polymerase. Under these conditions, the ratio of the* gag *protein to the fusion protein (about 20 to 1) is similar to that previously observed in infected cells, the frameshifting is specific for the* gag-pol *junction, and it is unaffected by large deletions in* gag. *In addition, synthesis of the fusion protein is ten times less efficient in an* Escherichia coli *cell-free translation system and cannot be explained by transcriptional errors or in vitro modification of the RNA. Ribosomal frameshifting may affect production of other proteins in higher eukaryotes, including proteins encoded by several retroviruses and transposable elements.*

Eukaryotic viruses have several strategies to generate multiple protein products from their genomes, including genome segmentation, polycistronic messenger RNA's (mRNA's), proteolytic cleavage of polyproteins, production of multiple mRNA's from a single template by differential splicing or polyadenylation, and termination suppression (*1*, *2*). Expression of the *pol* gene of several retroviruses is known to require synthesis of a polyprotein composed of regions encoded both by *pol* and by the preceding coding domain, *gag*, and later processing by proteolytic cleavage. The *gag-pol* polyprotein is produced, moreover, about 5 percent as efficiently as the translation product of *gag* alone (*3*).

The difficulties of occasionally circumventing the *gag* termination signal to make the *gag-pol* polyprotein were brought into focus by nucleotide sequencing of the genome of Rous sarcoma virus (RSV): the reading frames for *gag* and *pol* were found to differ, with *pol* in the −1 frame with respect to *gag*, and the open frames overlapping for 58 nucleotides before the *gag* amber termination codon (*4*) (Fig. 1A). Under these circumstances, synthesis of the *gag-pol* fusion protein could be achieved by either of two mechanisms: a low frequency RNA splicing event (despite the absence of conventional splice donor and acceptor sites) or occasional ribosomal frameshifting during translation, a phenomenon without precedent in higher eukaryotes (Fig. 1, B and C).

We now present strong evidence in favor of the frameshifting model, based on the synthesis of *gag* and *gag-pol* proteins in a cell-free rabbit reticulocyte translation system, when programmed with RNA transcribed in vitro from cloned RSV DNA by the bacteriophage SP6 RNA polymerase.

The rationale for our experimental approach was twofold: (i) Yoshinaka *et al.*, (*2*) demonstrated that occasional translational suppression of the amber stop codon separating the *gag* and *pol* genes of Moloney murine leukemia virus (Mo-MLV) is responsible for synthesis of the *gag-pol* polyprotein. Since retroviruses commonly use similar strategies for replication and gene expression (*3*), these results suggested that suppression of translational termination by ribosomal frameshifting might be used for those viruses whose *gag* and *pol* genes are out of frame. (ii) Several years ago it was reported (*5*) that RNA extracted from particles of RSV could direct the synthesis of both *gag* and *gag-pol* proteins in appropriate proportions in rabbit reticulocyte lysates (RRL). These experiments were not decisive with respect to the mechanism for generating the *gag-pol* protein, because they did not exclude the possibility that a low abundance, spliced *gag-pol* mRNA was adventitiously packaged into virions. However, if the frameshifting mechanism were correct and the spliced mRNA did not exist, the result implies that unspliced RSV RNA can be successfully translated to generate *gag-pol* polyprotein in reticulocyte lysates.

An in vitro test for frameshifting. To synthesize RNA in which the RSV *gag* and *pol* coding regions are in their genomic (out-of-frame) configuration, a portion of cloned RSV DNA, extending from the Sac I site 125 base pairs (bp) upstream from the *gag* initiation site to a Xba I site within *pol*, 895 bp downstream from the *gag* termination codon, was recloned in the plasmid pSP65 (*6*), in a polylinker downstream from the bacteriophage SP6 promoter (Fig. 2A). When this plasmid is linearized at the Xba I site and added to a reaction mixture containing SP6 RNA polymerase and ribonucleotide triphosphates, abundant amounts of *gag-pol* RNA (GP RNA) are synthesized. Translation of GP RNA in lysates of rabbit reticulocytes is expected to produce mainly the 76-kilodalton (kD) product of the *gag* gene, $Pr76^{gag}$. If some fraction of the ribosomes is able to shift into the −1 frame within the 58-nucleotide overlap of the *gag* and *pol* frames, a 108-kD *gag-pol* fusion protein will also be made.

Typical [^{35}S]methionine-labeled translation products of GP RNA are shown in Fig. 2B after fractionation by electrophoresis in sodium dodecyl sulfate (SDS)-polyacrylamide gels. Even without prior immunoprecipitation, there are clearly discernible proteins of the predicted sizes for *gag* and *gag-pol* products at the expected ratio of approximately 20:1 (Fig. 2B, lane 1). Both the 76-kD and 108-kD proteins are immunoprecipitable with serum specific for RSV $p19^{gag}$ (lane 2), but only the larger species is recognized by antiserum to reverse transcriptase, a product of *pol* (lane 3). The relatedness of the 76-kD and 108-kD proteins was further demonstrated by partial digestion with staphylococcal V8 protease. Some of the additional minor reaction products (Fig. 2B) can also be immunoprecipitated with antiserum to $p19^{gag}$ and presumably are proteins initiated at methionine codons internal to the *gag* region; for simplicity of

T. Jacks is a graduate student in the Department of Biochemistry and Biophysics and H. E. Varmus is a professor in the Departments of Microbiology and Immunology and Biochemistry and Biophysics at the University of California, San Francisco 94143.

presentation, these are not considered further, although at least one of these proteins is also precipitable by antiserum to reverse transcriptase and it presumably results from a frameshifting event similar to that which produces the 108-kD protein.

The results presented in Fig. 2B indicate that RNA containing *gag* and *pol* in different translational reading frames can direct synthesis of both *gag* and *gag-pol* proteins in vitro. Moreover, the efficiency of synthesis of the 108-kD *gag-pol* protein (about 5 percent of the level of $Pr76^{gag}$) is consistent with ratios of *gag* and *gag-pol* proteins observed in RSV-infected cells (*7*).

To begin to define the portion of GP RNA responsible for ribosomal frameshifting at 5 percent efficiency, we made the two deletion mutations of pGP shown in Fig. 3A. In GP-ΔB, we removed 463 codons of *gag* by deleting an internal Bam HI fragment from pGP. The construction maintains 51 codons at the beginning of *gag* and 189 codons at its end, so that GP-ΔB RNA can encode a 26-kD *gag* protein. If the deleted *gag* sequences are not required for frameshifting, then translation of the GP-ΔB RNA should also yield a 58-kD *gag-pol* fusion protein. Production of the predicted 58-kD *gag-pol* protein, as well as $p26^{gag}$, does occur in the absence of most of the *gag* coding domain (Fig. 3B, lane 1). The identity of the *gag-pol* protein was again confirmed by immunoprecipitation with antiserum to reverse transcriptase.

We were also able to remove a substantial portion of the 58-nt overlap region and still observe frameshifting. The plasmid pGP-ΔA was derived from pGP by deleting the 24-nt between the two Avr II sites in the overlap region (Fig. 3A). In GP-ΔA RNA, the window for frameshifting has been closed down to 34 nt. In vitro translation of this RNA produces an abundant protein that is slightly smaller than the *gag* protein of GP RNA, as was expected from the size of the deletion mutation (Fig. 3B, lanes 2 and 3). There is no apparent difference in the amounts of *gag-pol* fusion proteins produced, implying that the remaining 34 nt of the overlapping reading frames suffice to mediate frameshifting.

Tests for RNA homogeneity. Our conclusion that the *gag-pol* fusion protein is produced via ribosomal frameshifting relies on the assumption that the RNA synthesized and translated in vitro constitutes a homogeneous population in which *gag* and *pol* are out of frame. We have examined this assumption by addressing the unlikely possibility that a minor class of RNA, in which *gag* and *pol* occupy the same reading frame, has been generated in our experiments. This could conceivably occur either via transcriptional frameshifting by SP6 polymerase or via an unconventional type of mRNA splicing (either autogenous or induced by the in vitro conditions).

Since correction of frame could occur by the addition of one nucleotide or the deletion of two nucleotides, changes that would probably not be detectable by nuclease protection analysis, we have depended instead on a functional test for the in-frame subclass of RNA, using a cell-free translation system derived from *Escherichia coli*. To promote the efficiency of this test, we first modified the *gag-pol* DNA by the addition of a bacterial translational initiation site derived from the *E. coli* lipoprotein gene (*lpp*) (*8*); in addition, we took advantage of the proven dispensability of most of the *gag* domain (Fig. 3) and removed all of the *gag* sequences upstream of the Eco RI site preceding the *gag* terminator. The resulting construction, pLGP, and a similar plasmid, pLGP-IF, in which *gag* and *pol* are aligned in frame (Fig. 4A), were transcribed by the SP6 RNA polymerase to yield LGP and LGP-IF RNA's, which were subsequently translated in an *E. coli*–derived translation system (legend to Fig. 4).

Translation in the *E. coli* system could be directly compared to translation in the RRL because LGP RNA functions as an efficient mRNA in both systems. Moreover, the apparent efficiency of frameshifting in the RRL (approximately 15 percent) is significantly greater than observed previously. (This probably reflects a selective loss of the 10-kD *gag* protein from the polyacrylamide gel during preparation for autoradiography and not an actual increase in the level of frameshifting.) When LGP RNA is translated in both systems, similar amounts of the 10-kD *gag* protein are made, but the RRL generates at least ten times more of the 41-kD *gag-pol* protein (Fig. 4B, lanes 1 and 2). The

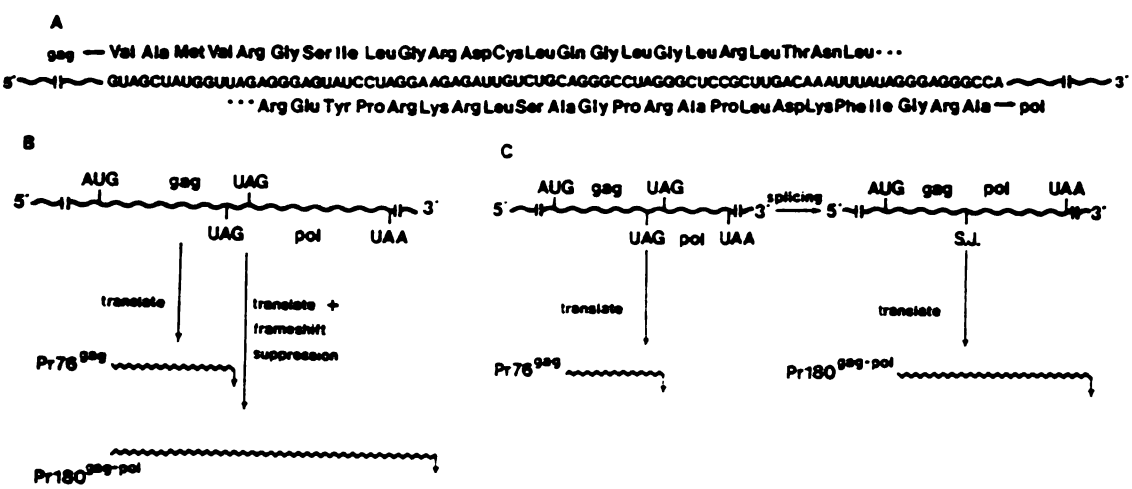


Fig. 1. Models for synthesis of the RSV proteins $Pr76^{gag}$ and $Pr180^{gag-pol}$. (A) Diagram of PR-C RSV RNA emphasizing the nucleotides shared by the *gag* and *pol* open reading frames and showing the amino acids encoded by the 3' end of the *gag* frame and the 5' end of the *pol* frame (*4*). The relevant stop codons are indicated by "···". (B) The translational suppression (one-mRNA) hypothesis. In this model, a single mRNA species directs synthesis of both $Pr76^{gag}$ (following normal translation) and $Pr180^{gag-pol}$ (after translational suppression of the frame difference between the *gag* and *pol* genes). (C) The splicing (two-mRNA) hypothesis. Here, infrequent splicing of the genome-length message gives rise to a second class of mRNA with a splice junction (S.J.) in which the *gag* and *pol* genes are in frame. Normal translation of these two mRNA's yields the desired proteins. (The distance between the termination codons has been exaggerated in parts B and C to emphasize the overlap region.)

reduced amount of the 41-kD fusion protein is not due to instability of the protein in the *E. coli* lysate, since translation of LGP-IF RNA yields the expected quantity of an almost identical protein (Fig. 4B, lane 3).

These results indicate that alignment of the *gag* and *pol* frames during transcription by SP6 RNA polymerase or by autocatalytic splicing cannot explain results in the RRL, since the in-frame RNA would direct equally efficient synthesis of the fusion protein in either translation system. Furthermore, the prokaryotic system appears to respond poorly to the presumptive eukaryotic signals for frameshifting at the *gag-pol* boundary.

The experiments shown in Fig. 4 do not exclude the remote possibility that synthesis of the fusion protein reflects modification (such as splicing) of the GP or LGP RNA on exposure to the RRL. However, when we retranslated RNA recovered from a RRL translation reaction in both *E. coli* system and the RRL, we observed results qualitatively similar to those obtained with newly synthesized RNA (*9*). Thus we conclude that ribosomal frameshifting during translation, rather than transcriptional frameshifting or RNA processing, is responsible for the synthesis of *gag-pol* fusion proteins in our experiments.

Sequence specificity of the frameshifting mechanism. It seems probable that ribosomal frameshifting in the RRL occurs in response to signals in the *gag-pol* mRNA; this would explain the differential efficiency with which the fusion proteins are made in *E. coli* and mammalian systems and would prevent promiscuous frameshifting during translation in eukaryotic cells. To test this notion more explicitly and to exclude the possibility that

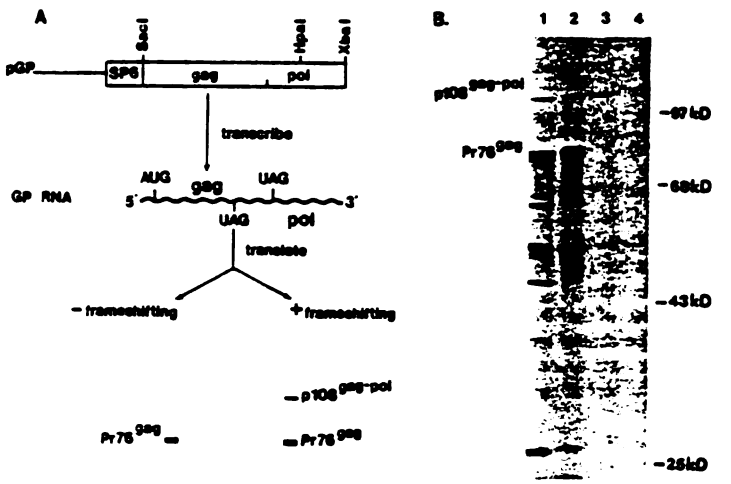


Fig. 2. An in vitro test of the translational suppression hypothesis. (A) The plasmid pGP was constructed by cloning a 3.1-kb fragment from the RSV *gag-pol* region into the SP6 vector SP65 (*6*). The RSV insert is composed of a 2.5-kb Sac I to Hpa I fragment, including the *gag-pol* overlap region, from cloned DNA of strain PR-C (*4*) and a 0.6 kb-Hpa I to Xba I fragment from strain SR-A (*31*). After linearization with Xba I, 2.5 μg of DNA were transcribed with 8 units of SP6 RNA polymerase (Promega) in a 25-μl reaction mixture (*6*). The resulting GP RNA (500 ng) was translated in a 50-μl RRL translation reaction (Promega) supplemented with [^{35}S]methionine (65 μCi;1200 Ci/mmol). The expected products of the translation with and without frameshifting are indicated as they would appear after gel electrophoresis. As in Fig. 1, the distance between the termination codons shown on the GP RNA is not to scale. (B) Autoradiogram of a 10 percent SDS–polyacrylamide gel of ^{35}S-labeled products of GP RNA translation. (Lane 1) Unprecipitated (2 μl); (lane 2) immunoprecipitated (4 μl) with rabbit antiserum to p19^{gag} (*32*); (lane 3) immunoprecipitated (10 μl) with rabbit anti–reverse transcriptase serum (*7*); (lane 4), immunoprecipitated (10 μl) with nonimmune rabbit serum. Immunoprecipitations were performed as described (*32*). The gel was exposed to Kodak XAR-5 film for 24 hours. Positions of labeled protein size markers are indicated.

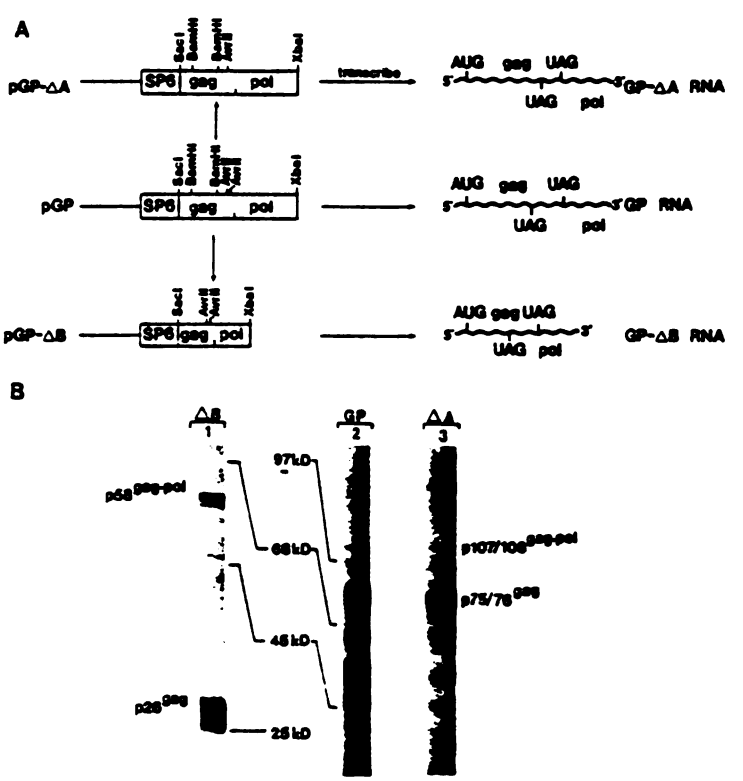


Fig. 3. Analysis of deletion mutants of pGP. (A) RNA's GP-ΔA (resulting from a 24-nt deletion in the *gag-pol* overlap region) and GP-ΔB (representing a 1386-nt internal deletion in *gag*) were transcribed from linearized plasmids pGP-ΔA and pGP-ΔB and translated (legend to Fig. 2). The plasmid pGP-ΔA was derived from pGP by cleavage with Avr II to delete 24-nt in the overlap region, followed by religation. The plasmid pGP-ΔB was constructed by cutting pGP at the two Bam HI sites within *gag*, filling in the cohesive ends with the Klenow fragment of *E. coli* DNA polymerase I (P-L Biochemicals), and religating. (B) Autoradiogram of 10 percent SDS–polyacrylamide gels of RRL translation products. (Lane 1) GP-ΔB RNA translation (3 μl); (lane 2) GP RNA translation (2 μl); (lane 3) GP-ΔA RNA translation (2 μl). Translations were performed as described (legend to Fig. 2) and products were analyzed without immunoprecipitation. Positions of molecular weight markers are indicated.

frameshifting occurred nonspecifically in the RRL, we challenged the RRL with an RNA that contains overlapping reading frames that are arbitrarily designed and not expected to induce frameshifting.

To do this, we replaced RSV sequences downstream of the Eco RI site near the end of *gag* in the plasmid pGP with DNA from the genome of human hepatitis B virus (HBV) (*10*) (Fig. 5A). The new plasmid, pGS-OF, is constructed so that the open reading frame for the HBV surface antigen ("sur") is −1 with respect to *gag*. The sur frame contains a termination codon 241 bp upstream of the first termination codon in the *gag* frame; thus GS-OF RNA offers a four times larger window for frameshifting to produce a fusion protein than does GP RNA. Nonspecific frameshifting during translation in vitro or errant synthesis by SP6 polymerase, putting the *gag* and sur genes in frame, would produce a 100-kD *gag*-sur fusion protein in addition to a 73-kD *gag* protein. As a control for the stability of the fusion protein, we constructed an additional plasmid, pGS-IF, in which the *gag* and sur genes are joined in a single long reading frame that encodes a 100-kD fusion protein (Fig. 5A).

The in vitro translation products from SP6 transcripts of pGS-OF and pGS-IF are shown in Fig. 5B. GS-OF RNA directs synthesis of abundant amounts of the 73-kD *gag* protein, without detectable synthesis of the 100-kD *gag*-sur fusion protein (lane 1). [A faint band in the relevant region of the gel is also seen after translation reactions without added RNA, and the labeled material is not immunoprecipitable with antiserum to $p19^{gag}$ (lane 2).] Translation of GS-IF RNA yields a stable 100-kD protein that is immunoprecipitable with antiserum to $p19^{gag}$ (lanes 3 and 4). Thus, within the sensitivity of these assays (about 0.1 percent), we find no evidence for unexpected frameshifting, either by transcriptional or translational mechanisms when unrelated genes are joined to produce extensive overlapping reading frames. We conclude that the production of *gag-pol* protein must depend on a mechanism that specifically recognizes some property of the *gag-pol* coding domain.

Possible mechanisms and implications. Our data show that ribosomal frameshift-

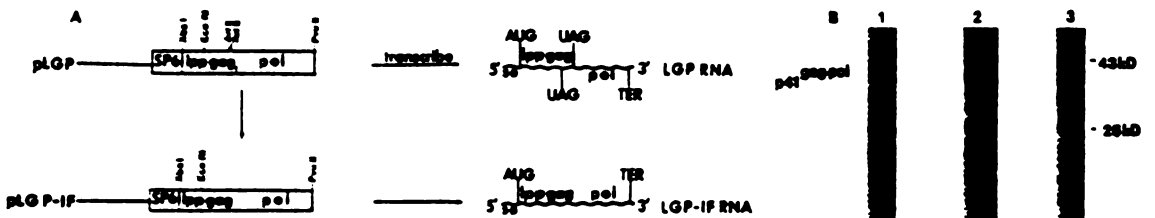


Fig. 4. Test for mRNA homogeneity. (A) Plasmid pLGP was constructed by replacing *gag* sequences upstream of the Eco RI within *gag* in pGP with sequences from the 5′ portion of the *E. coli* lipoprotein (*lpp*) gene (*8*) including the Shine-Dalgarno sequence (S-D), initiator AUG, and additional 28 codons. The plasmid pLGP-IF was derived from pLGP by cleaving with Avr II, filling in the cohesive ends with Klenow fragment, and religating. Transcription from Pvu II linearized plasmids was as described in the legend to Fig. 2. [Pvu II cuts within the plasmid vector several hundred nucleotides downstream of the RSV insert. The *pol* frame is shown closed by a stop codon (TER) presumed to be present in the vector-derived sequences.] (B) Autoradiogram of 15 percent SDS–polyacrylamide gel of translation products of LGP and LGP-IF RNA's: lane 1, products of LGP RNA (0.5 μg) translated in RRL (2 μl); lane 2, products of LGP RNA (1 μg) translated in the *E. coli* cell-free system (2 μl); lane 3, products of LGP RNA (1 μg) + LGP-IF RNA (0.15 μg) translated in the *E. coli* cell-free system (2 μl). The RRL translation was as described in legend to Fig. 2. The *E. coli* cell-free translation reactions (18 μl) were performed with reagents from an *E. coli* in vitro translation kit (Amersham) substituting an S135 extract (*34*) for the commercial S30 extract. Autoradiographic exposures were adjusted to give approximately equal intensities of the *gag* protein band.

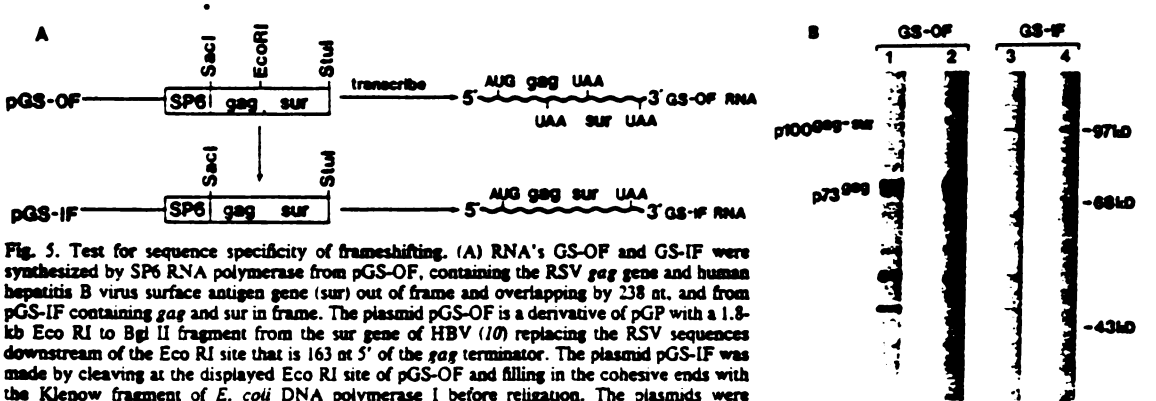


Fig. 5. Test for sequence specificity of frameshifting. (A) RNA's GS-OF and GS-IF were synthesized by SP6 RNA polymerase from pGS-OF, containing the RSV *gag* gene and human hepatitis B virus surface antigen gene (sur) out of frame and overlapping by 238 nt, and from pGS-IF containing *gag* and sur in frame. The plasmid pGS-OF is a derivative of pGP with a 1.8-kb Eco RI to Bgl II fragment from the sur gene of HBV (*10*) replacing the RSV sequences downstream of the Eco RI site that is 163 nt 5′ of the *gag* terminator. The plasmid pGS-IF was made by cleaving at the displayed Eco RI site of pGS-OF and filling in the cohesive ends with the Klenow fragment of *E. coli* DNA polymerase I before religation. The plasmids were linearized with Stu I prior to transcription, and the RNA was translated as described (Fig. 2). (B) Autoradiogram of a 10 percent SDS–polyacrylamide gel of translation products. (Lanes 1 and 2) GS-OF RNA translation products; (lane 1) unprecipitated (2 μl); (lane 2) immunoprecipitated (10 μl) with rabbit antiserum to $p19^{gag}$ (*32*); (lanes 3 and 4) GS-IF RNA translation products; (lane 3) unprecipitated (1 μl); (lane 4) immunoprecipitated (5 μl) with rabbit antiserum to $p19^{gag}$. Immunoprecipitations were performed as described (*33*). Positions of molecular weight markers are indicated.

ing occurs on an RSV RNA to produce two proteins from a single message in a mammalian cell-free translation system. The level of the *gag-pol* fusion protein produced in vitro (5 percent the level of the *gag* protein) is consistent with the ratio of these proteins in RSV-infected cells, making it likely that this form of translational suppression occurs in vivo as well. Thus, RSV may provide the first example of a ribosomal frameshift to control gene expression in a higher eukaryotic system. This mechanism has several precedents in prokaryotic systems and offers a simple way to produce a fixed ratio of two proteins. For RSV, frameshifting allows a small but predictable number of reverse transcriptase molecules to be incorporated into virus particles. Since the viral core is thought to be assembled through interactions between the genomic RNA and *gag* protein (*3*), inclusion of *pol* products in the virion reflects the ratio of *gag* to *gag-pol* proteins, which in turn, depends upon the efficiency of frameshifting. The bacteriophage T7 appears to use a very similar strategy to package a fixed ratio of two products of gene 10 into its phage heads (*11*).

Any of several mechanisms could account for the RSV frameshift.

1) Several host or viral factors might disrupt maintenance of the proper reading frame. For example, interactions between an RNA binding domain of the *gag* polyprotein and either the mRNA or a ribosomal RNA could cause ribosomes to stall during translation of the overlap region and occasionally change frame. However, the fact that the large deletions of *gag* in pGP-ΔB and in pLGP do not impair frameshifting argues against a role for *gag* protein.

2) In both bacteria and yeast, abnormal transfer RNA's (tRNA's) promote ribosomal frameshifting (*12*). These frameshift suppressor tRNA's are found in cells selected for phenotypic reversion of frameshift mutations in important genes; carry an extra nucleotide near the anticodon; and can cause −1 frameshifts at sequences related to the codons read by their normal homologs. Suppressor tRNA's of this type could not produce the RSV frameshift, which is in the −1 direction, but a role for a different type of abnormal tRNA is still possible. Any unusual tRNA would have to be present in many cell types and species though, since RSV *gag* and *gag-pol* proteins are produced in similar ratios in a wide spectrum of cells (*13*).

3) Certain normal tRNA's of *E. coli* can promote ribosomal frameshifting both in vitro (*14*) and in vivo (*15*) if present in high concentrations. These so-called "shifty" tRNA's may cause improper translocation of either two or four nucleotides resulting from a type of offset anticodon:codon pairing proposed by Weiss (*16*). Nucleotide context has been shown to influence both nonsense and missense suppression (*17*), and it almost certainly has an effect on whether a given codon will be the site of frameshifting. Short homopolymeric stretches of nucleotides have been implicated in the leakiness of frameshift mutations in the yeast mitochondrial *oxy-1* gene (*18*) and in the suspected frameshift during T7 gene 10 expression (*11*). The ribosomal frameshift that allows synthesis of *E. coli* release factor II (RFII) occurs just upstream of an amber stop codon, perhaps during a prolonged translational pause at this terminator (*19*). The RSV overlap region contains no long homopolymeric runs, but the UUA codon just 5′ of the *gag* amber stop codon is immediately preceded by another U residue (Fig. 1A). Slippage of a P-site leucine tRNA reading this UUA codon followed by mispairing to the UUU codon in the −1 frame would accomplish the necessary frameshift, and the position of the stop codon is reminiscent of the frameshift site in RFII. Placement of the frameshift at this point is also consistent with results of Rettenmier *et al.* (*20*) that suggest that the *gag-pol* fusion protein includes the arginine residue encoded in the *gag* frame just five codons upstream of the *gag* terminator (Fig. 1A). Site-directed mutagenesis of overlap nucleotides and amino acid sequencing of the fusion protein are now needed to examine these possibilities more rigorously.

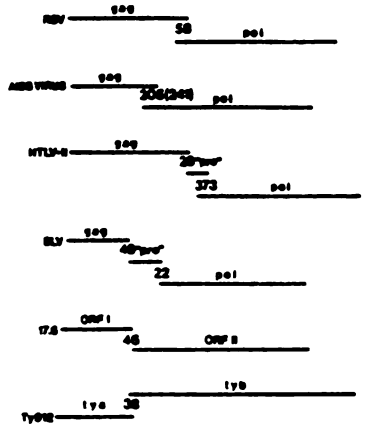


Fig. 6. Retrotransposons with overlapping of *gag* and *pol* genes. The *gag-pol* region of the AIDS retrovirus (*21*), HTLV-II (*22*), and BLV (*23*, *24*), and transposable elements Ty912 (*25*) and 17.6 (*26*) are represented. Placement of one open reading frame (bar) below another signifies a reading frame in the −1 position with respect to the overlying frame. Numbers indicate nucleotides shared by the two overlapping open reading frames. (Two isolates of the AIDS virus, LAV and ARV-2, have a 205-nt *gag-pol* overlap; it is 241-nt in HTLV-III.) The protease coding domains (pro) of HTLV-II and BLV separate the *gag* and *pol* genes but overlap both. Also, the Ty912 open reading frame tyb overlaps tya in the −1 direction.

Whatever the mechanism, frameshifting in higher eukaryotes is not likely to be limited to RSV. Certain other retroviruses and some related transposable elements already appear as probable candidates (Fig. 6). The AIDS (acquired immune deficiency syndrome) retrovirus (*21*), human T-cell leukemia virus type II (HTLV-II) (*22*), and bovine leukemia virus (BLV) (*23*, *24*) could all use a frameshifting mechanism to produce their (as yet unidentified) *gag-pol* fusion proteins. The latter two viruses would require two frameshift events in order to move out of the *gag* frame, through a segment encoding the viral protease, and into the *pol* frame.

The transposable elements Ty of yeast (*25*) and 17.6 of *Drosophila* (*26*), which resemble retroviruses in many other ways, also have overlapping open reading frames apparently encoding *gag*-like and *pol*-like functions. (Ty differs from the other elements by requiring a +1 shift to move from tya to tyb.) Two groups have shown that the product of the second open reading frame of Ty is expressed as a fusion protein with the product of the first, and no spliced mRNA has been detected by S1 nuclease protection analysis (*25*, *27*). Ribosomal frameshifting must be considered the favored explanation for expression in these systems as well.

Is frameshifting in eukaryotes restricted to viruses and transposable elements which must subvert normal cellular machinery in order to meet their requirements for successful replication? We think this unlikely. Still, strong evidence for the cellular counterpart of the RSV frameshift is lacking. An allele of the mitochondrial *Cox* II gene, encompassing two reading frames, is conserved between two species of trypanosomes (*28*, *29*), but, at least in the case of *Trypanosoma brucei*, a frame-corrected version of the gene exists and is active (*30*). The sequences of many nuclear genes include unassigned open reading frames either within or overlapping the major open reading frame. Analyses like the one described in this article might identify some of these sequences as substrates for frameshifting. Also, genetic screens could be specifically designed to uncover "shifty" sequences from eukaryotic genomes.

References and Notes

1. B. N. Fields, Ed., *Virology* (Raven, New York, 1985); H. R. B. Pelham, *Nature (London)* 272, 469 (1978); *Virology* 97, 256 (1979); M. March, G. Dragnea, C. Benicourt, *ibid.* 119, 193 (1982); E. G. Strauss, C. M. Rice, J. H. Strauss, *ibid.* 133, 92 (1984).
2. Y. Yoshinaka *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 82, 1618 (1985).
3. R. Weiss, N. Teich, H. Varmus, J. Coffin Eds., *Molecular Biology of Tumor Viruses: RNA Tumor Viruses* (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1982), chaps. 4–6.
4. D. E. Schwartz, R. Tizard, W. Gilbert, *Cell* 32, 853 (1983).
5. K. Beemon and E. Hunter, *Proc. Natl. Acad. Sci. U.S.A.* 74, 3302 (1977); A. Purchio, E. Erikson, R. L. Erikson, *ibid.*, p. 4661; S. R. Weiss *et al.*, *Cell* 15, 607 (1978).
6. D. A. Melton *et al.*, *Nucleic Acids Res.* 12, 7035 (1984).
7. H. Oppermann, J. M. Bishop, H. E. Varmus, L. Levintow, *Cell* 12, 993 (1977).
8. J. Ghrayeb and M. Inouye, *J. Biol. Chem.* 259, 463 (1984).
9. T. Jacks and H. E. Varmus, unpublished data. RNA was recovered from 100-μl RRL reaction mixtures with two phenol extractions and subsequent ethanol precipitation in the presence of 2.5*M* ammonium acetate and used to program either 18-μl *E. coli* or 25-μl RRL translation reactions as in Fig. 4.
10. P. Valenzuela, M. Quiroga, J. Zaldivar, P. Gray, W. J. Rutter, in *Animal Virus Genetics*, B. Fields, R. Jaenisch, D. F. Fox, Eds. (Academic Press, New York, 1980), pp. 57–70.
11. J. J. Dunn and F. W. Studier, *J. Mol. Biol.* 166, 477 (1983).
12. J. R. Roth, *Cell* 24, 601 (1981).
13. V. M. Vogt, D. A. Bruckstein, A. P. Bell, *J. Virology* 44, 725 (1982).
14. J. F. Atkins, R. F. Gesteland, B. R. Reid, C. W. Anderson, *Cell* 18, 1119 (1979).
15. R. Weiss and J. Gallant, *Nature (London)* 302, 389 (1983).
16. R. Weiss, *Proc. Natl. Acad. Sci. U.S.A.* 81, 5797 (1984).
17. M. Yarus and R. Thompson, in *Gene Function in Prokaryotes*, J. Beckwith, J. Davies, J. A. Gallant, Eds. (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1983), pp. 23–63.
18. T. D. Fox and B. Weiss-Brummer, *Nature (London)* 288, 60 (1980).
19. W. J. Craigen, R. G. Cook, W. P. Tate, C. T. Caskey, *Proc. Natl. Acad. Sci. U.S.A.* 82, 3616 (1985).
20. C. W. Rettenmier, R. E. Karess, S. M. Anderson, H. Hanafusa, *J. Virology* 32, 102 (1979).
21. S. Wain-Hobson, P. Sonigo, D. Danos, S. Cole, M. Alizon, *Cell* 40, 9 (1985); C. A. Rattner *et al.*, *Nature (London)* 331, 277 (1985); R. Sanchez-Pescador *et al.*, *Science* 227, 506 (1985).
22. K. Shimotohno *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 82, 3101 (1985).
23. N. Segata *et al.*, *ibid.*, p. 677.
24. N. R. Rice, R. M. Stevens, A. Buzy, R. V. Gilden, *Virology* 142, 357 (1985).
25. J. Clare and P. Farabaugh, *Proc. Natl. Acad. Sci. U.S.A.* 82, 2829 (1985).
26. K. Saigo *et al.*, *Nature (London)* 312, 659 (1985).
27. J. Meller *et al.*, *ibid.* 313, 243 (1985).
28. V. F. de la Cruz, W. Naveilmann, L. Simpson, *J. Biol. Chem.* 259, 15136 (1984).
29. L. Henegens *et al.*, *Nucleic Acids Res.* 12, 7327 (1984).
30. R. Benne, *Trends Genet.* 1, 117 (1985).
31. W. J. Delorbe *et al.*, *J. Virology* 36, 50 (1980).
32. H. Oppermann, J. M. Bishop, H. E. Varmus, unpublished data. Rabbit antisera were raised against gel-purified RSV $p19^{gag}$.
33. G. Ramsay, G. I. Evan, J. M. Bishop, *Proc. Natl. Acad. Sci. U.S.A.* 81, 7742 (1984).
34. M. Muller and G. Blobel, *ibid.*, p. 421.
35. We thank Irene Fecycz and Gunter Blobel for supplying the *E. coli* S135 extract and Pablo Garcia and Peter Walter for providing the *lpp* clone. We thank John Atkins, Bob Weiss, and Ray Gesteland for helpful discussions and materials and Janine Marinos for assistance in preparing the manuscript. Supported by a UCSF Earle C. Anthony/Dena's Fellowship (T.E.J.), American Cancer Society Research Professorship (H.E.V.), and grants from the National Institutes of Health and the American Chemical Society.

28 August 1985; accepted 30 October 1985

Chapter 2

Two Efficient Ribosomal Frameshifting Events Are Required for Synthesis of Mouse Mammary Tumor Virus *gag*-Related Polyproteins

Proc. Natl. Acad. Sci. USA
Vol. 84, pp. 4298–4302, June 1987
Microbiology

Two efficient ribosomal frameshifting events are required for synthesis of mouse mammary tumor virus *gag*-related polyproteins

(protease gene/*pol*/translation/retrovirus/type D retrovirus)

TYLER JACKS*, KIM TOWNSLEY†, HAROLD E. VARMUS*, AND JOHN MAJORS†

*Departments of Biochemistry and Biophysics and Microbiology and Immunology, University of California School of Medicine, San Francisco, CA 94143; and †Department of Biochemistry, Washington University School of Medicine, St. Louis, MO 63110

Contributed by Harold E. Varmus, February 12, 1987

ABSTRACT The primary translation products of retroviral *pol* genes are polyproteins initiated in an upstream gene (*gag*). To investigate the manner in which the *gag*-initiated polyproteins of the mouse mammary tumor virus are produced, we determined the nucleotide sequence of a 1.8-kilobase DNA fragment that spans the region between *gag* and *pol* in the C3H strain of mouse mammary tumor virus. The sequence reveals three overlapping open reading frames: the first encodes products of *gag* ($p27^{gag}$ and $p14^{gag}$); the second encodes a protein domain of unknown function (termed *X*) that is highly related to a similarly positioned sequence in simian type D retroviruses and the viral protease (*pro*); and the third encodes the reverse transcriptase. The reading frames are organized to permit uninterrupted readthrough from *gag* to *pol* if ribosomal frameshifts occur in the −1 direction within each of the two overlapping regions, one of which is 16 nucleotides in length and the other 13 nucleotides. Cell-free translation of RNA containing these overlap regions shows that fusion of the reading frames by ribosomal frameshifting occurs efficiently: about one-fourth of the ribosomes traversing the *gag–X/pro* overlap and one-tenth traversing the *X/pro–pol* overlap shift frames, generating *gag*-related polyproteins in ratios similar to those observed *in vivo*. Synthetic oligonucleotides containing either of the overlap regions inserted into novel contexts do not induce frameshifting; hence the overlapping portions of the reading frames are not sufficient to induce a frameshift event, and a larger sequence context or secondary structure may be implicated.

The mouse mammary tumor virus (MMTV) is unusual among retroviruses in that it can propagate and act as a carcinogenic agent in mammary epithelial tissue, it is transcriptionally regulated by steroid hormones, and it has a type B morphology. Nevertheless it utilizes strategies for macromolecular synthesis similar to those observed with other retroviruses: (*i*) *gag*-encoded viral core proteins are coordinately synthesized as components of a large precursor protein that is subsequently processed by a virus-encoded protease; (*ii*) the *pol*-encoded reverse transcriptase and integrase proteins are expressed at lower levels by similar processing of a large, fused *gag–pol* precursor; and (*iii*) *env*-encoded glycoproteins are expressed from a spliced, subgenomic mRNA (for review, see ref. 1).

MMTV also has biochemical features that distinguish it from most other retroviruses: (*i*) the presence of a protein coding domain of unknown function in the long terminal repeat (2, 3); (*ii*) the assembly of core particles (type A particles) in the cytoplasm of infected cells (4); and (*iii*) the synthesis of three precursor polyproteins (Pr77, Pr110, and Pr160) that possess *gag* antigenic determinants (5–10). The latter two properties are shared with type D primate viruses, such as Mason–Pfizer monkey virus (11, 12).

The three MMTV *gag*-related polyproteins have been detected both in virus-infected cells and following cell-free translation of viral mRNAs and are present at ratios of ≈30:10:1 (5–7, 9, 10). Tryptic peptide and immunological analyses have shown that Pr77 is the *gag* precursor, which is processed by proteolytic cleavage to yield viral core proteins (p10, p21, p27, and p14; for review, see ref. 1). Pr110 is thought to result from COOH-terminal extension of Pr77 (5, 6). A minor core protein, p30, probably cleaved from Pr110, contains peptides derived both from the COOH terminus of Pr77 and from sequences unique to Pr110; therefore, the junction between Pr77 and Pr110 should lie within p30 (6, 8). By analogy with other retroviruses, Pr160 should result from extended synthesis into the MMTV *pol* region and should be processed into reverse transcriptase and integrase activities.

In several retroviral systems, synthesis of large *gag–pol* precursor proteins results from inefficient suppression (at the level of 3–5%) of translation termination signals at the end of the *gag* region, either by in-frame nonsense-codon suppression (13) or by translation frameshifting (14). The relative abundance of the three MMTV *gag* polyproteins suggests that, if termination suppression is used in their generation, it must be efficient. In the studies described here, we have determined the nucleotide (nt) sequence of an MMTV genome in the region of *gag* and *pol* and shown that it contains a third gene, *X/pro*, that lies between *gag* and *pol* and briefly overlaps them both. Using cell-free translation of SP6 RNA polymerase transcripts of the same DNA, we also show that the extended products are efficiently generated *in vitro* by ribosomal frameshifting.

MATERIALS AND METHODS

MMTV DNA. The substrate for sequence analysis was initially isolated as a 4-kilobase (kb) *Pst* I fragment from unintegrated circular MMTV DNA purified from rat XC cells infected with the C3H strain of MMTV (15). The fragment was cloned directly into the *Pst* I site of pBR322.

Sequence Analysis. Sequencing was done by the method of Maxam and Gilbert (16) using both sets of overlapping deletions generated by BAL-31 nuclease and subfragments generated with various restriction endonucleases. Sequence comparisons between MMTV and simian retrovirus-1 (SRV-1) were done using the program ALIGN. $p27^{gag}$, $p14^{gag}$, and *pro* domains were delineated from protein-sequencing data of Hizi and Oroszlan (personal communication).

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: MMTV, mouse mammary tumor virus; SRV, simian retrovirus; RSV, Rous sarcoma virus; nt, nucleotide(s); HIV, human immunodeficiency virus; *Aha*-RNA, RNA generated by the restriction of pMGPP with *Aha* III; p10, p21, p27, and p14, viral core proteins of 10,000, 21,000, 27,000, and 14,000 daltons, respectively; Pr, precursor protein.

***In Vitro* Transcription and Translations.** SP6 transcription, rabbit reticulocyte translation, and immunoprecipitation reactions were done as described (14).

Plasmid Constructions. Plasmids were constructed as described in the figure legends using T4 DNA ligase (International Biotechnologies, New Haven, CT), *Escherichia coli* DNA polymerase I Klenow fragment (Boehringer Mannheim), and various restriction enzymes purchased from New England Biolabs. Oligonucleotides were synthesized by the Biomolecular Resource Center, University of California, San Francisco.

RESULTS AND DISCUSSION

The DNA and deduced protein sequences of the MMTV *gag–pol* region reveal three overlapping reading frames. To gain insight into the genetic organization of the MMTV *gag–pol* region, to assess the mechanism of synthesis of the nested polyproteins, and to study the relationship between MMTV and the type D retroviruses, we have determined the nucleotide sequence of a 1.8-kb DNA fragment that extends from midway through *gag* beyond the beginning of *pol*. The position of this *Pst* I–*Bgl* II fragment in the C3H MMTV genome is shown in Fig. 1.

Computer-assisted analysis of the nucleotide sequence reveals three extended and overlapping translational reading frames (Fig. 2). By comparing the amino acid sequences of these three reading frames with the sequences of other retroviral proteins and by taking into account the known pattern of MMTV polypeptide synthesis, we can assign each of the reading frames to known proteins. The first reading frame should code for the COOH terminus of $Pr77^{gag}$. By aligning our sequence with one previously deduced for the NH_2 terminus of the *gag* protein of the GR strain of MMTV (17), we were able to generate a hybrid GR/C3H sequence that predicts a *gag* protein with a molecular mass of 66 kDa, considerably smaller than its apparent molecular mass of 77 kDa in polyacrylamide gels. The portion of the *gag* sequence presented here codes for part of the major core protein $p27^{gag}$ and for all of $p14^{gag}$, a small nucleic acid-binding protein proteolytically cleaved from $Pr77^{gag}$ (18–20). Consistent with the postulated role for $p14^{gag}$, we find in its sequence two copies of the Cys-X_2-Cys-X_4-His-X_4-Cys peptide motif common to the small basic nucleic acid-binding proteins of all retroviruses (21).

Beginning 16 nucleotides upstream of the *gag* termination codon, in the −1 frame with respect to *gag*, is a second open reading frame that we have called *X/pro*. It extends for 304 codons beyond the *gag* terminus; fusion of the two frames would permit the synthesis of a protein of 95 kDa, again smaller than the experimentally determined size of Pr110. Contained within the COOH-terminal portion of this extended protein are two peptides, Asp-Thr-Gly-Ala-Asp and Gly-Arg-Asp, found in the presumed protease domains of most retroviruses (22). The protein domain, encoded by the first part of the second frame and called *X* in this discussion, is likely fused to $p14^{gag}$ to form the minor virion protein $p30^{gag-X}$ (5, 7).

A third reading frame begins 13 nucleotides upstream of the *X/pro* termination codon and should code for reverse transcriptase. Our sequence in this region is nearly identical to that previously determined for the beginning of the *pol* region of an MMTV endogenous element, *mtv-8* (23). The domain boundaries that we have drawn within our sequence are identical to those drawn by Moore *et al.* from a similar analysis of the sequence of the same region of the BR6 strain of MMTV (24).

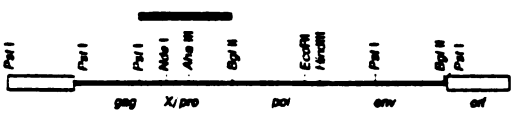


FIG. 1. A physical and genetic map of the C3H MMTV genome. The nature of the *X/pro* domain is discussed in the text. Solid bar, *Pst* I–*Bgl* II fragment the sequence of which is presented; *orf*, open reading frame in the long terminal repeat (box).

Similarity of MMTV *gag*, *X*, and *pro* Sequences to Analogous Sequences in Type D SRVs. MMTV and the type D SRVs both make intracytoplasmic type A particles and also have similar patterns of *gag*-related proteins (4–12). To assess the sequence relatedness of proteins encoded by the *gag–X/pro* region of MMTV to those encoded by the equivalent region of a type D virus, we compared the amino acid sequences of the MMTV *gag* and *X/pro* reading frames with those of the corresponding region of the type D virus SRV-1 (25). [The homologies found with SRV-1 hold for another type D virus, Mason–Pfizer monkey virus, whose amino acid sequences are >97% identical to SRV-1 in the regions analyzed (26).] Fig. 3 (*Upper*) compares the amino acid sequences of the first open reading frame in our sequence. The major core protein domains, $p27^{gag}$ for MMTV and $p24^{gag}$ for SRV-1, show 33% amino acid identity. This contrasts with <20% identity between the NH_2-terminal *gag* protein domains of MMTV-GR and SRV-1 (data not shown). The nucleic acid-binding proteins ($p14^{gag}$) of the two viruses are more closely related (about 45% amino acid identity), with the second of the cysteine repeats being more similar than the first. Fig. 3 (*Lower*) compares the sequences of the second open reading frame, *X/pro*. The *pro* domains of the two viruses show >50% amino acid identity, as do the two *X* domains (with one identical peptide of 11 amino acids). The conserved nature of the *X* domains of the two viruses suggests that *X* has a similar and probably important role in their respective replication cycles, perhaps in the formation of intracytoplasmic A particles, an unusual property shared by these two viruses. Also, the homology to MMTV *X* begins in the *X/pro* reading frame of SRV-1 upstream of the *gag* termination codon. Thus, in the synthesis of the SRV-1 *gag–X/pro* fusion protein, Pr110, the transition from the *gag* to *X/pro* reading frame is likely to occur upstream of or within the last coding domain of *gag* (Fig. 3 *Lower*).

Frameshifting *in Vitro*. If Pr110 is the product of *gag* and *X/pro* and Pr160 is encoded by these genes plus *pol*, either mRNA splicing or ribosomal frameshifting must occur in order to align the reading frames. In Rous sarcoma virus (RSV), the protease is encoded mainly at the end of *gag*, and *pol* overlaps *gag* directly (27). We have recently used *in vitro* transcription and translation methods to show that frameshifting during translation occurs in the RSV *gag–pol* overlap to produce a *gag–pol* fusion protein at about 5% efficiency (14). We have used a similar strategy to demonstrate that the MMTV *X/pro* and *pol* genes are also expressed via translational frameshifting.

Fig. 4*A* shows relevant regions of the plasmid pMGPP, the template for *in vitro* transcription by SP6 RNA polymerase. An MMTV DNA fragment extending from the *Nde* I site near the end of *gag* to a *Hin*dIII site downstream of *pol* (Fig. 1) was inserted into an SP6 vector containing the 5′ portion of the RSV *gag* gene such that the two *gag* genes are in-frame. (The RSV *gag* segment provides the translational initiation codon and a convenient antigen for immunoprecipitation of the translation products.) Linearization of pMGPP at any of the restriction enzyme sites shown in Fig. 4*A*, followed by transcription with bacteriophage SP6 RNA polymerase, yields RNAs that extend to different points within the *X/pro* and *pol* genes. Restriction of pMGPP with *Aha* III generates an RNA (*Aha*-RNA) that includes the hybrid *gag* gene and, in the −1 frame, ≈1/3 of the *X/pro* domain. Normal translation of *Aha*-RNA will yield a 41-kDa *gag* protein.

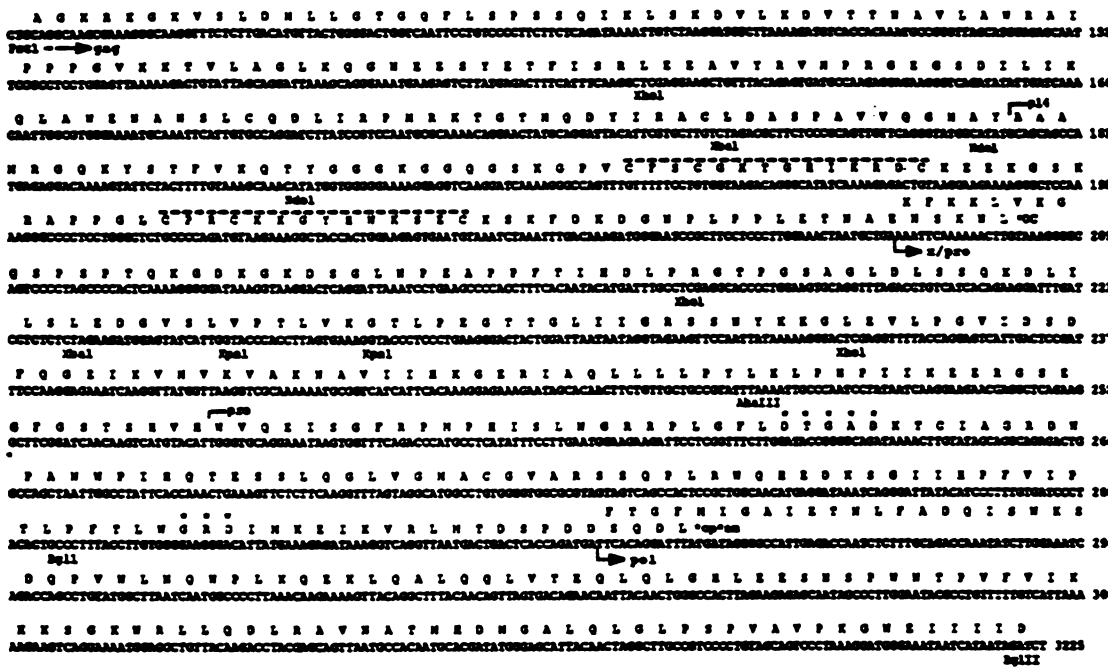


FIG. 2. DNA sequence of the *Pst* I–*Bgl* II fragment. Numbering is with respect to the start site of transcription as inferred from the sequence of Fasel *et al.* (17). The start points for the *X/pro* and *pol* open reading frames and the p14 and protease (pro) proteins as determined by Hizi and Oroszlan (personal communication) are shown, as are the cysteine-rich domains in p14 (—) and the conserved peptides in the protease domain (•••). op, opal; am, amber; and oc, ochre termination codons.

However, if some fraction of translating ribosomes are able to shift into the −1 frame within the 16-nucleotide *gag–X/pro* overlap region, a 55-kDa *gag–X/pro* fusion protein will also be produced. Similarly, upon translation of the RNA synthesized from *Bgl* II-digested pMGPP (*Bgl*-RNA), a ribosomal frameshift within the *gag–X/pro* overlap will result in a full-length 70-kDa *gag–X/pro* fusion protein, and successive frameshifts, first at the *gag–X/pro* overlap and then at the *X/pro–pol* overlap (13 nucleotide and requiring a −1 shift), will yield a *gag–X/pro–pol* fusion protein of ≈82 kDa. The protein products of *Hin*d-RNA translation should be the same as those produced from *Bgl*-RNA, except that the *gag–X/pro–pol* fusion will be 120 kDa.

The size and distribution of the actual ^{35}S-labeled products of the translation of these RNAs in a rabbit reticulocyte lysate are exactly as predicted by efficient frameshifting at the *gag–X/pro* and *X/pro–pol* overlaps (Fig. 4*B*, lanes 1, 4, and 7). All of the proteins are precipitated by an anti-RSV p19^{gag} serum (lanes 2, 5, and 8), but not by nonimmune serum (lanes 3, 6, and 9).

We have estimated the efficiency of frameshifting at the *gag–X/pro* overlap by calculating the ratio of the *gag* to *gag–X/pro* proteins produced in the same translation. The amount of radioactivity in the excised gel slices, after correcting for differential methionine content, reveals a frameshifting efficiency of ≈23%. Of those ribosomes that do shift into the −1 frame at the *gag–X/pro* overlap, ≈8% also shift at the *X/pro–pol* overlap (data not shown). These efficiencies, while remarkably high, are consistent with those required to produce the observed levels of the *in vivo* analogues (6). Furthermore, Moore *et al.* (24) have employed a similar strategy using a DNA clone of MMTV-BR6 and also observe single- and double-frameshifting at efficiencies consistent with those reported here.

Three Types of Translational Suppression Control Synthesis of Retroviral *gag–pol* Proteins. Our demonstration that the MMTV *gag*-fusion proteins are produced via single- and double-frameshifting events provides the third type of translational control over the synthesis of retroviral *gag* fusion proteins. A single ribosomal frameshift is sufficient to express the coding potential of the RSV (14, 27) and human immunodeficiency virus (HIV) *gag–pol* regions (28–30; T.J. and H.E.V., unpublished work). Yoshinaka *et al.* (13, 31) have shown that suppression of an amber codon separating the *gag* and *pol* domains of murine leukemic virus (MuLV) and feline leukemia virus (FeLV) occurs to synthesize a *gag–pol* fusion protein encoded in a single reading frame.

The existence of a separate open reading frame between *gag* and *pol* to encode the viral protease is not unique to MMTV. Bovine leukemia virus (BLV), human T-cell leukemia virus type 2 (HTLV-2), and the type D simian viruses (Mason–Pfizer monkey virus, SRV-1, and SRV-2) also have a three-tiered *gag–pro–pol* arrangement (25, 26, 32, 33). Experimental verification of the predicted frameshifts, however, is not yet available.

Determining the Signals for Frameshifting: The Overlap Nucleotides Are Not Sufficient. We presume that the highly efficient frameshifting observed in the MMTV overlaps occurs in response to one or more special codons that are either themselves unusual or are in unusual contexts. Moreover, "frameshift signals" might exist in the overlapping reading frames of other viruses that have been shown to, or are believed to, utilize frameshifting. In fact, a simple nucleotide sequence search has uncovered two potential signals. The MMTV *gag–X/pro* overlap includes the sequence A AAA AAC (where the triplets denote *gag* codons) (Fig. 5*A* and *C*); this sequence is also present in the upstream overlaps of BLV (32) and HTLV-2 (33). Furthermore, the amino acid sequence of the MMTV protein p30 as determined

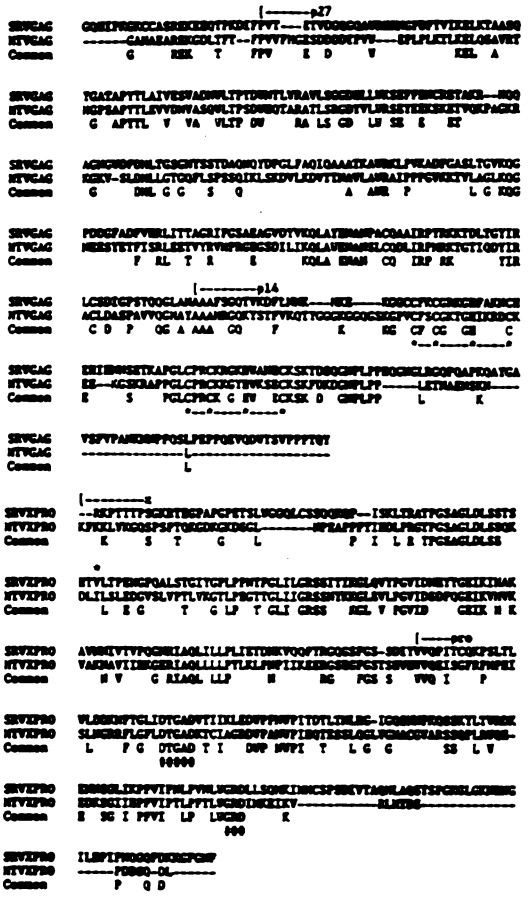


FIG. 3. (*Upper*) A comparison of the $p27^{gag}$ and $p14^{gag}$ domains of MMTV (MTVGAG) and a type D virus, SRV-1 (SRVGAG) (25). The MMTV p27 protein sequence is a hybrid GR/C3H sequence derived by joining the sequence of Fasel *et al.* (17) with ours. The NH_2 termini of MMTV $p27^{gag}$ and $p14^{gag}$ are shown, and the repeated cysteine-rich motif in p14 is indicated (•——•). (*Lower*) A comparison of the amino acid sequences of the *X/pro* domains of MMTV (MTVXPRO) and SRV-1 (SRVXPRO). The asterisk (*) above the valine in the SRV sequence indicates the end of the SRV-1 *gag* reading frame (25). The conserved *X* sequence TPGSAGLDLSS lies upstream of this site. Conserved *pro* peptides are indicated by (♦).

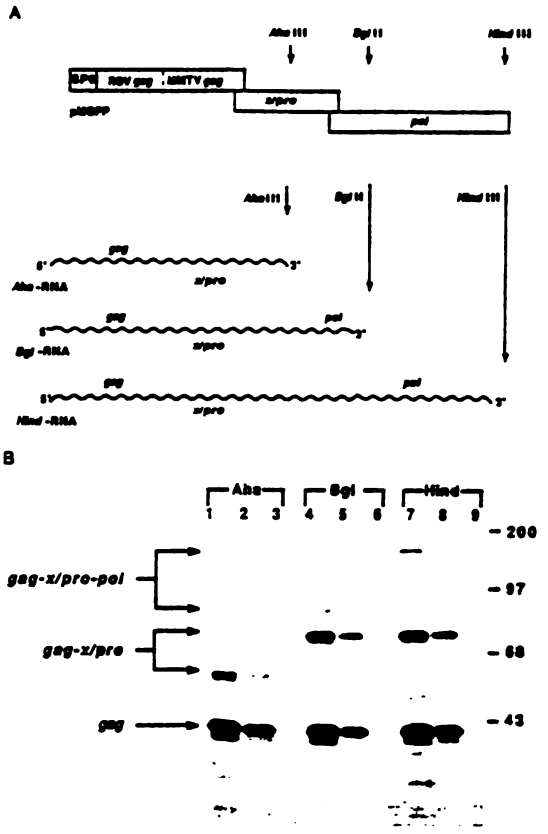


FIG. 4. (*A*) The plasmid pMGPP was constructed by ligating a 3.2-kb DNA fragment extending from a *Nde* I site within the p27 domain of MMTV *gag* (position 1796 in Fig. 2; also see Fig. 1) to a *Hin*dIII site downstream of *pol* (Fig. 1) to the plasmid pGP (14), previously digested with *Nde* I and *Hin*dIII. pGP is an SP6 vector containing the *gag–pol* domain of RSV; the *Nde* I site is in the p27 domain of *gag* (position 1290 in ref. 27), and the *Hin*dIII site is in the downstream polylinker. Cleavage of pMGPP with *Aha* III, *Bgl* II, or *Hin*dIII, followed by *in vitro* transcription with SP6 polymerase, generates the three mRNA species shown. (*B*) Fluorogram of ^{35}S-labeled proteins produced from rabbit reticulocyte lysate translation of *Aha*-, *Bgl*-, and *Hin*d-RNAs. (Lanes 1, 4, and 7) Unprecipitated proteins; (lanes 2, 5, and 8) proteins precipitated with anti-RSV $p19^{gag}$ serum; (lanes 3, 6, and 9) proteins precipitated with nonimmune rabbit serum. The positions of the expected products are indicated by arrows, and the positions of the molecular mass markers are indicated in kDa.

by Hizi and Oroszlan is consistent with the *gag–X/pro* frameshift occurring at this sequence (personal communication). The second potential signal, U UUA, is present in the downstream overlaps of MMTV (Fig. 5*A* and *C*), BLV (32), and HTLV-2 (33), as well as the single overlaps of RSV (27) and HIV (28–30). We have recently used amino acid sequencing and site-directed mutagenesis to show that this sequence is the frameshift site in RSV RNA (T.J., F. Masiarz, H.E.V., unpublished work). Given these potential signals and the fact that all of the retroviral frameshifts are in the −1 direction, a simple model for frameshifting would call for the tRNA reading the 0-frame codon (AAA, AAC, or UUA) to slip back one nucleotide and pair with the codon in the −1 frame.

Regardless of the details of the frameshifting mechanism, it seemed possible that all of the information required for frameshifting would reside in the signals described above or in the signals plus the adjacent nucleotides within the overlapping portions of the reading frames. To test whether the overlaps are sufficient to induce frameshifting, we cloned synthetic oligonucleotides corresponding to the two MMTV overlaps in between two new genes. As shown in Fig. 5*A*, the *gag–X/pro* and *X/pro–pol* overlaps join a portion of the 5′ end of the RSV *gag* gene and part of the 3′ region of the HIV *pol* gene in the plasmids pOL1 and pOL2; the plasmids are constructed so that the production of a *gag–pol* fusion protein is dependent on −1 frameshifting in the overlap segments.

The fluorogram in Fig. 5*B* shows the unprecipitated products of the translation of pOL1 and pOL2 RNAs. Despite copious amounts of the expected 43-kDa *gag* protein, neither RNA yields significant amounts of the *gag–pol* fusion (predicted to be 53 kDa). It appears, therefore, that in this new context the MMTV overlaps are not sufficient to promote efficient frameshifting. The simplest explanation for this failure is that sequences that border the overlaps in wild-type MMTV mRNA are involved in the frameshifting process.

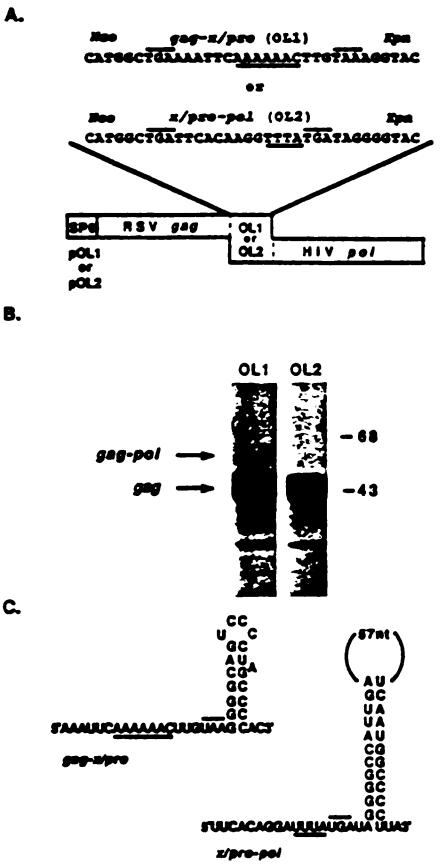


FIG. 5. (*A*) The plasmids pOL1 and pOL2. Synthetic oligonucleotides containing the plus strand of the *gag–X/pro* (pOL1) or *X/pro–pol* (pOL2) overlaps were ligated between an *Nco* I site in the 5' portion of the RSV *gag* and a *Kpn* I site in the 3' portion of the HIV *pol* in an SP6 vector. (The oligonucleotides were synthesized with *Nco* I- and *Kpn* I-compatible ends.) Ligation was followed by filling in the single-stranded region using the Klenow fragment of *E. coli* DNA polymerase I. The DNA sequences in the region of the overlaps were verified using the method of Chen and Seeburg (34). The termination codons that delineate the overlaps are overlined, and the proposed frameshift signals are underlined. (*B*) Fluorogram of unprecipitated, ^{35}S-labeled proteins produced from rabbit reticulocyte lysate translation of pOL1 and pOL2 RNAs. The predicted positions of the *gag* and *gag–pol* proteins are shown (arrows), and the positions of molecular mass markers are indicated in kDa. (*C*) Potential stem-loop structures located 3' to the MMTV *gag–X/pro* and *X/pro–pol* overlaps. The complete overlaps are shown, with the *gag* and *X/pro* termination codons overlined and the proposed frameshift signals underlined.

Alternatively, there could be a negative effect on frameshifting exerted by the sequences that now surround the overlaps. Interestingly, just 3' to both MMTV overlaps in their natural setting are potential stem-loop structures. Should these stem-loop structures be involved in frameshifting, they may act by stalling translating ribosomes, thereby promoting the tRNA slippage postulated above.

We thank Clive Dickson, Paul Luciw, Gordon Peters, and Steve Oroszlan for communicating results prior to publication and Janine Marinos for help with preparation of the manuscript. This work was supported by a University of California, San Francisco Earle C. Anthony/Dean's Fellowship (T.J.) and Public Health Service Grants CA38994 (J.M.) and CA39832 (H.E.V.). H.E.V. is an American Cancer Society Research Professor.

1. Weiss, R., Teich, N., Varmus, H. & Coffin, J., eds. (1982) *RNA Tumor Viruses* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
2. Dickson, C., Smith, R. & Peters, G. (1981) *Nature (London)* 291, 511–513.
3. Donehower, L. A., Huang, A. L. & Hager, G. L. (1981) *J. Virol.* 37, 226–238.
4. Bernhard, W. (1958) *Cancer Res.* 18, 491–509.
5. Anderson, S. J., Naso, R. B., Davis, J. & Bowen, J. M. (1979) *J. Virol.* 32, 507–516.
6. Dickson, C. & Atterwill, M. (1979) *Cell* 17, 1003–1012.
7. Dickson, C. & Peters, G. (1981) *J. Virol.* 37, 36–47.
8. Gautsch, J., Lerner, R., Howard, D., Teramoto, Y. & Schlom, J. (1978) *J. Virol.* 27, 688–699.
9. Massey, R. J. & Schochetman, G. (1979) *Virology* 99, 358–371.
10. Sen, G. C., Racevskis, J. & Sarkar, N. G. (1981) *J. Virol.* 37, 963–975.
11. Bradac, J. & Hunter, E. (1984) *Virology* 138, 260–275.
12. Chatterjee, S., Bradac, J. & Hunter, E. (1985) *Virology* 141, 65–76.
13. Yoshinaka, Y., Katoh, I., Copeland, T. D. & Oroszlan, S. (1985) *Proc. Natl. Acad. Sci. USA* 82, 1618–1622.
14. Jacks, T. & Varmus, H. E. (1985) *Science* 230, 1237–1242.
15. Majors, J. & Varmus, H. E. (1981) *Nature (London)* 289, 253–258.
16. Maxam, A. & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* 74, 560–564.
17. Fasel, N., Buetti, E., Firziaff, J., Pearson, K. & Diggelmann, H. (1983) *Nucleic Acids Res.* 11, 6943–6955.
18. Arthur, L., Long, C., Smith, G. & Fine, D. (1978) *Int. J. Cancer* 22, 433–440.
19. Dickson, C. & Peters, G. (1983) *Curr. Topics Microbiol. Immunol.* 106, 1–34.
20. Long, C. W., Berzinksi, R. & Gilden, R. V. (1977) *Int. J. Cancer* 19, 843–850.
21. Copeland, T., Oroszlan, S., Kalyanaraman, V., Sarngadharan, M. & Gallo, R. (1983) *FEBS Lett.* 162, 390–395.
22. Toh, H., Ono, M., Saigo, K. & Miyata, T. (1985) *Nature (London)* 315, 691.
23. Deen, K. & Sweet, R. (1986) *J. Virol.* 57, 422–432.
24. Moore, R., Dixon, M., Smith, R., Peters, G. & Dickson, C. (1987) *J. Virol.* 61, 480–490.
25. Power, M., Marx, P., Bryant, M., Gardner, M., Barr, P. & Luciw, P. (1986) *Science* 231, 1567–1572.
26. Sonigo, P., Barker, C., Hunter, E. & Wain-Hobson, S. (1986) *Cell* 45, 375–385.
27. Schwartz, D. E., Tizard, R. & Gilbert, W. (1983) *Cell* 32, 853–869.
28. Ratner, L., Haseltine, W., Patarca, R., Livak, K. J., Starcich, R., Josephs, S. F., Doran, E. R., Rafalski, J. A., Whitehorn, E. A., Baumeister, K., Ivanoff, L., Petteway, S. R., Jr., Pearson, M. L., Lautenberger, J. A., Papas, T. S., Ghrayeb, J., Chang, N. T., Gallo, R. C. & Wong-Staal, F. (1985) *Nature (London)* 313, 277–284.
29. Wain-Hobson, S., Sonigo, P., Danos, O., Cole, S. & Alizon, M. (1985) *Cell* 40, 9–17.
30. Sanchez-Pescador, R., Power, M. D., Barr, P., Steimer, K. S., Stempien, M. M., Brown-Shimer, S. L., Gee, W. W., Renard, A., Randolph, A., Levy, J. A., Dina, D. & Luciw, P. A. (1985) *Science* 227, 484–492.
31. Yoshinaka, Y., Katoh, I., Copeland, T. D. & Oroszlan, S. (1985) *J. Virol.* 55, 870–873.
32. Rice, N. R., Stephens, R. M., Burny, A. & Gilden, R. V. (1985) *Virology* 142, 357–377.
33. Shimotohno, K., Takahashi, Y., Shimizu, N., Gojobori, T., Golde, D. W., Chen, I. S. Y., Miwa, M. & Sugmura, T. (1985) *Proc. Natl. Acad. Sci. USA* 82, 3101–3105.
34. Chen, E. Y. & Seeburg, P. H. (1985) *DNA* 4, 165–170.

Chapter 3

Characterization of Ribosomal Frameshifting in HIV-1 *gag-pol* Expression

Characterization of ribosomal frameshifting in HIV-1 *gag-pol* expression

Tyler Jacks*, Michael D. Power†, Frank R. Masiarz†, Paul A. Luciw†‡, Philip J. Barr† & Harold E. Varmus*

* Departments of Biochemistry and Biophysics and Microbiology and Immunology, University of California Medical Center, San Francisco, California 94143, USA
† Chiron Corporation, 4560 Horton, Emeryville, California 94608, USA

Based on precedents from other retroviruses[1], the precursor of the human immunodeficiency virus (HIV-1) reverse transcriptase is predicted to be a polyprotein with a relative molecular mass (M_r) of 160,000 (160K) encoded by both the viral *pol* gene and the upstream *gag* gene. These two genes lie in different translational reading frames, with the 3′ end of *gag* overlapping the 5′ end of *pol* by 205 or 241 nucleotides[2–4]. Thus, production of the gag-pol fusion protein would require either messenger RNA processing or translational frameshifting. The latter mechanism has been shown in the synthesis of the gag-pol proteins of two other retroviruses, Rous sarcoma virus (RSV)[5] and mouse mammary tumour virus (MMTV)[6,7]. Here we report that translation of HIV-1 RNA synthesized *in vitro* by SP6 RNA polymerase yields significant amounts of a gag-pol fusion protein, indicating that efficient ribosomal frameshifting also occurs within the HIV-1 *gag-pol* overlap region. Site-directed mutagenesis and amino-acid sequencing localized the site of frameshifting to a UUA leucine codon near the 5′ end of the overlap.

Many, and probably all, retroviruses synthesize gag-pol fusion proteins, which are later cleaved by a virus-encoded protease to yield the mature pol proteins responsible for reverse transcription and integration[1]. The genetic structure of the *gag-pol* domains of retroviral genomes apparently precludes synthesis of this fusion protein, however, as *gag* and *pol* are either separated by an in-frame termination codon, overlapping in different reading frames, or are interrupted by a third gene (encoding the protease) which overlaps them both[1]. To circumvent these apparent blocks to synthesis of gag-pol fusion proteins, the four retroviruses so far examined use three different strategies: stop-codon readthrough in the case of murine leukaemia virus (MLV)[8] and feline leukaemia virus (FeLV)[9], and single and double ribosomal frameshifting in RSV[5] and MMTV[6,7] respectively.

The *gag-pol* domain of HIV-1 resembles most closely that of RSV in that the two genes overlap directly (with *pol* in the −1 frame with reference to *gag*), although the *gag-pol* overlap is considerably larger in HIV-1 (205 or 241 nucleotides as compared to 58 nucleotides for RSV)[2–4,10] (Fig. 1*a*). This region of the HIV-1 genome is apparently represented in a single mRNA species, the genome-length RNA, which is presumed to encode two proteins: the gag precursor, Pr55 gag, and a gag-pol fusion (160K)[11]. Typically, retroviral gag and gag-pol precursor proteins are synthesized at a ratio of about 10–20:1 (ref. 1).

To test whether the HIV-1 gag-pol fusion protein is also produced by ribosomal frameshifting, we used an experimental protocol that had previously allowed us to show frameshifting during the expression of the RSV and MMTV *pol* genes[5,6]. A DNA clone of HIV-1 (strain SF-2 (ref. 2)) encompassing the complete *gag-pol* domain was inserted downstream of the SP6 promoter[12] to form the plasmid pAGP (Fig. 1*a*). Linearization of pAGP at an *Nde*I site downstream of *pol* followed by transcription by SP6 RNA polymerase yields a unique species of RNA (N1-RNA) in which *gag* and *pol* are in their genomic, out-of-frame configuration. Translation of N1-RNA in a rabbit reticulocyte lysate system will, by normal translation, produce the gag protein, $Pr55^{gag}$; if some fraction of ribosomes shift into the −1 frame during translation of the 205 nucleotide (nt) *gag-pol* overlap, the 160K gag-pol fusion protein (referred to here as $Pr160^{gag\text{-}pol}$) will also be produced.

As shown in Fig. 1*b*, the unprecipitated translation products of N1-RNA include proteins of the correct molecular mass for $Pr55^{gag}$ and $Pr160^{gag\text{-}pol}$ (lane 1). Both of these proteins can be precipitated by an antiserum against the HIV-1 gag protein (lane 2); as expected, only the larger is recognized by an antiserum

‡ Present address: Department of Pathology, University of California, Davis, California 95616, USA.

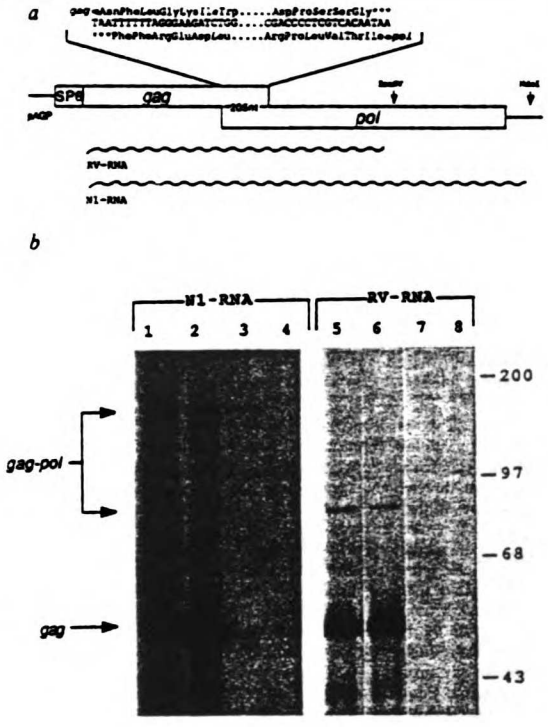


Fig. 1 Ribosomal frameshifting *in vitro*. *a*, The plasmid pAGP was constructed by inserting a 5.0 kilobase (kb) fragment from a DNA clone of HIV-1 (strain SF-2) extending from a *Sac*I site upstream of the *gag* initiator (position 225 in ref. 2) to an *Eco*RI site downstream of *pol* (position 5,296 in ref. 2) into the plasmid pSP64 (ref. 12) that had previously been digested with *Sac*I and *Eco*RI. The first 19 and last 16 nt of the 205 nt *gag-pol* overlap are shown translated in both the *gag* (above) and *pol* (below) frames. Linearization of pAGP with either *Eco*RV or *Nde*I, followed by *in vitro* transcription with SP6 RNA polymerase, yields the two RNAs shown (RV-RNA and N1-RNA). *b*, Translation of N1- and RV-RNA. The ^{35}S-labelled *in vitro* translation products of N1- and RV-RNA were electrophoresed through a 10% SDS-polyacrylamide gel either without immunoprecipitation (lanes 1 and 5) or after precipitation with anti-gag (lanes 2 and 6), anti-pol (lanes 3 and 7) or non immune rabbit serum (lanes 4 and 8). The gel was soaked in AMPLIFY (Amersham) and the proteins visualized by fluorography. The expected sizes of the gag and gag-pol proteins and positions of relative molecular mass standards (K) are shown. *In vitro* transcription, rabbit reticulocyte translations and immunoprecipitation reactions were carried out as described[5], except that the translation reactions were stopped after 15 min due to the instability of the gag-pol proteins. The anti-gag and anti-pol antisera were raised against purified HIV-1 proteins produced in *Escherichia coli*. The gag antigen corresponds to p25^{gag}, the central gag protein[17]; the pol antigen is analogous to p31^{pol}, the presumed integrase, encoded at the 3′ end of *pol*[18].

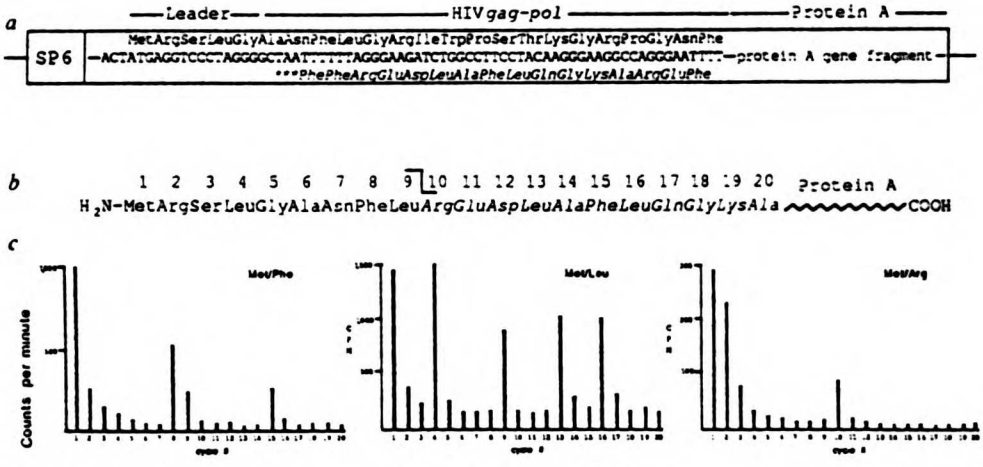


Fig. 2 The amino-acid sequence at the HIV-1 frameshift site. *a*, The plasmid pHSS was constructed by ligating a double-stranded oligonucleotide containing the first 50 nt of the HIV-1 overlap into an SP6 vector[12] between a short leader sequence (an initiator methionine codon plus five additional codons) and a portion of the *Staphylococcus* A-protein *A* gene (extending from position 742 to 2,001 in ref. 19). (The *A*-gene-containing vector pRIT2T was purchased from Pharmacia.) The *A*-gene segment is continuous with the *pol* frame of the HIV-1 insert. The leader sequence, which is in the *gag* frame, includes the first four amino acids of the chick pre-lysozyme protein[20], one of a small number of proteins known not to be amino-terminally acetylated during rabbit-reticulocyte lysate translation[21]. The two open reading frames are translated through the HIV-1 insert (*gag* above the nucleotide sequence, *pol* below it and in italics). *b*, The predicted amino-terminal amino-acid sequence of the fusion protein produced upon translation of pHSS-directed RNA, based on the model of frameshifting presented in the text. The predicted point of transition from the *gag* to *pol* frames is indicated (↴); the *pol*-encoded amino acids are shown in italics. *c*, Radioactivity profiles of automated Edman degradation of pHSS-encoded protein synthesized *in vitro* in the presence of [^{35}S]methionine and either [^{3}H]phenylalanine (left), [^{3}H]eucine (centre) or [^{3}H]arginine (right). 500 μl rabbit-reticulocyte translations were performed as described[5], except that 25 μg ml^{-1} α_2-macroglobulin (Boehringer) was added to inhibit proteolysis. The resulting fusion protein was purified using rabbit IgG-Sepharose (Pharmacia)[22]. The purified protein was subjected to 20 cycles of Edman degradation on an Applied Biosystems model 470A gas-phase protein sequencer using standard 03CATZ cycles; the products of each cycle were dried under vacuum, resuspended in scintillation fluid and counted in the ^{3}H channel.

Fig. 3 Two point mutations in the HIV-1 U UUA sequence strongly inhibit frameshifting. A single-stranded DNA vector (M13) containing an HIV-1 *gag-pol* insert was used as the template for site-directed mutagenesis[23]. Each of the first two thymidine residues in the overlap sequence T TTA (see Fig. 1*a*) were separately converted to cytosines, using 22-residue synthetic oligonucleotides containing the desired mutations. Double-stranded DNA fragments encompassing these mutations were isolated from the replicative form of the mutant vectors and used to replace the analogous fragment in the plasmid pAGP for *in vitro* analysis. The presence of the mutations were verified by DNA sequencing[24]. Translation of wild-type and mutant RV-RNAs (Fig. 1*a*) was followed by immunoprecipitation with anti-p25^{gag} antiserum (see legend to Fig. 1) and separation of the products on a 10% SDS-polyacrylamide gel. The expected size of the gag and gag-pol proteins and positions of relative molecular mass standards (K) are indicated.

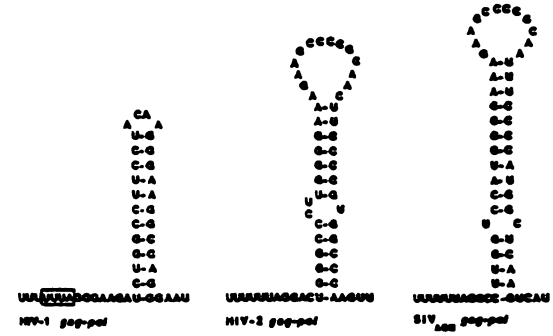


Fig. 4 Comparison of the HIV-1 frameshift site and potential stem-loop structure with homologous regions from the HIV-2 (ref. 13) and SIV$_{AGM}$ (ref. 14). The HIV-1 frameshift site is shown boxed.

directed againt pol (lane 3). The other minor, high molecular-mass protein species observed after N1-RNA translation (Fig. 1*b*) probably result from premature transcriptional or translational termination or internal translational initiation within *pol*.

To measure more accurately the ratio of the gag to gag-pol proteins and thereby estimate the frameshifting efficiency, we translated a shorter HIV-1 gag-pol RNA, RV-RNA (synthesized from *Eco*RV-linearized pAGP; Fig. 1*a*). RV-RNA also yields Pr55^{gag} as well as the predicted gag-pol fusion of 80K (Fig. 1*b*, lane 5). The two proteins are precipitable with the anti-gag serum (lane 6). But because the domain recognized by our anti-pol serum is encoded near the 3' end of the gene, the shorter gag-pol fusion is not recognized by this serum (lane 7). The ratio of the two proteins (as determined from the amount of radioactivity in the excised gel slices after correction for differential methionine content) is about 8:1, indicating a frameshifting efficiency of 11%. This efficiency is higher than the 5% we observed within the RSV *gag-pol* overlap[3], although not as high as was measured for the upstream (*gag-x/pro*) overlap in MMTV, where about one in four ribosomes changes frame[4].

To identify the site at which frameshifting occurs within the HIV-1 *gag-pol* overlap, we first noted the sequence U UUA (where the UUA is a codon in the *gag* frame) near the 5' end of the overlap (Fig. 1). This sequence also appears in the *gag-pol* overlap of RSV[3] and the *pro-pol* overlap of MMTV[4]. We have previously proposed[4] that −1 frameshifting at this sequence might be mediated by slippage of the leucyl transfer RNA reading the UUA codon back to the −1-frame UUU codon. (Amino-acid sequencing and site-directed mutagenesis is consistent with this model for RSV (T.J., F.R.M. and H.E.V., in preparation).) To test this model with HIV-1, we first determined a portion of the amino-acid sequence of a protein produced *in vitro* through frameshifting on an HIV-1 RNA. We cloned a synthetic DNA fragment corresponding to the first 50 nt of the HIV-1 overlap into an SP6 vector designed to allow rapid purification and sequencing of *in vitro*-synthesized proteins, producing the plasmid pHSS (Fig. 2*a*). The mechanism of frameshifting proposed above would result in a pHSS-encoded fusion protein with the amino-acid sequence Phe-Leu-Arg at positions 8–10 (Fig. 2*b*), where the arginine is the first amino acid encoded in the *pol* frame.

The histograms shown in Fig. 2*c* record the amount of radioactivity in the first 20 cycles of Edman degradation performed on purified, differentially labelled fusion protein. The peaks of radioactivity at positions 8, 9 and 10 in the phenylalanine-, leucine- and arginine-labelled samples, respectively, confirm that the shift into the *pol* frame occurs primarily after the decoding of the leucine UUA codon at position 9 and are consistent with the model for frameshifting presented above. The other peaks, including that at position 1, are also predicted by the methionine, phenylalanine, leucine and arginine codons, either upstream of the frameshift site in the *gag* frame or downstream of this site in the *pol* frame (Fig. 2*b*). It appears, however, that ~30% of frameshifting ribosomes do not insert leucine at position 9, based on the amount of radioactive leucine at this position and that expected given the amounts at positions 4 and 13 (Fig. 2*c*, middle). Those ribosomes that do not insert leucine at position 9 may incorporate phenylalanine, as there is a small peak at this position in the [^{3}H]phenylalanine label (Fig. 2*c*, left) that cannot be completely accounted for from carry-over of the [^{3}H]phenylalanine at position 8 (Fig. 2*b*). Perhaps after slipping from the UUA codon and before peptidyl transfer and translocation, a small fraction of the leucyl tRNA is released from the A site, and the exposed A-site UUU triplet then decoded by a phenylalanyl tRNA.

Although the amino-acid sequence at the transition from the *gag* to *pol* frames is consistent with the slippage model, other mechanisms, for example two-nucleotide translocation by the tRNA reading the last 0-frame codon (UUA), would produce the same amino-acid sequence. To test more directly whether leucyl tRNA slips back to the −1 frame, we constructed two site-directed mutations, converting each of the first two uracyl residues of the sequence U UUA to cytosines. We would expect these mutations to inhibit frameshifting by disrupting the site to which the UUA-reading tRNA slips (in the C UUA mutant) or by changing the UUA codon itself and thereby specifying a tRNA with an anticodon that is less likely to pair in the −1 frame (the U CUA mutant). As is shown in Fig. 3, both of these mutations strongly inhibit frameshifting *in vitro*, supporting the slippage model. Also, the absence of *gag-pol* expression in these two mutants suggests that frameshifting in the HIV-1 *gag-pol* overlap is confined to this single site near the 5' end.

The positioning of the HIV-1 frameshift site results in translation of both *gag* and *pol* sequences in the *gag-pol* overlap. The viral protease is encoded in the *pol* frame beginning 55 amino-acids downstream of the frameshift site (S. Oroszlan and S. Venkatesan, personal communications); the function of the *pol*-encoded amino acids between the frameshift site and the protease domain is unclear.

The amino-acid sequence and site-directed mutagenesis implicate the U UUA sequence as necessary for frameshifting within the HIV-1 overlap. We do not, however, believe that this sequence is sufficient to cause such a high degree of ribosomal shifting. More extensive mutational analysis of RSV has shown the importance of nucleotides just 5' to this virus's U UUA sequence (T.J., F.R.M. and H.E.V., in preparation) and we have previously proposed that stem-loop structures situated just 3' to the frameshift sites might be important in the retroviral frameshifting mechanism[6]. (Deletion and site-directed mutational analysis of the RSV stem-loop supports this proposal (T.J., H. Madhani and H.E.V., in preparation).) The HIV-1 frameshift site is followed closely by a G-C rich stem-loop structure (Fig. 4). Two other retroviruses, HIV-2 (ref. 13), and the simian immunodeficiency virus (SIV) (ref. 14, 15) also contain within their *gag-pol* overlaps U UUA sequences followed by G-C rich stem-loop structures (Fig. 4). The U UUA sequences are included in regions of identity with HIV-1 that are 9 nt in length, but the sequence composition of the stem-loop structures are very different. We expect that these other viruses also use ribosomal frameshifting at this site to produce their gag-pol fusion proteins. A complex set of transcriptional and post-transcriptional mechanisms has been proposed to regulate HIV-1 genes[16]. As has been suggested for some of the other mechanisms, a specific inhibitor of the frameshifting process could be an effective means of blocking HIV-1 replication.

This work was supported in part by a grant from the NIH. We thank J. Marinos for help with preparation of the manuscript and H. Madhani and P. Pryciak for helpful discussions. T.J. is a University of California Graduate Dean's Fellow. H.E.V. is an American Cancer Society Research Professor.

Received 22 September; accepted 23 November 1987.

1. Weiss, R., Teich, N., Varmus, H. & Coffin, J. *Molecular Biology of Tumor Viruses, RNA Tumor Viruses* (Cold Spring Harbor, 1985).
2. Sanchez-Pescador, R. *et al. Science* 227, 484–492 (1985).
3. Wain-Hobson, S., Sonigo, P., Danos, O., Cole, S. & Alizon, M. *Cell* 40, 9–17 (1985).
4. Rasmer, L. *et al. Nature* 313, 277–284 (1985).
5. Jacks, T. & Varmus, H. E. *Science* 230, 1237–1242 (1985).
6. Jacks, T., Townsley, K., Varmus, H. E. & Majors, J. *Proc. natn. Acad. Sci. U.S.A.* 84, 4298–4302 (1987).
7. Moore, R., Dixon, M., Smith, R., Peters, G. & Dickson, C. *J. Virol.* 61, 480–490 (1987).
8. Yoshinaka, Y., Katoh, I., Copeland, T. D. & Oroszlan, S. *Proc. natn. Acad. Sci. U.S.A.* 82, 1618–1622 (1985).
9. Yoshinaka, Y., Katoh, I., Copeland, T. D. & Oroszlan, S. *J. Virol.* 55, 870–873 (1985).
10. Schwartz, D. E., Tizard, R. & Gilbert, W. *Cell* 32, 853–869 (1983).
11. Mensing, M. A. *et al. Nature* 313, 450–458 (1985).
12. Mellon, D. A. *et al. Nucleic Acids Res.* 12, 7035–7056 (1984).
13. Guyader, M. *et al. Nature* 326, 662–669 (1987).
14. Franchini, G. *et al. Nature* 328, 539–543 (1987).
15. Chakraborti, L. *et al. Nature* 328, 543–547 (1987).
16. Chen, I. S. Y. *Cell* 47, 1–2 (1986).
17. Stoltzfus, K. S. *et al. Virology* 150, 283–290 (1986).
18. Stoltzfus, K. S. *et al. J. Virol.* 58, 9–16 (1986).
19. Ulmer, M. *et al. J. biol. Chem.* 259, 1695–1702 (1984).
20. Palmiter, R. D., Gagnon, J., Ericsson, L. H. & Walsh, K. A. *J. biol. Chem.* 252, 6386–6393 (1977).
21. Tsunasawa, S., Stewart, J. W. & Sherman, F. *J. biol. Chem.* 260, 5382–5391 (1985).
22. Ulmer, M. *et al. Gene* 23, 369–378 (1983).
23. Zoller, M. & Smith, M. *Nucleic Acids Res.* 10, 6487–6500 (1982).
24. Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* 74, 5463–5467 (1977).

Chapter 4

A Detailed Analysis of the Ribosomal Frameshift Site in the Rous Sarcoma Virus *gag-pol* Gene

Summary

The *gag-pol* protein of Rous sarcoma virus (RSV), the precursor to the enzymes responsible for reverse transcription and integration, is expressed from two genes that lie in different translational reading frames by ribosomal frameshifting. In this paper we localize the site of frameshifting to the last codon of *gag*, a UUA leucine, and show that the frameshifting reaction in RSV is mediated by slippage of two adjacent tRNAs by a single nucleotide in the 5′ direction. The *gag* terminator, which immediately follows the frameshift site, is not required for frameshifting. Other suspected retroviral frameshift sites, similarly structured to that of RSV, also mediate frameshifting when placed at the end of RSV *gag*.

Introduction

The vast majority of eukaryotic messenger RNAS (mRNAs) are monocistronic (Kozak, 1987). Unlike their prokaryotic counterparts, eukaryotic ribosomes tend not to initiate at internal methionine codons, and thus translation on a eukaryotic mRNA is usually limited to the open reading frame that follows the first AUG codon (Kozak, 1978). Consequently, there are very few examples of coordinate synthesis of multiple protein products from individual eukaryotic mRNA species (Kozak, 1986). There is an emerging class of eukaryotic mRNAs, however, that do encode multiple proteins, not by controlling where ribosomes begin translating but where they finish, either by the suppression of in-frame termination codons or by ribosomal frameshifting (Pelham, 1978, 1979; Yoshinaka et al., 1985a,b; Jacks and Varmus, 1985; Moore et al., 1987; Jacks et al., 1987; Brierly et al., 1987; Jacks et al., 1988).

In all known retroviruses, the *pol* gene (encoding the reverse transcriptase and integrase functions) lies downstream of the *gag* gene, which codes for the virus core proteins (Weiss et al., 1982). As shown in Figure 1, retroviruses arrange their *gag* and *pol* genes in one of three ways: in the same reading frame, separated by a single termination codon; in different reading frames, with *pol* briefly overlapping *gag* in the -1 direction; or with a third gene (encoding the viral protease, termed *pro*) intervening *gag* and *pol* and overlapping both. Despite these apparent blocks to continuous translation, all retroviruses initially express *pol* by first synthesizing a *gag-pol* (or *gag-pro-pol*) fusion protein that is later cleaved during virus assembly to yield the mature products. The ratio of this fusion protein to the product of the *gag*

gene alone is approximately 1:20 (Weiss et al., 1982).

Yoshinaka et al. (1985a) first showed by direct amino acid sequencing that the termination codon separating the murine leukemia virus (MLV) *gag* and *pol* genes is efficiently suppressed by a glutamine-charged tRNA. *In vitro* transcription and translation methods were then used to demonstrate ribosomal frameshifting during expression of the *gag-pol* protein of Rous sarcoma virus (RSV) (Jacks and Varmus, 1985) and human immunodeficiency virus type 1 (HIV-1) (Jacks et al., 1988) and double frameshifting in the synthesis of the mouse mammary tumor virus (MMTV) *gag-pro-pol* protein (Jacks, et al., 1987; Moore et al., 1987). The genomic sequences of several other retroviruses indicate that they utilize one of these three strategies to express their *pol* genes (Fig. 1).

In this report and the one that follows, we examine the sequence requirements for ribosomal frameshifting during translation of retroviral RNAs, using RSV as a model system. Radiolabelled amino acid sequencing and site-directed mutagenesis were used to localize the precise site of frameshifting in RSV RNA to the last *gag* codon, a UUA leucine, and suggest that the -1 frameshift is mediated by the simultaneous slippage of two tRNAs, the UUA-reading $tRNA^{Leu}$ and the one preceding it, by one nucleotide in the 5′ direction. Certain other sequences will functionally substitute for the natural RSV sequence at the frameshift site, including the sequences A AAA AAC and U UUA AAC, which are suspected to be the sites of frameshifting in other retroviral RNAs.

These studies demonstrate that frameshifting during retroviral gene

expression is dependent on specific nucleotide sequences at the frameshift site. However, we have previously provided evidence that these sequences are not sufficient to cause ribosomes to shift reading frame (Jacks et al., 1987). We demonstrate in the accompanying paper that an RNA secondary structure downstream of the frameshift site in RSV RNA is also required for efficient frameshifting.

Results

Common sequences within different retroviral overlap regions

Site-specific frameshifting within the various retroviral overlap regions was first suggested by the observation that each of them contains one of three common sequences. (The overlap regions are delineated on the 3′ side by the termination codon of the upstream (e.g. *gag*) open reading frame and on the 5′ side by the termination codon that demarcates the beginning of the downstream open reading frame (e.g. *pol*).) As shown in Table 1, several overlaps, including the *gag/pol* overlaps of RSV and HIV-1, contain the sequence U UUA (where the UUA is a leucine codon in the 0 frame). Two other common sequences U UUU and A AAC, appear in the remaining overlaps (Table 1). Amino acid sequencing has shown that two of these sequences, U UUA and A AAC, are the sites of frameshifting during HIV-1 *gag-pol* (Jacks et al., 1988) and MMTV *gag-pro* expression (Hizi et al., 1987), respectively.

In each of the overlaps, save one, these four nucleotide sequences are preceded by runs of three U, A, or G residues. (The U UUA sequence in the MMTV *pro-pol* overlap is preceded by the sequence GGA.) Thus, all of the retroviral overlaps include similar sequence motifs that are seven nucleotides in length (Table 1). Table 1 also includes putative frameshift site in the *gag/pol* overlaps of two retrotransposons of *Drosophila*, 17.6 and gypsy, and the mouse intracistronic A particle.

A simple model for frameshifting

Given the arrangement of nucleotides in these heptameric sequences and the fact that all of the retroviral frameshifts are in the -1 direction (Fig. 1), we have proposed that they might function by

allowing the tRNAs reading the 0-frame codons to occasionally slip into the -1 frame (Jacks et al., 1987, 1988). This model is shown in some detail for the RSV sequence A AAU UUA in Figure 2A, along with alternative models utilizing this same RNA sequence (Fig. 2B-D). According to the simultaneous slippage model (Fig. 2A), normal translation delivers a ribosome to the final two codons of *gag* such that the UUA codon is in the ribosomal A site being read by $tRNA^{Leu}$. The nascent protein is carried by the $tRNA^{Asn}$ reading the AAU codon in the P site (step I). Simultaneous slippage of these two tRNAs by one nucleotide in the 5′ direction leads to the conformation shown in step II, where both tRNAs are base-paired to the mRNA in two out of three anticodon positions. This interaction is made possible by the A and U residues 5′ to the AAU and UUA codons, respectively. Next, normal peptidyl transfer of the nascent protein to the $tRNA^{Leu}$ and translocation of this tRNA to the P site brings the first *pol* frame codon (AUA) into the A site, where it is normally decoded by $tRNA^{Ile}$ (step III). The other suspected frameshift sites listed in Table 1 would allow slippage by these or other tRNA species in a similar manner. Alternative models for frameshifting at this sequence include: two-base translocation by the $tRNA^{Leu}$ (Fig. 2B), and -1 or +2 slippage by the $tRNA^{Leu}$ while in the ribosomal P site (Fig. 2C and D).

Amino acid sequencing at the gag-pol junction

Before attempting to ascertain the mechanism of frameshifting, we first used amino acid sequencing to demonstrate that the proposed frameshift site is in fact the point where ribosomes begin translation in the *pol* frame. We replaced nearly the entire RSV *gag* gene with an

initiator methionine and two additional codons such that the position of translational initiation is just 10 codons from the proposed site of frameshifting (Fig. 3A). This plasmid, pGP-S, also has a portion of the *Staphylococcus aureus* protein A gene replacing the carboxy terminus of RSV *pol* in order to facilitate purification of the resulting "transframe" protein (the product of frameshifting).

If the *gag* frame UUA-leucine codon is the site of frameshifting, translation of pGP-S-encoded mRNA should proceed normally until the ribosomes decode this codon (the eleventh). Three of the four models shown in Figure 2 (parts A-C) call for the *pol* frame AUA-isoleucine codon to be the next decoded; the fourth model (Fig. 2D) would have the UUA codon followed by the *pol* frame GGG-glycine codon. Thus, the amino acid sequence of IgG-sepharose-purified material from translation of GP-S RNA should include leucine at position 11 and either isoleucine or glycine at position 12. If frameshifting on GP-S RNA occurs upstream of the UUA codon, the eleventh decoded triplet would be the *pol* frame UUU-phenylalanine (Fig. 3A), and the resulting transframe protein would contain phenylalanine at position 11 (Fig. 3A). Because the UUA-leucine is the last codon in the overlap (Fig. 3A), productive frameshifting cannot occur downstream of this site.

The histograms shown in Figure 3B display the amounts of radioactivity present in the first twenty cycles of Edman degradation of purified, pGP-S-encoded transframe protein synthesized *in vitro* in the presence of ^{35}S-methionine and either ^{3}H-leucine (panel I), ^{3}H-isoleucine (panel II), or ^{3}H-phenylalanine (panel III). The peaks of radioactive leucine and isoleucine at positions 11 and 12 and the lack

of radioactive phenylalanine confirm that the site of frameshifting is the terminal *gag* codon, UUA. The other observed peaks correspond to the methionine residue at position one and leucine residues at positions four, six, and eight in the *gag* frame and position 18 in the *pol* frame (Fig. 3A and B). The amino acid sequence at the frameshift site is consistent with three of the models shown in Figure 2 (parts A-C) and excludes the +2 slippage model (Fig. 2D), as well as models calling for a five-nucleotide translocation by the $tRNA^{Leu}$ or the action of a non-cognate tRNA (not shown).

Site-directed mutations in the frameshift site

In order to test the remaining three models of frameshifting (Fig. 2A-C), we constructed a series of site-directed mutations in and around the RSV frameshift site. To facilitate discussion of these mutations, the nucleotide positions have seen numbered as shown in Figure 4A. (The first position of the UUA codon is designated +1, with positive and negative integers proceeding 3′ and 5′, respectively.) The mutations were constructed by oligonucleotide-directed mutagenesis (see Experimental Procedures) in an SP6-promoter- containing plasmid carrying the complete RSV *gag* gene and about one-third of the *pol* gene; frameshifting was assayed by the ability of RNAs transcribed from these mutants to direct synthesis of a 108kD *gag-pol* fusion protein in a rabbit reticulocyte lysate *in vitro* translation system.

According to the simultaneous -1 slippage model (Fig. 2A), the seven nucleotides extending from the A residue at position -4 through to the A residue at position +3 (Fig. 4A) participate in the frameshift event as part of the 0- and -1 frame codons read by the frameshift-mediating

tRNAs. This model would predict that mutations in these seven positions would be inhibitory. Conversely, the nucleotides neighboring this heptameric sequence play no obvious role in this mechanism, and thus mutations in these positions should be silent.

The two remaining alternative models, two-nucleotide translocation and P site -1 slippage, predict a different spectrum of mutational effects. The nucleotides of the UUA codon (positions +1 to +3) should be required for both models. However, P site slippage (Fig. 2C) also demands the presence of the U residue at position -1 and could potentially be influenced by the 3′ neighboring *gag* termination codon (positions +4 to +6). In *E. coli*, frameshifting by P site slippage on certain homopolymeric sequences is greatly enhanced by an adjacent stop codon (Weiss et al., 1988). The requirements for the two-nucleotide translocation model (Fig. 2B) are more difficult to predict, but, again based on precedents from *E. coli*, there may be a context effect on both the 5′ and 3′ sides of the UUA codon. Nonsense suppression in *E. coli* is greatly influenced by the first and, to a lesser extent, second nucleotides following the suppressed stop codon (Bossi and Roth, 1980; Bossi, 1983; Miller and Alberti, 1983). The efficiency of missense suppression can be affected by the 5′ neighboring codon (Murgola et al., 1984).

Mutations in positions -1 through +2 abolish frameshifting; Support for $tRNA^{Leu}$ slippage

Slippage by the $tRNA^{Leu}$ from the *gag* frame into the *pol* frame, an essential component of two models shown in Figure 2 (A and C), requires the integrity of the run of three U residues in positions -1 to +2. A

mutation in the 5'-most U residue would impair the ability of the $tRNA^{Leu}$ to slip back. Mutation of the following two U residues would change the 0-frame codon and thereby specify a tRNA that would be less likely to slip back given a U at position -1. As shown in Figure 4B, mutation of any of these three U residues to any other nucleotide severely inhibits frameshifting efficiency. Translation of RNAs carrying these mutations results in undetectable amounts of the *gag-pol* protein (Fig. 4B). Thus, -1 slippage of the $tRNA^{Leu}$ is indicated. The model of two-nucleotide translocation by the $tRNA^{Leu}$ (Fig. 2B) would also predict severe effects by mutations in the U residues of the UUA codon (positions +1 and +2), but the preceding U residue (position -1) would not be expected to be equally critical.

Frameshifting is not affected by mutations in the gag terminator

Whether the $tRNA^{Leu}$ slips into the *pol* frame in concert with the $tRNA^{Asn}$ (Fig. 2A) or while resident in the P site after the $tRNA^{Asn}$ has exited the ribosome (Fig. 2C) can be deduced from the effects of mutations in the upstream AAU-asparagine codon (and the A that precedes it) and in the downstream *gag* termination codon.

The proximity of the *gag* terminator to the frameshift site is provocative, especially in light of the enhancement of frameshifting in *E. coli* by 3' neighboring stop codons (Weiss et al., 1988). An inhibitory effect of mutations in the *gag* terminator would support the P site slippage model (Fig. 2C), since the stop codon should only exert its effect while resident in the A site. However, whether the *gag* terminator was changed to a sense codon (positions +4 and +5 mutations and +6U), another stop codon (+6A), or was followed by a second stop

codon (+7U), the observed frameshifting efficiency was unchanged (Fig. 4B and C).

With the exception of the +6A mutation, all changes in the *gag* termination codon extend the *gag* open reading frame by 111 nt (Schwartz et al., 1983), resulting in a larger *gag* protein (Fig. 4B) and providing a significantly increased window in which frameshifting would yield a *gag-pol* fusion protein. In order to show that frameshifting occurs at the same site on these mutants as on wild-type RNA, we replaced the RSV sequences in the protein sequencing vector pGP-S with the analogous sequences from the +6U mutant. The transframe protein encoded by this mutant includes the *pol* frame isoleucine residue in the twelfth amino acid position, indicating that the transition to the *pol* frame occurs at the wild type location (data not shown).

Mutations in positions -4 through +2 inhibit frameshifting; a mutation further upstream does not

That frameshifting on RSV RNA occurs while the $tRNA^{Leu}$ is in the ribosomal A site and involves the simultaneous slippage of this tRNA and the P site $tRNA^{Asn}$ is most strongly supported by the reduction in frameshifting efficiency observed upon mutation of the three A residues in positions -4 to

-2. Converting any of these A residues to C (-4C, -3C, and -2C) reduces frameshifting efficiency from the wild-type value of 5% to approximately 1% (Fig. 4B and C). The -2U mutation has a similar inhibitory effect. As with the inhibition caused by mutations in the run of U residues, we attribute these deleterious effects to the specification of a tRNA with a decreased probability of slipping back (position -3 and -2 mutations)

or the disruption of the site to which the $tRNA^{Asn}$ normally slips (-4C).

The nucleotides 5′ to the -4 position should not influence frameshifting according to the simultaneous slippage model (Fig. 2A). Indeed, conversion of the wild type AC dinucleotide at positions -6 and -5 to GG does not alter the ratio of the *gag* to *gag-pol* protein (6G-5G, Fig. 4B and C).

Mutations are tolerated in the +3 position

In addition to favorable -1 frame base pairing by the $tRNA^{Leu}$ and $tRNA^{Asn}$, we considered the possibility that frameshifting might require the action of specialized isoacceptor tRNA species having the unusual ability to slip into the -1 frame. The effects of the three mutations in the +3 position suggest that if such a requirement exists, it is not absolute. The mutations that convert the UUA leucine codon to either UUG leucine or UUC phenylalanine, still allow efficient frameshifting (Fig. 4B, +3G and +3C), and the +3U mutation (creating a UUU phenylalanine codon) actually enhances the frameshifting efficiency two-fold.

More efficient frameshifting in the +3U mutant could mean that base pairing potential in the -1 frame is solely responsible for how often tRNAs and, consequently, ribosomes shift into the alternate reading frame. ($tRNA^{Phe}$ would have three of three anticodon position paired in the -1 frame rather than two of three for the wild-type $tRNA^{Leu}$.) Alternatively, frameshifting on the wild type and all three +3 mutant RNAs might involve a specialized $tRNA^{Leu}$ capable of decoding all codons with the sequence UUN (where N can be any nucleotide). To distinguish between these possibilities, we placed the +3U mutation into pGP-S (Fig.

5A) and determined the amino acid sequence at the frameshift site. As shown in Figure 5B, the +3U-encoded transframe protein contains phenylalanine at position 11 followed by leucine at position 12 (encoded by the *pol* frame codon UUA; Fig. 5A). Therefore, in the +3U mutant, frameshifting is mediated by a $tRNA^{Phe}$. As in the wild type, all three nucleotides of the 0-frame codon are read by the cognate tRNA followed by slippage of this tRNA by a single nucleotide in the 5′ direction.

Other retroviral frameshift signals functionally replace the RSV signal

The sequence at the end of the SRV-1 *pro* gene exactly matches the last seven nucleotides of RSV *gag*, except that the UUA codon is substituted with UUU (Table 1), the same substitution as in the RSV +3U mutant. While the SRV-1 sequence has not been tested for frameshifting in its native context, we strongly suspect that it is functional, and further that the frameshift is mediated in part by a $tRNA^{Phe}$.

Given the successful substitution by the presumed SRV-1 frameshift site, we next tested two other suspected frameshift sites, A AAA AAC and U UUA AAC (Table 1), for their ability to functionally replace the natural RSV frameshift sequence. As shown in Figure 6, frameshifting does occur on RNAs in which the last seven nucleotides of *gag* match these two sequences. The frameshifting efficiencies on these two RNAs are approximately 10% (lanes 2 and 3). Converting the last residue of these heptanucleotide sequences from C to A causes a ten-fold reduction in frameshifting efficiency (Fig. 6, lanes 4 and 5). It is noteworthy that one of these mutations (lane 4) produces in a run of seven consecutive A residues, yet the frameshifting efficiency is greatly reduced. This result argues that simple nucleotide redundancy is not

sufficient to mediate frameshifting in this context and suggests that only certain A-site tRNAs may be competent to shift into the -1 frame. This point is strengthened by the failure of the final RSV mutant, one which replaces the RSV U UUA sequence with G GGG, to allow any detectable frameshifting (Fig. 6, lane 7).

Discussion

The discovery of ribosomal frameshifting in RSV and other retroviruses has brought to light a previously unrealized mechanism for gene expression in higher eukaryotic cells. Understanding the details of the frameshifting reaction as it occurs in retroviral gene expression may lead to the discovery of programmed frameshifts in cellular genes and should address a more general problem in translation: the accurate maintenance of reading frame.

Mechanisms of frameshifting: Slippery codons

Homopolymeric or "slippery" sequences have been proposed to account for frameshifting in many genes in many systems. Runs of U residues have been implicated in the -1 frameshifting during translation of gene 10 of bacteriophage T7 (Dunn and Studier, 1983) and in the +1 and -1 frameshifts inferred from the activity of leaky frameshift alleles of the yeast mitochondrial gene *oxi*1 (Fox and Weiss-Brummer, 1980). The very efficient frameshift in the *RF*II gene of *E. coli* involves mispairing of the terminal 0-frame tRNA to the overlapping +1 frame codon (Craigen et al., 1985; Weiss et al., 1988). tRNA slippage by one or a few nucleotides in the 5′ and 3′ direction along several synthetic homopolymeric runs has recently been observed in *E. coli* by Weiss et al. (1988).

Simultaneous slippage

The amino acid sequence at the RSV *gag-pol* frameshift site and the results of the site-directed mutagenesis presented here indicate that ribosomal frameshifting in RSV (and, by analogy, other retroviruses) is also mediated by slippage of tRNAs along homopolymeric sequences.

However, the mechanism of frameshifting as it occurs in retroviral genes differs from those discussed above in that two adjacent tRNAs slip into the alternate (-1) frame. Thus, for RSV an A-site $tRNA^{Leu}$ and P-site $tRNA^{Asn}$ move from the last two *gag* codons into the *pol* frame, adopting a two-out-of-three base pair, anticodon-codon configuration. The requirement for at least two-of-three complementarity between the A site tRNA and the -1 frame codon seems absolute since any change that disrupts the run of three U residues that determines this pairing abolishes frameshifting. Complete complementarity (in all three anticodon positions) with the -1 frame codon can improve frameshifting efficiency, as evidenced by the two-fold enhancement obtained upon changing the wild-type RSV sequence U UUA to U UUU. The potential for alternative base-pairing interactions by the A site tRNA is not sufficient to explain frameshifting at this site, though, since the sequence G GGG cannot functionally replace the U UUA sequence. Also, while the bona fide retroviral frameshift site A AAA AAC can substitute for the wild-type RSV site, the sequence A AAA AAA is ten-fold less effective. These results suggest that specialized tRNAs might mediate frameshifting (see below).

The role of the tRNA reading the P site codon at the RSV frameshift site, while important, is less critical. Mutations in the run of three A residues responsible for the 0- and -1 frame interactions of the $tRNA^{Asn}$ lower the frameshifting efficiency by approximately five-fold, but the *gag-pol* protein is still readily observed. Consistent with the more relaxed P site requirements is the presence of several different P site codons in various retroviral frameshift sites while only three A

site codons are observed (see Table 1). The weakest apparent P site interaction occurs in the presumed frameshift site in the MMTV *pro/pol* overlap where a $tRNA^{Tyr}$ reading a 0-frame GAU codon is expected to mispair with the overlapping GGA codon. (Maintaining two-of-three base pair anticodon-codon contact here would require a wobble-like G:U base pair in the central anticodon position.) Further mutagenesis is needed to better define the requirements for the P site codon-anticodon interaction.

According to this model, slippage at the RSV frameshift site occurs prior to the translocation of the $tRNA^{Leu}$ to the P site. Despite the fact that this tRNA is presumably base paired to only two nucleotides after slippage, three nucleotides are translocated along with it to the P site in order that the next *pol* frame codon can occupy the A site. Thus, three-nucleotide translocation is not dependent on three base pairs between the codon and anticodon.

A neighboring termination codon is not required

In light of the fact that the RSV frameshift site encompasses two adjacent sense codons, it is not surprising that the *gag* termination codon, which lies immediately 3′ to the site, can be mutated without affecting frameshifting efficiency. Also, the majority of suspected frameshift sites listed in Table 1 do not directly precede a stop codon. In contrast, Weiss et al. (1988) have found in *E. coli* a strong influence on frameshifting along homopolymeric sequences by neighboring terminators. In addition, the presence of the ochre termination codon adjacent to the *RF*II frameshift site can stimulate frameshifting by as much as ten-fold (Weiss et al., 1988). In these examples, the stop

codons may act while resident in the ribosomal A site by arresting translation and allowing increased time for the P site tRNA to adopt a new alignment with the mRNA. In RSV a translational pause occurs at the frameshift site in the absence of the *gag* terminator by virtue of downstream RNA structure (see accompanying paper). The principle of broadening the time window for tRNA re-alignment may be similar, however.

Mechanism of frameshifting: Shifty tRNAs

Although the complementarity between the -1 frame codons and the anticodons of the tRNAs responsible for frameshifting on the RSV site is necessary for efficient frameshifting, such complementarity alone is insufficient to account for frameshifting in this setting as discussed above. The suggestion that only certain, specialized "shifty" tRNAs are competent to sample the alternative reading frames for suitable base-pairing interactions is also supported by the observation that in all of the documented or suspected retroviral frameshift sites (one of which is present in each of the retroviral overlaps), only three A site codons are found: UUA, UUU, and AAC (Table 1). Two of these three (UUA and UUU) are also present as P site codons in certain frameshift sites (Table 1). Discovery of the special features (if any) of the tRNAs that mediate frameshifting in retroviral genes must await their purification and sequencing.

The efficiency of frameshifting

The ratio of the RSV *gag-pol* protein to the *gag* protein is determined by the frameshifting efficiency, and this efficiency is at least partially determined by the nucleotide sequence at the frameshift

site. The natural frameshift site, A AAU UUA, allows approximately one in 20 ribosomes to shift into the *pol* frame. This efficiency can be increased two-fold by a point mutation (A to U) in the last position of this site. A similar increase is observed when the natural site is replaced by two other retroviral frameshift sites U UUA AAC and A AAA AAC. Conversely, the frameshifting efficiency can be reduced (to approximately 1%) by mutations in the first three A residues of the wild-type RSV site. Thus, by changing the nucleotide sequence at the frameshift site, *gag* to *gag-pol* ratios ranging from 10-100:1 can be achieved. The actual frameshift site has presumably been maintained because a *gag* to *gag-pol* ratio of 20:1 is optimal at some step in virus replication, most likely when the two proteins are associating with viral RNA to form the immature core particle. We are currently testing the effect on virus viability of mutations that increase or decrease the frameshifting efficiency and thereby alter this ratio.

Frameshift sites in other genes

Eukaryotic cells utilize several mechanisms to overcome the limitations of constrained translational initiation in order to express multiple protein products from individual genes. These mechanisms include: polyprotein synthesis, the production of multiple mRNAs (through the use of alternative sites of transcriptional initiation, splicing, or polyadenylation, or mRNA editing), and termination suppression. The potential for high-level ribosomal frameshifting introduces yet another means to generate multiple proteins from individual genes and, in fact, individual mRNAs.

Frameshifting in eukaryotic cells is not limited to retroviruses and

their related transposable elements. Brierley et al. (1987) have recently reported high level frameshifting in the F1/F2 overlap of the coronavirus avian infectious bronchitis virus (IBV). Although the site of this -1 frameshift has not been identified, the U UUA AAC sequence contained in the F1/F2 overlap is a likely candidate. This sequence is also present in two retroviral overlaps (Table 1) and, as shown above, allows efficient frameshifting when placed at the end of RSV *gag*.

To begin to investigate whether frameshifting occurs in other non-retroviral genes, we have recently conducted a computer-assisted search of eukaryotic gene sequences for the four heptanucleotide frameshift sites shown in this report to allow efficient frameshifting in RSV *gag* (R. Colgrove, T. Jacks, and H.E. Varmus, unpublished). While these sequences are much less frequent than would be expected from statistical considerations, they are found, in the correct reading frame, in many cellular and viral genes. We think it is unlikely that more than a few of these genes actually engage in frameshifting, however. As we have shown previously (Jacks et al., 1987) and investigate more thoroughly in the accompanying paper, at least three of the retroviral frameshift sites are by themselves insufficient to direct frameshifting; and for RSV a downstream stem-loop structure in the mRNA is also required. Only four of the potential frameshift sites uncovered in our search (all present in viral genes) are followed by stem-loop structures of significant stability. Three of these sites are present in the analogous position in three alpha virus genomes (Garoff et al., 1980; Rice and Strauss, 1981; Dalgarno et al., 1983); the fourth is located in the genome of tobacco etch virus (Allison et al., 1988). There is no

independent evidence that frameshifting occurs at any of these sites. We are currently assaying them for activity *in vitro*.

While our search failed to identify obvious cellular candidates for frameshifting, attention to those sequences that will allow efficient frameshifting in the proper context should hasten discovery of such genes.

Experimental Procedures

Amino acid sequencing

GP-S RNA and +3U-S RNA were translated in 500 μl rabbit reticulocyte translation reactions (Promega), supplemented with ^{35}S-methionine and either ^{3}H-leucine, -isoleucine, or -phenylalanine (Amersham); protein-A-containing products purified with rabbit IgG-Sepharose (Pharmacia); and amino acid sequence analysis performed as described (Jacks et al., 1988). The plasmids that code for the sequenced transframe proteins, pGP-S and p+3U-S, were derived from the plasmid pHSS (Jacks et al., 1988) by replacing the HIV sequences between the *Avr*II site located in six nucleotides from the initiator AUG and the *Bsh*II site that borders the protein A gene segment with an *Avr*II-*Bsh*II RSV *gag-pol* fragment. (These restriction sites are located at positions 2458 and 2724 in the sequence of Schwartz et al. [1983].) The RSV fragments were isolated from the wild-type plasmid pGP (Jacks and Varmus, 1985) (pGP-S) or the +3U mutant described here (+3U-S).

Site-directed mutagenesis

The protocol used for site-directed mutagenesis is an adaptation of that of Lewis et al. (1983). The plasmid pGP-S (or mutant derivatives) was linearized at a *Hpa*I site located 248 nt downstream of the *gag* terminator (position 2731 in the sequence of Schwartz et al. [1983]) and briefly digested with exonuclease III (New England Biolabs). (The extent of exonuclease III digestion was assayed using mung bean nuclease [New England Biolabs]; plasmids that had approximately 400 nt removed from each end were used as the substrates for mutagenesis.) Mutagenic oligonucleotides (10 pM) were added to 0.5 μg of *exo*III-treated plasmid

in a 5 μl reaction mixture containing 50 mM Tris-HCl (pH 8.0), 20 mM KCl, 7 mM $MgCl_2$, 0.1 mM EDTA, and 10 mM β-mercaptoethanol and heated to 65° for five minutes. After cooling to room temperature (five minutes), nucleotides (150μM dCTP, 150 μM dGTP, 150 μM TTP, 50 μM dATP, and 50 μM ATP), T4 DNA ligase (0.50 units, New England Biolabs) and Klenow (2.5 units, Boehringer) were added in a volume of 5 μl and the reaction incubated for 8-12 hrs at 15°C. The reactions were then ethanol precipitated in the presence of 2.5M NH_4Ac and used to transform *E. coli* HB101. Colonies harboring mutant plasmids were identified by differential screening using ^{32}P-labelled mutagenic oligonucleotides. The sizes of the mutagenic oligonucleotides were approximately 20 nt. The mutations were verified by double-stranded DNA sequencing (Chen and Seeburg, 1986).

SP6 transcriptions and rabbit reticulocyte translations were performed as described (Jacks and Varmus, 1985).

Acknowledgments

We thank Bruce Bowerman, Charles Craik, Michael Glotzer, Bob Weiss, John Atkins, and Ray Gesteland for helpful suggestions and discussion, and J. Marinos for reluctant assistance in preparing the manuscript. T.J. is a UCSF Graduate Dean's Fellow. H.E.V. is an American Cancer Society Professor of Molecular Virology.

References

Allison, R., Johnston, R.E. and Dougherty, W.G. (1986). The nucleotide sequence of the coding region of tobacco etch virus genomic RNA: Evidence for synthesis of a single polypeptide. Virology *154*, 9-20.

Bossi, L. (1983). Context effects: Translation of AUG codon by supressor tRNA is affected by the sequence following UAG in the message. J. Mol. Biol. *164*, 73-87.

Bossi, L. and Roth, J.R. (1980). The influence of codon context on genetic code translation. Nature *286*, 123-127.

Brierly, I., Boursnell, M., Birns, M., Bilmoria, B., Block, V., Brown, T. and Inglis, S. (1987). An efficient ribosomal frameshifting signal in the polymerase-encoding region of the coronavirus IBV. EMBO J. *6*:3779-3785.

Chakrabarti, L., Guyader, M., Alizon, M., Daniel, M.D., Desrosiers, R.C., Tiollais, P. and Sonigo, P. (1987). Sequence of simian immunodeficiency virus from macaque and its relationship to other human and simian retroviruses. Nature *328*, 543-547.

Chen, E.Y. and Seeburg, P.H. (1985). Supercoil sequencing: A fast and simple method for sequencing plasmid DNA. DNA 4, 165-170.

Craigen, W.J., Cook, R.G., Tate, W.P. and Caskey, C.T. (1985). Bacterial peptide chain release factors: Conserved primary structure and possible frameshift regulation of release factor 2. Proc. Natl. Acad. Sci. USA *82*, 3616-3620.

Dalgarno, L., Rice, C.M. and Strauss, J.H. (1983). Ross River virus 26S RNA: Complete nucleotide sequence and deduced sequence of encoded structural proteins. Virology *129*, 170-187.

Dunn, J.J. and Studier, F.W. (1983). Complete nucleotide sequence of bacteriophage T7 DNA and the locations of T7 genetic elements. J. Mol. Biol. *166*, 477-535.

Fox, T.D. and Weiss-Brummer, B. (1980). Leaky +1 and -1 frameshift mutations at the same site in a yeast mitochondrial gene. Nature *288*, 60-63.

Garoff, H., Frischauf, A.M., Simons, K., Lehrach, H. and Delius, H. (1980). Nucleotide sequence of cDNA coding for Semliki Forest virus membrane glycoproteins. Nature *288*, 236-241.

Guyader, M., Emerman, M., Sonigo, P., Claver, F., Montagnier, L. and Alizon, M. (1987). Genome organization and transcription of the human immunodeficiency virus type 2. Nature *326*, 662-669.

Hiramatsu, K., Nishida, J., Naito, A. and Yoshikura, H. (1987).

Molecular cloning of the closed circular provirus of human T cell leukemia virus type I: A new open reading frame in the *gag-pol* region. J. Gen. Virol. *68*, 213-218.

Hizi, A., Henderson, L.E., Copeland, T.D., Sowden, R.C., Hixson, C.V. and Oroszlan, S. (1987). Characterization of mouse mammary tumor virus *gag-pol* gene products and the ribosomal frameshift by protein sequencing. Proc. Natl. Acad. Sci. USA *84*, 7041-7046.

Jacks, T., Power, M.D., Masiarz, F.R., Luciw, P.A., Barr, P.J. and Varmus, H.E. (1988). Characterization of ribosomal mutations in HIV-1 *gag-pol* expression. Nature *331*, 280-283.

Jacks, T., Townsley, K., Varmus, H.E. and Majors, J. (1987). Two efficient ribosomal frameshift events are required for synthesis of mouse mammary tumor virus *gag*-related polypeptides. Proc. Natl. Acad. Sci. USA *84*, 4298-4302.

Jacks, T. and Varmus, H.E. (1985). Expression of the Rous sarcoma virus *pol* gene by ribosomal frameshifting. Science *230*, 1237-1242.

Kozak, M. (1987). An analysis of 5'-noncoding regions from 699 vertebrate messenger RNAs. Nucl. Acids Res. *15*, 8125-8148.

Kozak, M. (1978). How do eucaryotic ribosomes select initiation regions in messenger RNA? Cell *15*, 1109-1123.

Kozak, M. (1986). Bifunctional messenger RNAs in eukaryotes. Cell *47*, 481-483.

Lewis, E.D., Chen, S., Kumar, A., Blanck, G., Pollack, R.E. and Manley, J.L. (1983). A frameshift mutation affecting the carboxyl terminus of the simian virus 40 large tumor antigen results in a replication- and transformation-defective virus. Proc. Natl. Acad. Sci. USA *80*, 7065-7069.

Marlor, R.L., Parkhurst, S.M. and Corces, V.G. (1986). The *Drosophila melanogaster* gypsy transposable element encodes putative gene products homologous to retroviral proteins. Mol. Cell. Biol. *6*, 1129-1134.

Meitz, J.A., Grossman, Z., Lueders, K.K. and Kuff, E.L. (1987). Nucleotide sequence of a complete mouse intracisternal A-particle genome: No relationship to known aspects of particle assembly and function. J. Virol. *61*, 3020-3029.

Melton, D.A., Krieg, P.A., Rebagliati, M.R., Maniatis, T., Zinn, K. and Green, M.R. (1984). Efficient in vitro synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. Nucl. Acids Res. *12*, 7035-7056.

Miller, J.H. and Albertini, A.M. (1983). Effects of surrounding sequence on the suppression of nonsense codons. J. Mol. Biol. *164*, 59-71.

Moore, R., Dixon, M., Smith, R., Peters, G. and Dickson, C. (1987). Complete nucleotide sequence of a milk-transmitted mouse mammary tumor virus: Two frameshift suppression events are required for translation of *gag* and *pol*. J. Virol. *61*, 480-490.

Murgola, E.J., Pagel, F.T. and Hijazi, K.A. (1984). Codon context effects in missense suppression. J. Mol. Biol. *175*, 19-28.

Nilsson, B., Abrahmsen, L. and Uhlen, M. (1985). Immobilization and purification of enzymes with staphylococcal protein A gene fusion vectors. EMBO J. 4, 1075-1080.

Pelham, H.R.B. (1978). Leaky UAG termination codon in tobacco mosaic virus RNA. Nature *272*, 469-471.

Pelham, H.R.B. (1979). Translation of tobacco rattle virus RNAs in vitro: Four proteins from three RNAs. Virology *97*, 256-265.

Power, M.D., Marx, P.A., Bryant, M.L., Gardner, M.D., Barr, P.J. and Luciw, P.A. (1986). Nucleotide sequence of SRV-1, a type D simian acquired immune deficiency syndrome retrovirus. Science *231*, 1567-1572.

Ratner, L., Hazeltine, W., Patarca, R., Livak, K.J., Starcich, B., Josephs, S.F., Doran, E.R., Rafalski, J.A., Whitehorn, E.A., Baumeister, K., Ivanoff, L., Petteway, S.R., Pearson, M.L., Lautenberger, J.A.,

Papas, T.S., Ghrayeb, J., Chang, N.T., Gallo, R.C. and Wong-Staal, F. (1985). Complete nucleotide sequence of the AIDS virus, HTLV-III. Nature *313*, 277-284.

Rice, C.M. and Strauss, J.H. (1981). Nucleotide sequence of the 26S mRNA of Sindbis virus and deduced sequence of the encoded virus structural proteins. Proc. Natl. Acad. Sci. USA *78*, 1062-2066.

Rice, N.R., Stephens, R.M., Burny, A. and Gilden, R.V. (1985). The *gag* and *pol* genes of bovine leukemia virus: Nucleotide sequence and analysis. Virol. *142*, 357-377.

Sagata, N., Yasunaga, T., Tsuzuku-Kawamura, J., Ohishi, K., Ogawa, Y. and Ikawa, Y. (1985). Complete nucleotide sequence of the genome of bovine leukemia virus: Its evolutionary relationship to other retroviruses. Proc. Natl. Acad. Sci. USA *82*, 677-681.

Saigo, K., Kugimiya, W., Matsuo, Y., Inouye, S., Yoshioka, K. and Yuki, S. (1984). Identification of the coding sequence for a reverse transcriptase-like enzyme in a transposable genetic element in *Drosophila melanogaster*. Nature *312*, 659-661.

Sanchez-Pescador, R., Power, M.D., Barr, P.J., Steimer, K.S., Stempier, M.M., Brown-Shimer, S.L., Gee, W.W., Renard, A., Randolf, A., Levy, J.A., Dina, D. and Luciw, P.A. (1985). Nucleotide sequence and expression of an AIDS-associated retrovirus (ARV-2). Science *227*, 484-

492.

Shinnick, T.M., Lerner, R.A. and Sutcliffe, J.G. (1981). Nucleotide sequence of Moloney murine leukemia virus. Nature *293*, 543-548.

Schwartz, D.E., Tizard, R. and Gilbert, W. (1983). Nucleotide sequence of Rous sarcoma virus. Cell *32*, 853-869.

Shimotohno, K., Takahashi, Y., Shimizu, N., Gojobori, T., Golde, D.W., Chen, I.S.Y., Miwa, M. and Sugimura, T. (1985). Complete nucleotide sequence of an infectious clone of human T-cell leukemia virus type II: An open reading frame for the protease gene. Proc. Natl. Acad. Sci. USA *82*, 3101-3105.

Sonigo, P., Alizon, M., Staskus, K., Klatzmann, D., Cole, S., Danos, O., Retzel, E., Tiollais, P., Haase, A. and Wain-Hobson, S. (1985). Nucleotide sequence of the visna lentivirus relationship with the AIDS virus. Cell *42*, 369-382.

Sonigo, P., Barker, C., Hunter, E. and Wain-Hobson, S. (1986). Nucleotide sequence of Mason-Pfizer monkey virus: An immunosuppressive D-type retrovirus. Cell *45*, 375-385.

Stephens, R.M., Casey, J.W. and Rice, N.R. (1986). Equine infectious anemia virus *gag* and *pol* genes: Relatedness to visna and AIDS virus. Science *231*, 589-594.

Uhlen, M., Nilsson, B., Guss, B., Lindberg, M., Gatenbeck, S. and Philipson, L. (1983). Gene fusion vectors based on the gene for staphylococcal protein A. Gene *23*, 369-378.

Wain-Hobsen, S., Sonigo, P., Danos, O., Cole, S. and Alizon, M. (1985). Nucleotide sequence of the AIDS virus, LAV. Cell *40*, 9-17.

Weiss, R.B., Dunn, D.M., Atkins, J.F. and Gesteland, R.F. (1988). Slippery runs, shifty stops, backward steps and forward hops: -2, -1, +5 and +6 ribosomal frameshifting. Cold Spring Harbor Symp. Quant. Biol. *52*, in press.

Weiss, R., Teich, N., Varmus, H. and Coffin, J. (1982). RNA Tumor Viruses. Cold Spring Harbor Laboratory.

Yoshinaka, Y., Katoh, I., Copeland, T.D. and Oroszlan, S.J. (1985a). Murine leukemia virus protease is encoded by the *gag-pol* gene and is synthesized through suppression of an amber termination codon. Proc. Natl. Acad. Sci. USA *82*, 1618-1622.

Yoshinaka, Y., Katoh, I., Copeland, T.D. and Oroszlan, S.J. (1985b). Translational readthrough of an amber termination codon during synthesis of feline leukemia virus protease. J. Virol. *55*, 870-873.

Figure Legends

Table 1. Demonstrated and suspected retroviral frameshift sites. Common heptanucleotide sequence motifs are present in all retroviral overlaps known or presumed to contain sites of frameshifting. The heptanucleotides are shown along with their neighboring sequences (in smaller type), and the distance (in nucleotides) between the 3' nucleotide of the heptameric sequence and the 3' end of its overlap (as delineated by the first nucleotide of the 0-frame termination codon). The sequences are grouped according to their final three nucleotides; these comprise a codon in the upstream (e.g. *gag*) gene. Two of these codons, UUA and AAC, have previously been identified as the sites of frameshifting (see text). Evidence that the entire heptanucleotide sequence may participate in the frameshifting reaction is presented here. 17.6 sequence by Saigo et al. (1985); gypsy by Marlor et al. (1986); and mouse intracisternal A particle (IAP) by Meitz et al. (1987). The references for remaining sequences can be found in legend to Figure 1.

Figure 1. The genetic structure of retroviral *gag-pol* domains. All retroviruses arrange their *gag* and *pol* genes in one of the three following ways. Type 1: *gag* and *pol* in the same translational reading frame separated by a single termination codon. MLV (Shinnick et al., 1981) and feline leukemia virus (FeLV) (Yoshinaka et al., 1985b) exhibit this arrangement. Type II: *pol* directly overlapping *gag* in the -1 reading frame. Examples of this type include RSV (Schwartz et al.,

1983), HIV-1 (Wain-Hobson et al., 1985; Ratner et al., 1985; Sanchez-Pescador et al., 1985), HIV-2 (Guyader et al., 1987), simian immunodeficiency virus (SIV) (Chakrabarti et al., 1987), Visna virus (Sonigo et al., 1985), and equine infectious anemia virus (EIAV) (Stephens et al., 1986). Type III: *gag* and *pol* separated by a third gene, *pro* (encoding the viral protease), that overlaps them both. The *pro* and *pol* genes lie in the -1 reading frame relative to the genes that precede them (*gag* and *pro*). Retroviruses in the Type III category include MMTV (Jacks et al., 1987; Moore et al., 1987), simian retrovirus type 1 (SRV-1) (Power et al., 1986), Mason-Pfizer monkey virus (MPMV) (Sonigo et al., 1986), bovine leukemia virus (BLV) (Sagata et al., 1986; Rice et al., 1985), and human T-cell leukemia virus type 1 (HTLV-1) (Hiramatsu et al., 1987) and type 2 (HTLV-2) (Shimotohno et al., 1985).

Figure 2. Models for frameshifting at the RSV frameshift site. (A) Simultaneous -1 slippage. $tRNA^{Asn}$ carrying the nascent peptide (jagged line) and $tRNA^{Leu}$ are shown bound to the *gag* frame codons, AAU and UUA, in the ribosomal P and A sites (step 1). Simultaneous slippage of the two tRNAs by one nucleotide in the 5′ direction results in their complexing with the adjacent *pol* frame codons, AAA and UUU, with base pairs (bars) in the first and second codon positions (step 2). Normal peptidyl transfer and three-nucleotide translocation brings the next *pol* frame codon, AUA, into the A site where it is decoded by $tRNA^{Ile}$ (step 3). Note that the slippage could also occur following peptidyl transfer and prior to translocation. Also, the sequence of the tRNA anticodons shown in this and other models are based on standard Watson-Crick base

pairs. The actual anticodon sequences are not known (see text).
(B) Two-nucleotide translocation. The same point in translation described for step 1 in model A above is followed by peptidyl transfer and the translocation of only the first two nucleotides on the UUA codon into the P site. The *pol*-frame AUA codon thus fills the A site, where it is decoded by $tRNA^{Ile}$.
(C) P site -1 slip. Following the configuration shown in step I for model A, a normal, three-nucleotide translocation brings the UUA codon and $tRNA^{Leu}$ into the site. $tRNA^{Leu}$ slips by one nucleotide in the 5′ direction, mispairing with the *pol* frame UUU codon. The *pol* frame AAU codon is then available to the $tRNA^{Ile}$ in the A site.
(D) P site +2 shift. As in model C above, the $tRNA^{Leu}$ correctly arrives at the P site, but then slips by two nucleotides in the 3′ direction, mispairing with the *pol* frame AUA codon with base pairs in the second and third codon positions. The A site is thus occupied by the successive *pol* frame codon, GGG, which is subsequently decoded by a $tRNA^{Gly}$.

Figure 3. The amino acid sequence at the RSV frameshift site.
(A) A portion of the protein sequencing vector pGP-S. Downstream of the SP6 promoter (Melton et al., 1984) was cloned a sequence composed of an initiator methionine codon and two additional codons (Arg and Ser) followed in frame by the 3′ end of the RSV *gag* gene, beginning with the leucine codon located seven codons upstream of the *gag* terminator. (The N-terminus of the protein encoded by RNA transcribed from pGP-S, Met-Arg-Ser-Leu, is not acetylated in the rabbit reticulocyte lysate system

[Jacks et al., 1988].) Following approximately 250 nt of RSV *pol* sequence in pGP-S is a segment of the *Staphylococcus aureas* protein A gene (Uhlen et al., 1983) in frame with *pol*. Thus, the transframe protein encoded by GP-S RNA is readily purified using IgG-Sepharose (Nilsson et al., 1985). The nucleotide sequence of the 5′ end of this hybrid gene is shown along with the translation in the *gag* frame (above the nucleotide sequence) and *pol* frame (below the sequence and in italics).

(B) Possible N-terminal amino acid sequences of the transframe protein synthesized from GP-S RNA. The sequence of both *gag* and *pol* frame (below and in italics) amino acids are shown. The amino acid positions are numbered.

(C) Histograms recording the amount of radioactivity present in the first 20 cycles of Edman degradation performed on IgG-Sepharose-purified protein synthesized from GP-S RNA in a rabbit reticulocyte supplemented with ^{35}S-methionine and either ^{3}H-leucine (panel I), -isoleucine (panel II), or -phenylalanine (panel III). CPM refers to counts per minute above background.

Figure 4. The effects of point mutations in and around the RSV frameshift site.

(A) The nucleotides at the end of the RSV *gag* gene are designated numerically as shown.

(B) Fluorogram of a 10% SDS-polyacrylamide gel containing total ^{35}S-labelled products of rabbit reticulocyte translations directed by either wild-type (wt) or various mutant RSV RNAs. The specific mutations are

indicated by their position (according to part A, above) and nucleotide change. The position of the *gag* and *gag-pol* proteins and molecular mass markers (in kD) are indicated.

(C) Summary of mutational effects. The wild-type RSV sequence is shown horizontally, and the possible base changes listed vertically. Narrow and thick downward arrows indicate decreases in frameshifting efficiency of approximately five- and greater than ten-fold, respectively. A narrow, upward arrow indicates a two-fold increase in frameshifting efficiency. NC symbolizes no change in efficiency. Frameshifting efficiencies were calculated from the amount of radioactivity in excised gel slices containing the *gag* and *gag-pol* proteins after correction for differential methionine content. Blank entries indicate either that mutations were not constructed or that the nucleotide corresponds to the wild-type sequence.

Figure 5. The amino acid sequence at the frameshift site in the RSV +3U mutant.

(A) A segment of the RNA sequence synthesized from the plasmid p+3U-S (a derivative of pGP-S including the +3U mutation; see Experimental Procedures) and its translation in the *gag* frame (above the sequence) and *pol* frame (below the sequence and in italics). The numbers refer to amino acid positions in the resulting transframe protein.

(B) Histograms recording the amount of radioactivity in the first 20 cycles of Edman degradation performed on IgG-Sepharose purified protein synthesized from +3U-S RNA in a rabbit reticulocyte system supplemented with ^{35}S-methionine and either ^{3}H-leucine (panel I) or -phenylalanine

(panel II). CPM refers to counts per minute above background.

Figure 6. Frameshifting on functional and defective replacements of the RSV frameshift site.

Fluorogram of a 10% SDS-polyacrylamide gel containing total ^{35}S-labelled products of rabbit reticulocyte lysate translation of wild-type RSV RNA (lane 1) or various mutant derivatives (lanes 2-6). The mutations affect all or a part of the RSV frameshift site (the last seven nucleotides of the *gag* gene). The sequences of the RSV frameshift site and mutant sites are shown above the lanes. The position of the *gag* and *gag-pol* proteins and molecular mass standards (in kD) are indicated.

Virus or transposable element	Overlap	Sequence	Distance from 3' end of overlap
RSV	gag/pol	ACA AAU UUA UAG	0
HIV-1	gag/pol	AAU UUU UUA GGG	198
HIV-2	gag/pol	GGU UUU UUA GGA	267
SIV	gag/pol	GGU UUU UUA GGC	213
Gypsy	gag/pol	AAU UUU UUA GGG	51
MMTV	pro/pol	CAG GAU UUA UGA	0
SRV-1	pro/pol	GGA AAU UUU UAA	0
MPMV	pro/pol	GGA AAU UUU UAA	0
17.6	gag/pol	GAA AAU UUU CAG	30
Mouse IAP	gag/pol	CUG GGU UUU CCU	3
MMTV	gag/pro	UCA AAA AAC UUG	3
BLV	gag/pro	UCA AAA AAC UAA	0
HTLV-1	gag/pro	CCA AAA AAC UCC	18
HTLV-2	gag/pro	GGA AAA AAC UCC	18
EIAV	gag/pro	CCA AAA AAC GGG	195
BLV	pro/pol	CCU UUA AAC UAG	0
HTLV-1	pro/pol	CCU UUA AAC CAG	156
HTLV-2	pro/pol	CCU UUA AAC CUG	18
SRV-I	gag/pro	CAG GGA AAC GAC	147
MPMV	gag/pro	CAG GGA AAC GGG	147
Visna	gag/pol	CAG GGA AAC AAC	45

Table 1.

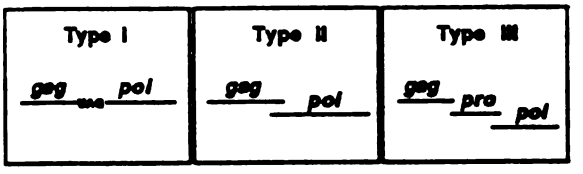


Figure 1.

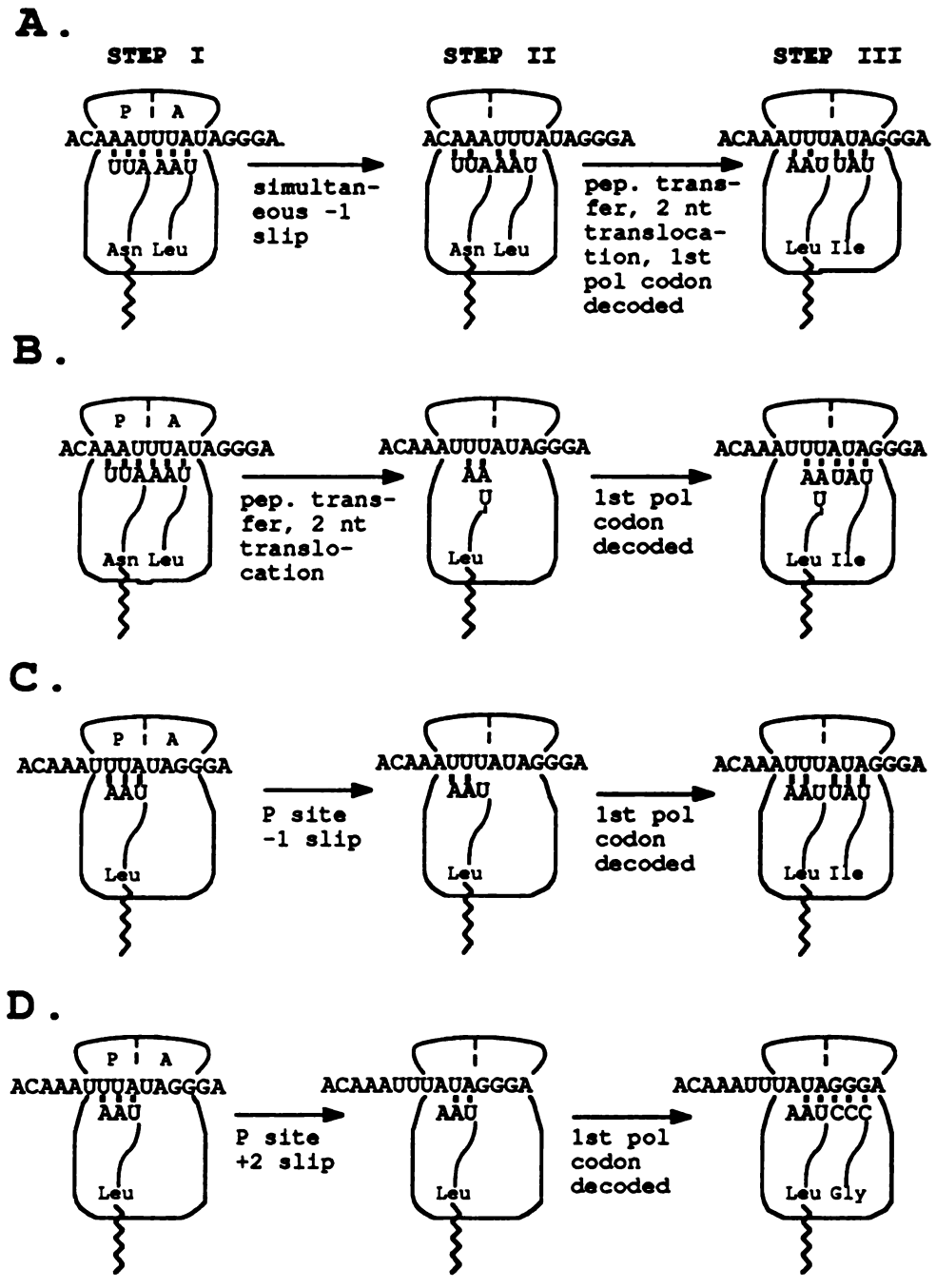
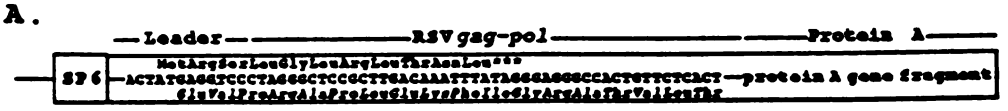

A.
STEP I
STEP II
STEP III
P A
ACAAAUUUAUAGGGA
UUAAAU
Asn Leu
simultaneous -1 slip
ACAAAUUUAUAGGGA
UUAAAU
Asn Leu
pep. transfer, 2 nt translocation, 1st pol codon decoded
ACAAAUUUAUAGGGA
AAUUAU
Leu Ile
B.
P A
ACAAAUUUAUAGGGA
UUAAAU
Asn Leu
pep. transfer, 2 nt translocation
ACAAAUUUAUAGGGA
AA
U
Leu
1st pol codon decoded
ACAAAUUUAUAGGGA
AAUAU
U
Leu Ile
C.
P A
ACAAAUUUAUAGGGA
AAU
Leu
P site -1 slip
ACAAAUUUAUAGGGA
AAU
Leu
1st pol codon decoded
ACAAAUUUAUAGGGA
AAUUAU
Leu Ile
D.
P A
ACAAAUUUAUAGGGA
AAU
Leu
P site +2 slip
ACAAAUUUAUAGGGA
AAU
Leu
1st pol codon decoded
ACAAAUUUAUAGGGA
AAUCCC
Leu Gly

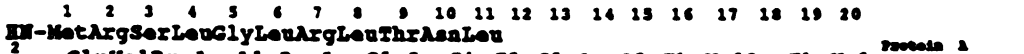
Figure 2.

A.



pGP-S

B.



C.

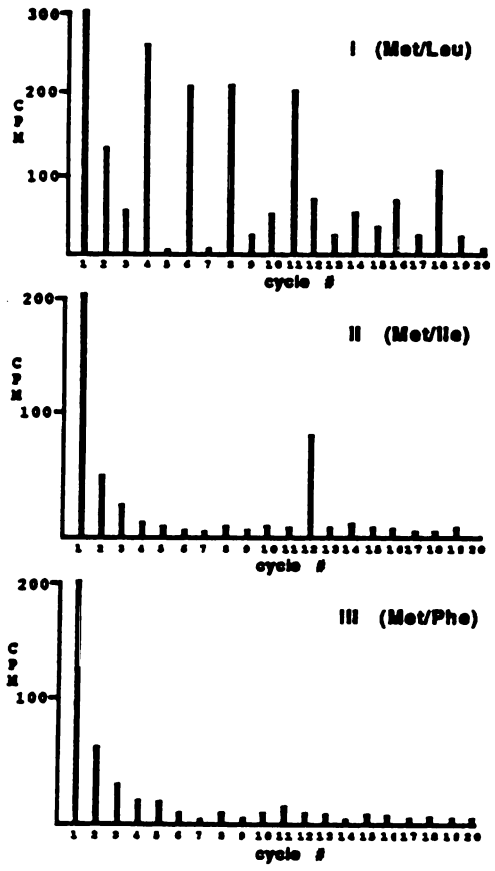
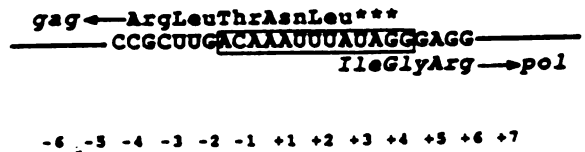
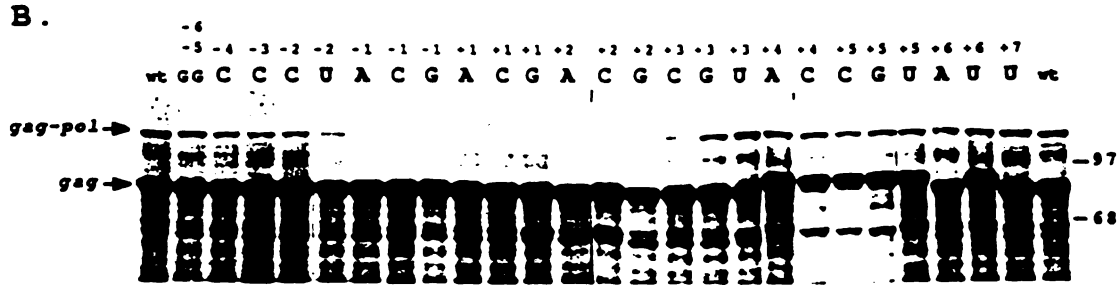


Figure 3.

A.



B.

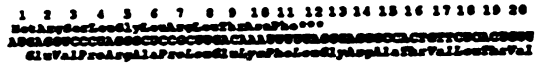


C.

	-6	-5	-4	-3	-2	-1	+1	+2	+3	+4	+5	+6	+7
	A	C	A	A	A	U	U	U	A	U	A	G	G
A						↓	↓	↓		NC		NC	
C			↓	↓	↓	↓	↓	↓	NC	NC	NC		
G	NC	NC				↓	↓	↓	NC		NC		
U					↓				↑		NC	NC	NC

Figure 4.

A.



B.

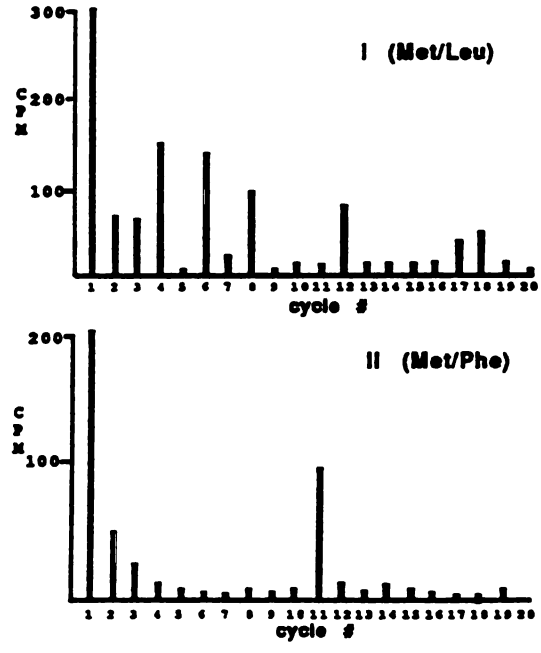

I (Met/Leu)
CPM
300
200
100
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
cycle #
II (Met/Phe)
CPM
200
100
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
cycle #

Figure 5.

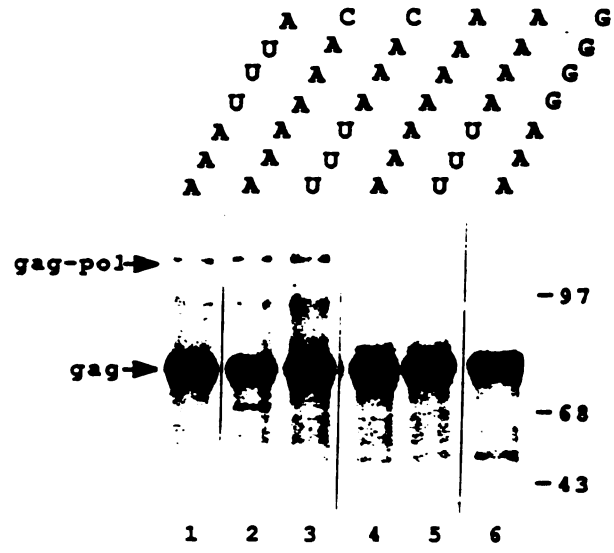

A C C A A G
U A A A A G
U A A A A G
U A A A A G
A A U A U A
A A U A U A
A A U A U A
gag-pol
gag
97
68
43
1 2 3 4 5 6

Figure 6.

Chapter 5

Ribosomal Frameshifting in the Rous Sarcoma Virus *gag-pol* Gene Reguires Downstream RNA Secondary Structure

Summary

The *gag-pol* protein of Rous Sarcoma virus (RSV) is produced via a specific ribosomal frameshift event at the last codon of the *gag* gene. We now describe mutations in RSV *pol*, downstream of the frameshift site, that affect synthesis of the *gag-pol* protein *in vitro*. Mutations that remove or alter part of a potential stem-loop RNA structure adjacent to the frameshift site inhibit frameshifting. Compensatory changes that restore the stem-forming potential of the RNA return frameshifting efficiency to near normal levels. In the presence of the stem-loop structure, ribosomes appear to pause at or near the frameshift site. Finally, a short sequence of RSV RNA, 147 nucleotides in length, containing the frameshift site and stem-loop structure, is sufficient to direct frameshifting in a novel genetic context.

Introduction

The precise maintenance of the reading frame during translation is required for efficient protein synthesis. The estimated frequency of ribosomal frameshifting during translation of typical messenger RNA (mRNA) is approximately to 10^{-5} per codon (Kurland, 1979). There are, however, certain specialized mRNAs on which frameshifting is remarkably efficient, resulting in the synthesis of multiple translation products from single mRNA species. Examples of genes that use programmed ribosomal frameshifts include the *E. coli* release factor II (*RFII*) gene (Craigen et al, 1985), gene 10 of bacteriophage T7 (Dunn and Studier, 1983), the *tya-tyb* gene of the yeast transposable element TY-1 (Clare and Farabough, 1985; Mellor et al., 1985), and the *F1-F2* gene of the coronavirus avian infections bronchial virus (IBV) (Brierly et al., 1987).

Retroviruses provide several more examples of ribosomal frameshifting. The *pol* genes of Rous sarcoma virus (RSV) and human immunodeficiency virus type 1 (HIV-1) are translated after a single frameshift in the upstream *gag* gene (Jacks and Varmus, 1985; Jacks et al., 1988a); two successive frameshifts are required for the expression of *pol* in mouse mammary tumor virus (MMTV) RNA (Jacks et al., 1987; Moore et al., 1987); and other retroviruses are presumed to utilize one or two ribosomal frameshifts in the expression of their *pol* genes (see Jacks et al., 1988b).

In an effort to understand the mechanism of frameshifting in retroviral gene expression, we have concentrated on RSV, performing

amino acid sequencing to localize the frameshift site and site-directed and deletion mutagenesis to establish the sequences required for efficient frameshifting. As described in the accompanying paper, frameshifting in the RSV *gag-pol* overlap occurs at the last codon at *gag* by the slippage of the ultimate and penultimate tRNAs from the 0-frame into the -1 frame along short, homopolymeric sequences. Similar sequences have been shown or are expected to function analogously during frameshifting on other retroviral RNAs (Jacks et al., 1988b).

Not surprisingly, these short homopolymeric sequences are insufficient to induce ribosomes to change frame in a heterologous genetic context (Jacks et al. 1987). This finding led to the hypothesis that potential stem-loop structures positioned just downstream of all retrovirus frameshift sites are a second necessary element in the frameshifting process (Jacks et al. 1987, 1988a). We now demonstrate by deletion and site-directed mutagenesis of the 5′ region of the RSV *pol* gene that a stem-loop structure is required for efficient frameshifting *in vitro* for this virus. A 147 nucleotide sequence containing the RSV frameshift site and stem-loop is sufficient to cause efficient ribosomal frameshifting when placed in a heterologous context. Finally, translational time course experiments indicate that the stem-loop structure causes ribosomes to pause at or near the frameshift site, suggesting a biochemical role for RNA secondary structure in this process.

Results

Figure 1 shows a collection of potential stem-loop structures located 3' to the documented or suspected frameshift sites of several retroviruses. Although very different in primary sequence and somewhat variable in length, the stems are all G-C rich. The predicted ΔG values range from -16.9 K.Cal/mole for the weakest (MMTV *gag/pro*) and EIAV *gag-pol* to -35 K.Cal/mole for the strongest (RSV *gag/pol*). The structures also differ from one another in the presence of extra-helical nucleotides in the stem, the size of the loops, and slightly in the distance from the base of the predicted stem from the frameshift site, although none is further than eight nucleotides away.

Deletion analysis of the RSV stem-loop

To begin to assess whether these stem-loop structures are relevant to frameshifting, we constructed a series of plasmids harboring progressive truncations of the RSV *pol* gene. The mutants contain, downstream of the SP6 promoter, the entire RSV *gag* gene and variable amounts of RSV *pol*, followed by a constant segment derived from the 3' end of human immunodeficiency virus *pol* gene. (The HIV segment encodes an antigen for immunoprecipitation of the protein products.) The positions of the deletion endpoints and their designations are shown with respect to a detailed model of the RSV stem-loop in Figure 2A. Frameshifting efficiency was assayed by the ability of RNA synthesized *in vitro* from the mutant DNAs to yield a *gag-pol* (actually *gag-pol-HIV-pol*) fusion protein upon translation in a rabbit reticulocyte lysate system.

With one exception, those mutations that leave the predicted stem-loop structure intact produce wild-type levels of the *gag-pol* protein (Fig. 2B lanes 1-4). (The efficiency of frameshifting on a wild-type RSV RNA is approximately 5%.) Conversely, mutations that partially or completely disrupt the structure have much reduced levels of frameshifting (Fig. 2B, lanes 6-8). Unexpectedly, the mutant E, whose endpoint is the very last nucleotide of the predicted stem-loop structure, also shows reduced levels of frameshifting (Fig. 2B, lane 5). We will discuss this result in detail below.

A 147 nt RSV fragment is sufficient to cause frameshifting in a novel context

We next utilized these deletion mutations to determine the minimum-sized RSV RNA sequence sufficient to allow frameshifting in a novel genetic context. We replaced all but the last 11 codons at the *gag* gene in each of the original mutants with a portion of the ground squirrel hepatitis B virus surface antigen gene (GS-sAg) such that the only RSV sequences in the resulting plasmids extend from just upstream of the frameshift site (located at the end of the *gag* gene; see accompanying paper) to the deletion end points (Fig. 3A).

As shown in Figure 3B, GS RNAs containing the four longest RSV inserts yield significant amounts of the transframe protein (the product of frameshifting) upon in vitro translation (lanes 1-4). The efficiency of frameshifting is approximately 5%, similar to that obtained with wild-type RSV RNA. The shortest fully functional RSV sequence, present in the GS-D derivative, is 147 nucleotides (lane 4). It is likely that

the minimally-sized RSV sequence capable of conferring frameshifting ability is shorter than this: the first 26 nt of the RSV sequences lie upstream of the frameshift site and are presumably dispensable (see accompanying paper); also, the 3′ boundary for sufficiency probably lies between the endpoints of the fully functional D mutant and the defective E mutant (a distance of 23 nt).

As in the initial deletion analysis, the E mutation, which removes sequences up to the base of the predicted stem, has greatly reduced frameshifting with the GS-sAg gene segment in place of the RSV gene *gag* (Fig. 3B, lane 5). The final three GS derivatives carry still fewer of the RSV stem-loop nucleotides and make even less or no transframe protein (Fig. 3B, lanes 6-8).

Disrupting base pairs within the stem reduces frameshifting efficiency; restoring base pairs rescues it

The results presented above suggest that the RSV stem-loop is required for efficient frameshifting but show definitively only that certain sequences within *pol* are important in this event. The fact that these sequences can potentially form part of a stem-loop structure may be coincidental. Therefore, to directly test whether the stem-loop structure itself, and not merely its primary sequence, influences frameshifting, we investigated the effects of specific stem destabilizing and restabilizing mutations.

Beginning with a plasmid carrying the wild-type RSV *gag* gene and a portion of the RSV *pol* gene, we constructed two site-directed mutations that each disrupt the same five consecutive base pairs (located in the

center of the predicted stem) by converting to their complements the five relevant nucleotides in the 5′ arm (pSM1) or 3′ arm (pSM2) of the stem (Fig. 4A). These mutations should severely destabilize the stem structure. We also combined the two mutations in the plasmid pSM1+2; SM1+2 RNA should form a stem-loop structure similar in thermal stability to that of wild-type RSV RNA but different from wild-type in ten nucleotide positions in the central portion of the stem.

As shown in Figure 4B, frameshifting on an RSV RNA correlates with the presence of a stem-loop structure. The frameshifting efficiency of the SM-1 and SM-2 mutants is reduced greater than ten fold as compared to the wild-type level (lanes 1-3). When the two mutations are present together in the same RNA, restoring the potential for base pairing in the stem, the frameshift efficiency returns to approximately 2.5%, one half the wild-type value (Fig. 4B, lane 4).

The stem-loop structure causes ribosomal pausing near the frameshift site

The mechanism for frameshifting in retroviruses emerging from work presented in the accompanying paper calls for the simultaneous back-slippage of the two tRNAs complexed with the ribosome at the frameshift site. If the efficiency of frameshifting is limited by the frequency of slippage, the role of the stem-loop structure could be to impede the movement of ribosomes through the region near the frameshift site, prolonging the time spent at the relevant codon pair and allowing a greater percentage of the frameshift-mediating tRNAs to slip into the -1 frame. We tested this possibility by performing translational time

course experiments on a variety of RSV RNA derivatives.

In order to assay ribosome pausing on RSV RNA, we made two necessary changes in *gag* (Fig. 5A). We used an internally deleted *gag* gene (the ΔB deletion; Fig. 5A) to improve the degree of synchrony between ribosomes on different RNA m olecules in the same translation reaction. Also, the *gag* termination codon, normally present immediately 3′ to the frameshift site, was changed to a UAU-tryptophan codon. This +6U mutation does not affect the position or efficiency of frameshifting (see accompanying paper); this mutation is necessary because the wild type terminator is located at the base of the predicted stem-loop structure (Fig. 2A). Thus, on ΔB RNA the product of pausing at the stem-loop or (pause product) would be the same size as mature *gag* protein. (26 kD; Fig. 5A). The + 6U mutation extends the *gag* open reading frame by 111 nt. and as such +6UΔB RNA encodes a 30 kD *gag* protein (Fig. 5A. The position of the stem-loop does not change in +6UΔB RNA, however, the pause product generated of this RNA should remain 26 kD and co-migrate with that from ΔB RNA (Fig. 5A). In fact, the 26 kD *gag* protein encoded by ΔB RNA can serve as a marker for the pause product. During *in vitro* translation of +6UΔB RNA, low-level internal initiation at an AUG codon located 30 codons downstream of the true *gag* initiator gives rise to a 27 kD *gag* protein. This protein can be distinguished from the pause product both by size and kinetics of appearance.

Two minutes after addition of +6UΔB RNA to a standard rabbit reticulocyte lysate translation system, further initiation was inhibited with edeine (Garcia et al., 1988) and the progress of the elongating

ribosomes assayed at increasing times of incubation by immunoprecipitation of the nascent or completed polypeptide chains. A significant pause during translation will result in the appearance of a distinct, but transient protein intermediate that gives way to larger products at later time points. In addition to these properties, the pause product of interest co-migrate with the *gag* protein encoded by ΔB RNA that is terminated at the wild-type stop codon (Fig. 5A).

As shown in Figure 5B (panel I), among numerous lower molecular weight intermediates, a protein indicative of pausing at the frameshift site (p) appears at approximately two minutes after edeine addition; it peaks in abundance at approximately three minutes; and then progressively disappears at later times. The full-length (g) and internally initiated (i) *gag* proteins arise with slightly slower kinetics. These proteins then persist through the remainder of the time course. The *gag-pol* (gp) protein appears at approximately four minutes. It too persists thereafter.

To show that the observed ribosomal pausing near the frameshift site is due to the presence of the stem-loop structure, we performed a time course on an RNA containing the stem mutation, SM1 (see Fig. 4A). The SM1 mutation inhibits frameshifting (Fig. 4b) and, as shown in Figure 5B (panel II) also greatly reduces either the duration of the ribosomal pause at the frameshift site or the number of ribosomes that pause there. The relevant intermediate is approximately five- to ten-fold less abundant when the stem is disrupted. The distribution and intensity of the remaining intermediates and completed proteins is

similar between the two experiments, save the anticipated absence of the *gag-pol* protein in the +6UΔB(SM1) RNA translation. Translation of an RNA containing the stem-restoring mutation, +6UΔB(SM1+2), once again includes a significant pause near the frameshift site (Fig 5B, panel III). The relative intensity of this intermediate is approximately one half that of +6UΔB-encoded pause product, consistent with the two-fold reduction in the frameshifting efficiency on +6UΔB(SM1+2) RNA (Fig. 5B). (This drop in frameshifting efficiency was also observed on SM1+2 RNA (Fig. 4B).)

While the existence of the pause product correlates with the presence of an intact stem-loop structure, it also correlates with the frameshifting efficiency on the respective RNAs. Therefore, to show that this phenomenon is related to the stem-loop and not to an unrelated feature of the frameshifting process, we performed a time course experiment on an RNA containing the wild-type stem-loop but with a mutation in the frameshift site previously shown to prevent frameshifting. Except that the *gag-pol* protein is absent, the products of the translational time course of +1C+6UΔB RNA are indistinguishable from those of +6UΔB RNA (Fig. 5B, panels I and IV). Thus, pausing at the frameshift site is not dependent on active frameshifting, but rather on the downstream RNA secondary structure.

Discussion

The ribosomal frameshifting at the end of the RSV *gag* gene to allow synthesis of the *gag-pol* fusion protein is dependent on sequences at the frameshift site as well as an RNA secondary structure located just 3' to this site. The importance of the stem-loop structure is illustrated by the inhibition of frameshifting caused by stem-disrupting deletion and site-directed mutations and, most convincingly, by the recovery of high level frameshifting when two complementary mutations, which are separately deleterious, are combined in the same mRNA.

A role for the stem-loop: a translational barrier

The exact function of the stem-loop structure is yet to be determined, but one effect of it appears to be the slowing of ribosomes as they pass through the frameshift site. The presence of a distinct but transient nascent polypeptide of the appropriate size during translation of stem-containing RNAs and its relative scarcity upon translation of a stem-disrupted message suggest that this RNA structure inhibits ribosomal transit at or near the codons at which the frameshift occurs. Therefore, one role of the stem-loop structure might be to provide ribosomes or, more appropriately, their resident tRNAs increased time to realign with the mRNA at the frameshift site.

It is not yet possible to determine what fraction of ribosomes pause at the RSV stem-loop, or exactly how long those that do pause remain before continuing translation. Comparing the intensity of the pause product to that of the *gag-pol* protein in the time course, it is clear that not every ribosome that pauses goes on to synthesize the *gag-*

pol protein. Whether the remainder of the paused ribosomes actually do shift frame but return to the 0-frame before moving on is not known. From our current experiments, we can only estimate that in this *in vitro* system the duration of the pause at the RSV stem-loop is between 15 and 30 seconds.

The concept of increased "error" with decreased translation rate runs counter to the generally accepted notion that accuracy is sacrificed for increased speed of translation. Several lines of evidence suggest that the need for rapid protein synthesis prevents ribosomes from exercizing their full potential to discriminate between cognate and non-cognate tRNAs (reviewed in Yarus and Thompson, 1984). Reducing the rate of translation with drugs or ribosomal mutations can decrease the frequency of missense errors (Thompson and Karim; Thompson, 1988). However, Yarus and Thompson (1984) have argued that errors requiring kinetically slow reactions might be enhanced if translation were itself slowed.

In *E. coli*, ribosomal frameshifting along several homopolymeric sequences is enhanced by the presence of a 3′ neighboring stop codon (Weiss et al., 1988b). Although the mechanism differs, the concept of a translational pause (during the decoding of the stop codon) potentiating an otherwise unlikely event (the slippage of the tRNA) might pertain. If frameshifting on RSV RNA is dependent on a translational pause, general frame maintenance may normally be achieved, at least in part, by limiting the time for ribosome-bound tRNAs to sample the alternative reading frames. This hypothesis could be tested directly by examining

the effects on frameshifting efficiency of agents that artificially slow translation.

Alternative roles for the stem-loop

The observed ribosomal pause at the RSV stem-loop may merely be a reflection of a more primary role played by the structure in the frameshifting reaction. The pause presumably reflects the ribosome's difficulty in translocating through the stem. Rather than simply delaying ribosome movement to allow tRNAs to re-align with the message, this strained translocation might actually "push" a subset of ribosomes into the -1 frame. If this mechanism were correct, frameshifting efficiency should be very sensitive to mutations that alter the spacing between the frameshift site and stem-loop. In this regard, it is interesting that the proposed stem-loop structures downstream of the various retroviral frameshift sites differ slightly with respect to their distance from those sites (Fig. 1). These computer-folded structures are hypothetical, however, and may not represent that which forms when the ribosome is bound at the frameshift site. In fact, some of the point mutations described in the accompanying paper would disrupt the base of the fully-folded RSV stem-loop (which is closer to its frameshift site than the proposed structures of other retroviral RNAs; Fig. 1) and yet
do not affect frameshift efficiency. Presumably, these lower base pairs do not form when the ribosome is occupying the frameshift site.

Maximal frameshift efficiency on the *E. coli RFII* gene requires an intermolecular interaction between the 3' end of 16S rRNA and a Shine-

Dalgarno-like sequence just upstream of the frameshift site (Weiss et al., 1988a). As proposed above for the RSV stem-loop, Weiss *et al* suggest that this interaction in *RFII* might help force ribosomes into the alternative reading frame, in this case the +1 frame. In this system, a single nucleotide insertion between the Shine-Dalgarno sequence and the frameshift site reduces frameshift efficiency approximately 15-fold (Weiss et al., 1988a).

Additional possible roles for the stem-loop include direct interaction with a ribosomal RNA or protein or the binding of a soluble protein factor. Any of these might influence frameshifting by affecting the ribosome-tRNA-mRNA interaction. We feel such mechanisms unlikely though, since the various stem-loop structures show little primary sequence similarity.

The structure of the RNA

Throughout this text we have referred to the necessary RSV RNA secondary structure as a stem-loop, and it is clear from the effects of mutations and complementary mutations that the proposed major stem is a part of the active structure. There are indications, however, that the structure may be more complex. First, the 65 nucleotides between the two arms of the major stem are predicted to form two additional stem-loop structures (Fig. 2A). Secondly, the E deletion described above, which leaves the proposed structure intact, nevertheless greatly inhibits frameshifting. This result is consistent with an important tertiary interaction, for example, between unpaired nucleotides in the loop and nucleotides downstream of the major stem, a so-called pseudo-

knot structure (Pleij et at., 1985; Puglisi et al., 1988). In fact, there is abundant opportunity for base pairing between sequences in single-stranded regions of the computer-folded stem-loop structure and the 23 nucleotides located between the endpoints of the fully functional D mutant and the defective E mutant (Fig. 2B).

A frameshift cassette

The availability of a short RNA sequence capable of inducing high-level frameshifting is useful for many purposes, including the production of a fixed ratio of two N-terminally related proteins. As shown above, all of the sequences necessary for high-level frameshifting are contained in a 147 nucleotide RSV RNA sequence. We have previously reported production of a transframe protein directed by a 50 nucleodite sequence derived from HIV-1 (Jacks et al., 1988a). But while we have observed frameshifting on these cassettes in two settings, we do not expect them to function equally well in all contexts. At least for RSV, RNA structure is critical for frameshifting, and a perturbation of that structure by new surrounding sequence would be expected to lower frameshift efficiency. In fact, an alternative explanation for the poor efficiency of the E deletion mutant (rather than the tertiary interaction suggested above) is that the novel 3′ sequence abutting the stem destabilizes the structure. In addition, Trifonov (1987) has proposed a modelwhere by the selection atcodons upstream of a frameshift signal would affect frameshift efficiency. More constructions utilizing these and other cassettes are needed to evaluate these factors.

The generality of stem-loop involvement in retroviral frameshifting

All demonstrated retroviral frameshift sites are followed closely by stem-loop structures (Fig. 1 and T.J. unpublished observations). The nucleotide sequences that comprise the stem-loops are not well conserved between viruses, even closely related viruses such as HIV-1 and HIV-2, suggesting that the RNA structure has been maintained per se. Given the conservation of this RNA structure, the failure of isolated frameshift sites to mediate high level frameshifting (Jacks et al., 1987), and the now documented requirement for a stem-loop in frameshifting on RSV RNA, it is not unreasonable to assume that all retroviral frameshift events will require downstream stem-loop structures.

We have recent evidence that partially contradicts this view, however. In a series of constructions in which the HIV-1 *gag-pol* sequences downstream of the frameshift site were replaced by heterologous sequences, we observed see variable effects on frameshifting efficiency (Madhani et al., 1988). In one case the efficiency was reduced approximately ten fold, while in others (including specific stem destabilizing mutations) it was not significantly different from that determined for wild-type HIV-1 RNA (Madhani et al., 1988). These results demonstrate that sequences downstream of the HIV-1 frameshift site can influence frameshifting efficiency but also that high level frameshifting can occur at this site, at least *in vitro*, in the absence of an obvious downstream stem-loop structure. We have not yet tested these RNAs *in vivo* to see if an influence of the stem-loop can be observed in a more natural setting. A second caveat in these experiments is that the stem-removing mutations

were made in an heterologous context, with a foreign gene in place of HIV-1 *pol*. It is possible that, in addition to the possible negative context effects discussed above, novel 3′ sequence could functionally substitute for an absent stem-loop. In the future, the importance of the stem-loop of HIV-1 and other retroviral RNAs will be assessed by mutations in an otherwise wild-type context as we have done here for RSV.

Experimental Procedures

Construction of deletion mutants

The plasmid pGP (Jacks and Varmus, 1985) was first digested with Hind III (position 2740 in the sequence at Schwartz et al. (1983)) and treated with Bal 31 (IBI) according to the specifications the manufacturer. The ends of the DNA were then blunted with T4 DNA polymerase (Boeringer) and Kpn I linkers (Collaborative Research) added using T4 DNA ligase (IBI). After exhaustive digestion with Asp718 (a isoscizomer of KpnI) and PvuI (which cuts in the vector sequence), the resulting fragments were ligated to complementary fragments from the plasmid pAGP (Jacks et al, 1988) previously digested with Asp718 and PvuI. (The Asp718 site in pAGP is in the 3' end of the HIV-1 *pol* gene and corresponds to position 3707 in the sequence of Power et al. (1985).) The resulting plasmids were sequenced by the method of Chen and Seeburg (1985) using a primer complementary the HIV-1 *pol* sequences. In all but two of the deletion mutants tested, the RSV *pol* and HIV *pol* sequences were in frame. For mutants B and E the frame had to be corrected by digesting the plasmids with Asp718 and filling in the 5' overhang using the Klenow fragment of DNA polymerase I (New England Biolabs).

A second set of deletion mutants were constructed by replacing the RSV sequences upstream of the PstI site located near the end of RSV *gag* (position 2450 in Schwartz et al. (1983) with sequences from the 5' end of the ground squirrel hepatitis B virus surface antigen gene (GS-sAg). The original truncation plasmids were cleaved with PvuI (which cuts in

the vector) and PstI and were ligated to a complementary PvuI-PstI fragment from an SP6-promoter-containing plasmid (Melton et al, 1984) carrying the complete GS-sAg gene. The PstI site in GS-sAg corresponds to position 1518 in Seeger et al. (1984). SP6 transcriptions, rabbit reticulocyte translations, and immunoprecipitations were carried out as described (Jacks and Varmus, 1985).

Site-directed mutagenesis

The mutations SM1 and SM2 were constructed in the plasmid pGP (Jacks and Varmus, 1985) by the method described in the accompanying paper using mutagenic oligonucleotides 34 residues in length. pSM1 was then used as the substrate for mutagenesis using the SM2 oligonucleotide to construct pSM1+2.

The plasmids used in the translational time course experiments were constructed by first introducing the +6U mutation (that converts the *gag* terminator to a sense codon; see accompanying paper) into either pGP, pSM1, pSM1+2, or p+1C (+1C is a mutation in the frameshift site; see accompanying paper). Fragments from these mutants were then subcloned into the plasmid pGP-ΔB (Jacks and Varmus, 1985) such that the resulting plasmids contained a deleted *gag* gene and the mutation in the *gag* terminator (+6UΔB) plus either a frameshift site mutation (+1C+6UΔB) or a stem-loop mutation (+6UΔB(SM1) and +6UΔB(SM1+2)). All mutations were verified by DNA sequencing using the method of Chen and Seeburg (1985).

Translational time course experiments

150μl translation reactions supplemented with ^{35}S-methionine and RNA transcribed from one of the plasmids described above were incubated for

two minutes at 30°C. Edeine was then added to a final concentration of 5 μM to inhibit initiation (Garcia et al., 1988) and 8 μl aliquots were removed at increasing times. Aliquots were diluted in 50 μl RIPA buffer (10 μM Tris-HCl (pH 8.15), 5% Na deoxycholate (.5% NP40, .1% SDS, and 200 mM NaCl) and immunoprecipitated with 0.5 μl rabbit anti-p19^{gag} serum by incubating at 4°C for 1 hour. Immunoprecipitates were collected by adding 15 μl Pansorbin (Amersham) and rocking at 4°C for 1 hour. Pellets were washed twice with RIPA buffer and once with 50 mM Tris-HCl, pH 6.8. The antigen was released and any peptidyl-tRNA hydrolized upon boiling in Laemli sample buffer at pH 10.

Acknowledgements

We thank Titia de Lange, Christine Guthrie, and Peter Pryciak for helpful discussion; Volker Vogt for providing the α p19^{gag} antibodies used in these studies; Robin Colgrove for help with computer programs; Peter Walter and Pablo Garcia for providing edeine and useful advise; and Janine Marinos for typing the manuscript. T.J. is a UCSF Graduate Dean's Fellow. H.D.M. is a trainee of the Medical Scientist Training Program. H.E.V is an American Cancer Society Professor of Molecular Virology.

References

Brierly, I., Boursnell, M., Birns, M., Bilmoria, B., Block, V., Brown, T. and Inglis, S. (1987). An efficient ribosomal frameshifting signal in the polymerase-encoding region of the coronavirus IBV. EMBO J. *6*:3779-3785.

Chakrabarti, L., Guyader, M., Alizon, M., Daniel, M.D., Desrosiers, R.C., Tiollais, P. and Sonigo, P. (1987). Sequence of simian immunodeficiency virus from macaque and its relationship to other human and simian retroviruses. Nature *328*, 543-547.

Chen, E.Y. and Seeburg, P.H. (1985). Supercoil sequencing: a fast and simple method for sequencing plasmid DNA. DNA 4, 165-170.

Clare J. and Farabaugh, P. (1985). Nucleotide sequence of a yeast Ty element: evidence for a unusual mechanism of gene expression. Proc. Natl. Acad. Sci. USA *82*, 2829-2833.

Craigen, W.J., Cook, R.G., Tate, W.P. and Caskey, L.T. (1985). Bacterial peptide chain release factors: Conserved primary structure and possible frameshift regulation of release factor 2. Proc. Natl. Acad. Sci. USA *82*, 3616-3620.

Dunn, J.J. and Studier, F.W. (1983). Complete nucleotide sequence of

bacteriophage T7 DNA and the locations of T7 genetic elements. J. Mol. Biol. *166*, 477-535.

Fox, T.D. and Weiss-Brummer, B. (1980). Leaky +1 and -1 frameshift mutations at the same site in a yeast mitochondrial gene. Nature *288*, 60-63.

Garcia, P.O., Ou, J.H., Rutter, W.H. and Walter, P. (1988). Targeting of the hepatitis B virus precore protein to the endoplasmic reticulum membrane: after signal cleavage translocation can be aborted and the products released into the cytoplasm. J. Cell. Biol., in press.

Guyader, M., Emerman, M., Sonigo, P., Claver, F., Montagnier, L. and Alizon, M. (1987). Genome organization and transcription of the human immunodeficiency virus type 2. Nature *326*, 662-669.

Hiramatsu, K., Nishida, J., Naito, A. and Yoshikura, H. (1987). Molecular cloning of the closed circular provirus of human T cell leukemia virus type I: A new open reading frame in the *gag-pol* region. J. Gen. Virol. *68*, 213-218.

Jacks, T., Power, M.D., Masiarz, F.R., Luciw, P.A., Barr, P.J. and Varmus, H.E. (1988a). Characterization of ribosomal mutations in HIV-1 *gag-pol* expression. Nature *331*, 280-283.

Jacks, T., Masiarz, F.R. and Varmus, H.E. (1988b). A detailed analysis of the ribosomal frameshift site in the Rous sarcoma virus *gag-pol* gene. Cell, submitted.

Jacks, T., Townsley, K., Varmus, H.E. and Majors, J. (1987). Two efficient ribosomal frameshift events are required for synthesis of mouse mammary tumor virus *gag*-related polypeptides. Proc. Natl. Acad. Sci. USA *84*, 4298-4302.

Jacks, T. and Varmus, H.E. (1985). Expression of the Rous sarcoma virus *pol* gene by ribosomal frameshifting. Science *230*, 1237-1242.

Kurland, C.G. (1979). In *Nonsense Mutations and tRNA suppressors* (Celis, J.E. and Smith, J.D., eds.), pp. 98-108. Academic Press, New York.

Madhani, H.D., Jacks, T. and Varmus, H.E. (1988)). Signals for the expression of the HIV pol gene by ribosomal frameshifting. In *The Control of HIV Gene Expression* (Franza, R., Cullen, B. and Wong-Staal, F., eds.). Cold Spring Harbor Laboratory, Cold Spring Harbor, in press.

Mellor, J., Fulton, S.M., Dobson, M.J., Wilson, W., Kingsman, S.M. and Kingsman, A.J. (1985). A retrovirus like strategy for expression of a fusion protein encoded by the yeast transposon Tyl. Nature *313*, 243-246.

Melton, D.A., Krieg, P.A., Rebagliati, M.R., Maniatis, T., Zinn, K. and Green, M.R. (1984). Efficient in vitro synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. Nucl. Acids Res. *12*, 7035-7056.

Moore, R., Dixon, M., Smith, R., Peters, G. and Dickson, C. (1987). Complete nucleotide sequence of a milk-transmitted mouse mammary tumor virus: Two frameshift suppression events are required for translation of *gag* and *pol*. J. Virol. *61*, 480-490.

Pleij, C.W.A., Rietveld, K. and Bosch, L. (1985). A new principle of RNA folding based on pseudoknotting. Nucl. Acids. Res. *13*, 1717-1731.

Power, M.D., Marx, P.A., Bryant, M.L., Gardner, M.D., Barr, P.J. and Luciw, P.A. (1986). Nucleotide sequence of SRV-1, a type D simian acquired immune deficiency syndrome retrovirus. Science *231*, 1567-1572.

Puglisi, J.D., Wyatt, J.R. and Tinoco, I. (1988). A pseudoknotted RNA oligonucleotide. Nature *331*, 283-286.

Sagata, N., Yasunaga, T., Tsuzuku-Kawamura, J., Ohishi, K., Ogawa, Y. and Ikawa, Y. (1985). Complete nucleotide sequence of the genome of bovine leukemia virus: Its evolutionary relationship to other retroviruses. Proc. Natl. Acad. Sci. USA *82*, 677-681.

Sanchez-Pescador, R., Power, M.D., Barr, P.J., Steimer, K.S., Stempier, M.M., Brown-Shimer, S.L., Gee, W.W., Renard, A., Randolf, A., Levy, J.A., Dina, D. and Luciw, P.A. (1985). Nucleotide sequence and expression of an AIDS-associated retrovirus (ARV-2). Science *227*, 484-492.

Schwartz, D.E., Tizard, R. and Gilbert, W. (1983). Nucleotide sequence of Rous sarcoma virus. Cell *32*, 853-869.

Seeger, C., Ganem, D. and Varmus, H.E. (1984). Nucleotide sequence of an infectious molecularly cloned genome of ground squirrel hepatitis virus. J. Virol. *51*, 367-375.

Shimotohno, K., Takahashi, Y., Shimizu, N., Gojobori, T., Golde, D.W., Chen, I.S.Y., Miwa, M. and Sugimura, T. (1985). Complete nucleotide sequence of an infectious clone of human T-cell leukemia virus type II: An open reading frame for the protease gene. Proc. Natl. Acad. Sci. USA *82*, 3101-3105.

Sonigo, P., Alizon, M., Staskus, K., Klatzmann, D., Cole, S., Danos, O., Retzel, E., Tiollais, P., Haase, A. and Wain-Hobson, S. (1985). Nucleotide sequence of the visna lentivirus relationship with the AIDS virus. Cell *42*, 369-382.

Stephens, R.M., Casey, J.W. and Rice, N.R. (1986). Equine infectious

anemia virus *gag* and *pol* genes: Relatedness to visna and AIDS virus. Science *231*, 589-594.

Thompson, R.C. and Karim, A.M. (1982). The accuracy of protein biosynthesis is limited by its speed: high fidelity selection by ribosomes of aminoacyl-tRNA ternary complexes containing GTP [γS]. Proc. Natl. Acad. Sci. USA *79*, 4922-4926.

Thompson, R.C. (1988). Ef-Tu provides an internal kinetics standard for translational accuracy. Trends Biochem. Sci., in press.

Tinoco, I., Borel, P., Dengler, B., Levine, M., Uhlenbeck, O., Crothers, D.M. and Gralla, J. (1973). Improved estimation of secondary structure in ribonucleic acids. Nature New Biol. *246*, 40-41.

Trifonov, E.N. (1987). Translation framing code and frame-monitoring mechanism as suggested by the analysis of mRNA and rRNA nucleotide sequences. J. Mol. Biol. *194*, 643-652.

Weiss, R.B., Dunn, D.M., Dahlberg, A.E., Atkins, J.F. and Gesteland, R.F. (1988a). Reading frame switch caused by base-pair formation between the 3′ end of 16S rRNA and the mRNA durign elongation of protein synthesis in *Escherichia coli*. EMBO J., in press.

Weiss, R.B., Dunn, D.M., Atkins, J.F. and Gesteland, R.F. (1988b).

Slippery runs, shifty stops, backward steps and forward hops: -2, -1, +5 and +6 ribosomal frameshifting. Cold Spring Harbor Symp. Quant. Biol. *52*, in press.

Yarus, M. and Thomson, R.C. (1983). Precision of protein biosynthesis. In *Gene Function in Prokaryotes*, (Beckwith, J.R., Davies, J.E. and Gallat, J.A., eds.), pp. 23-63. Cold Spring Harbor Laboratory, Cold Spring Harbor.

Yoshinaka, Y., Katoh, I., Copeland, T.D. and Oroszlan, S.J. (1985a). A murine leukemia virus protease is encoded by the *gag-pol* gene and is synthesized through suppression of an amber termination codon. Proc. Natl. Acad. Sci. USA *82*, 1618-1622.

Yoshinaka, Y., Katoh, I., Copeland, T.D. and Oroszlan, S.J. (1985b). Translational readthrough of an amber termination codon during synthesis of feline leukemia virus protease. J. Virol. *55*, 870-873.

Figure Legends

Figure 1. Potential stem-loop structures located downstream of assorted retroviral frameshift sites.

Predicted stem-loop structure of RNA sequences downstream of several documented or suspected retroviral frameshift sites were determined with the aid of the Hairpins program (Hugo Martinez, University of California, San Francisco). The predicted free energy valves (ΔG) of the structures were determined according to rules of Tinoco et al. (1973) and are expressed in Kcal. The heptameric frameshift sites are shown boxed (see Jacks et al., 1988b.) Retroviral sequences are derived from the following: RSV (Schwartz et al., 1983), HIV-1 (Sanchez-Pesendor, 1985), HIV-2 (Guyader et al., 1987), MMTV (Jacks et al., 1987), bovine leukemia virus (BLV) (Sagata et al., 1985), (Human T-cell leukemia virus type 2 (HTLV-2) (Shimotohno et al., 1985). Visna virus (Sonigo et al., 1985), and equine infectious anemia virus (EIAV) (Stephens et al., 1986).

Figure 2. The effect of truncation of the RSV *pol* gene on frameshifting efficiency.

A. The endpoints of eight truncations (A-H) in RSV *pol* are indicated by arrows relative to the sequence and predicted RNA secondary structure in the region. The RSV frameshift site is boxed. The sequence presented on the second line is a combination of the *pol* sequence shown above it. In each of these mutants, the RSV sequence is followed by a constant

sequence derived from the HIV-1 *pol* gene (not shown; see Experimental Procedures).

B. Fluorogram of a 10% SDS - polyacrylamide gel containing total ^{35}S-labelled products of rabbit retriculocyte translation of RNAs transcribed from RSV truncation mutants. The names of the mutants (see part A) are shown above the lanes. The position at the *gag* and *gag-pol*-HIV *pol* (referred to as *gag-pol*) fusion proteins and molecular mass markers (in kD) are indicated.

Figure 3. Frameshifting on RSV sequences present in a heterologous context.

A. A portion of the plasmid pGS-A is diagrammed. The 5′ region of the ground squirrel hepatitits virus surface antigen gene (GS-*sAg*) was used to replace all but the final 11 codons of the RSV *gag* gene in the original truncation plasmid A (see Fig. 2, legend). Analogous plasmids, pGS-B through -H, were also constructed. The open reading frames are drawn to indicate that frameshifting in the RSV segment is required for production of a GS-*sAg*-RSV *gag-pol*-HIV *pol* fusion protein. The enlargement of the RSV region shows the relative position of the frameshift site (box) and nucleotides of the predicted major stem (inverted arrows). Details of plasmid constructions are given in Experimental Procedures.

B. Fluorogram of a 12.5% SDS-polyacrylamide gel containing anti-GS-sAg-serum-precipitated products of rabbit reticulocyte lysate translation of GS RNAs. The names of the corresponding RSV truncated mutants are shown

above the lanes (see Fig. 2A). The position of the uniframe protein (the N-terminal GS-sAg protein produced in the absence of frameshifting) and transframe protein (GS-*sAg*-RSV *gag-pol*-HIV *pol* fusion protein product of frameshifting) are shown along with the position of molecular mass standards.

Figure 4. The effect of specific stem destabilizing and restabilizing mutations.

A. The predicted stem-loop structure of wild type RSV RNA and three mutant derivatives. The wild type stem-loop structure shown is a simplification of that shown in Fig. 2A. The mutants SM1 and SM2 have five consecutive bases in the 5′ or 3′ arms of the stem changed to their complements. SM1+2 RNA carries both mutations present in SM1 and SM2 and thus can re-form a stem structure. The mutations were constructed as described in Experimental Procedures.

B. Fluorogram of a 10% SDS-polyacrylamide gel of total ^{35}S-labelled translation products of rabbit reticulocyte translation of wild type RSV (wt), SM1, SM2, or SM1+2 RNAs. The position of the *gag* and *gag-pol* proteins and molecular weight standards (in kD) are indicated.

Figure 5. Translational time course experiments.

A. The RNA diagrammedatleft derivesfrom the RSV ΔB mutant (Jacks and Varmus, 1985). It encodes *gag* (g) and *gag-pol* (gp) proteins of 26 and 34 kD, respectively. The presumed product of pausing at this structure (p) is the same size as the *gag* protein (gp). Initiation at the second

AUG shown in ΔB RNA would generate a 23 kD *gag* product (i). Conversion of the *gag* terminator to a sense codon leads to the +6UΔB RNA shown at right. The p and gp products are unchanged in size, whereas the g and i proteins both increase in size by 4 kD.

B. Fluorograms of 10-15% SDS-polyacrylamide gradient gels containing products of translational time course experiments programmed with +6UΔB (panel I) or three mutant derivatives of it: +6UΔB(SM1) (panel II), +6UΔB(SM1+2) (panel III), or +1C+6UΔB (panel IV) (see text). Rabbit reticulocyte translation reactions programmed by the appropriate RNA were incubated for two minutes at 30°C to allow initiation to occur. Initiation was then inhibited with edeine and aliquots were removed at the times indicated for immunoprecipitation with anti-p19^{gag} antiserum. The immunoprecipitation products were boiled in pH 10 buffer (to hydrolize any peptidy tRNA) before being loaded onto the gel. The lanes marked M contain anti-p19^{gag} - serum - precipitated products of ΔB RNA, which retains the wild-type *gag* terminator, to serve as a marker for the pause product (p). The position of the *gag* protein (g), *gag-pol* protein (gp), and internal initiation product (i) are also indicated (see Fig. 5A). The position of the molecular mass standards are shown in kD.

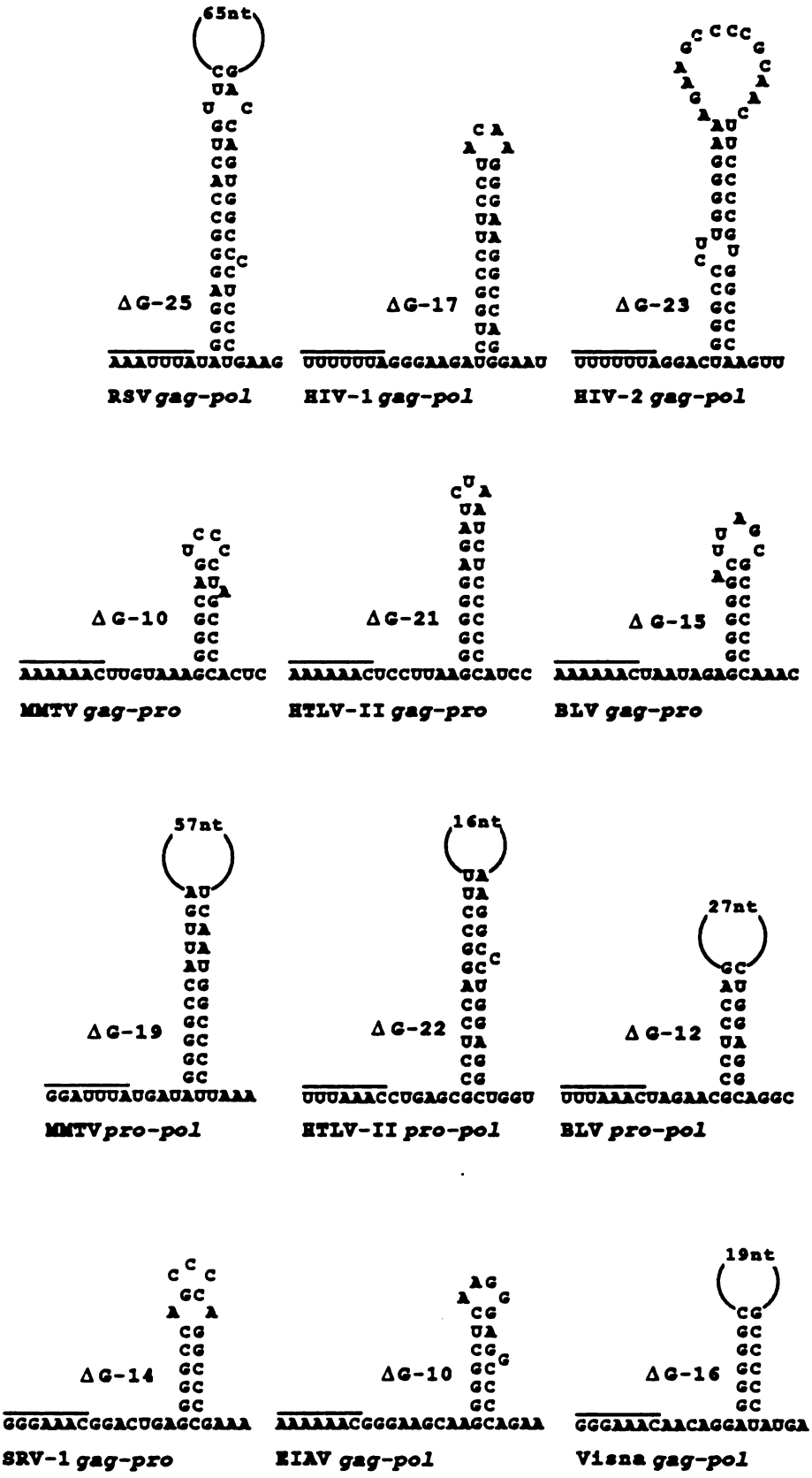

65nt
ΔG-25
RSV gag-pol
ΔG-17
HIV-1 gag-pol
ΔG-23
HIV-2 gag-pol
ΔG-10
MMTV gag-pro
ΔG-21
HTLV-II gag-pro
ΔG-15
BLV gag-pro
57nt
ΔG-19
MMTV pro-pol
16nt
ΔG-22
HTLV-II pro-pol
27nt
ΔG-12
BLV pro-pol
ΔG-14
SRV-1 gag-pro
ΔG-10
EIAV gag-pol
19nt
ΔG-16
Visna gag-pol

Figure 1.

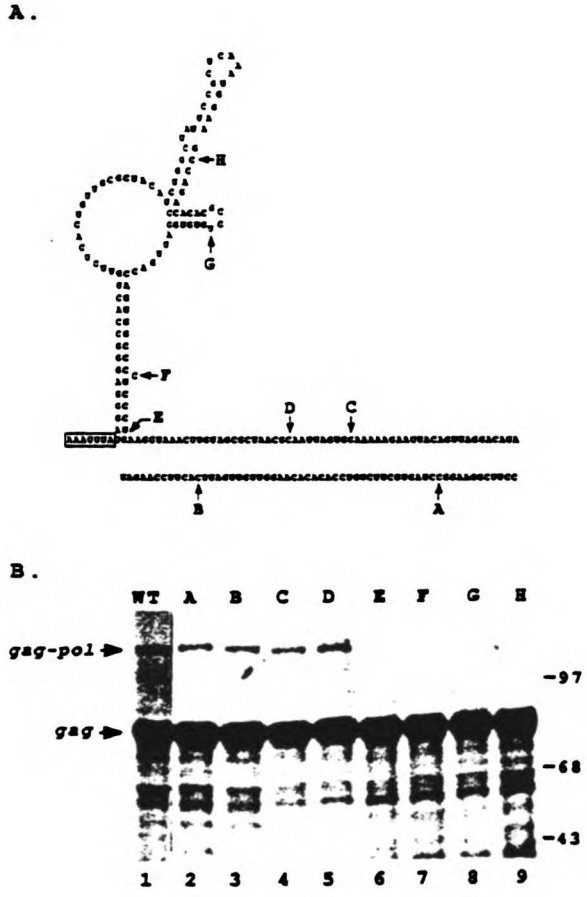


Figure 2.

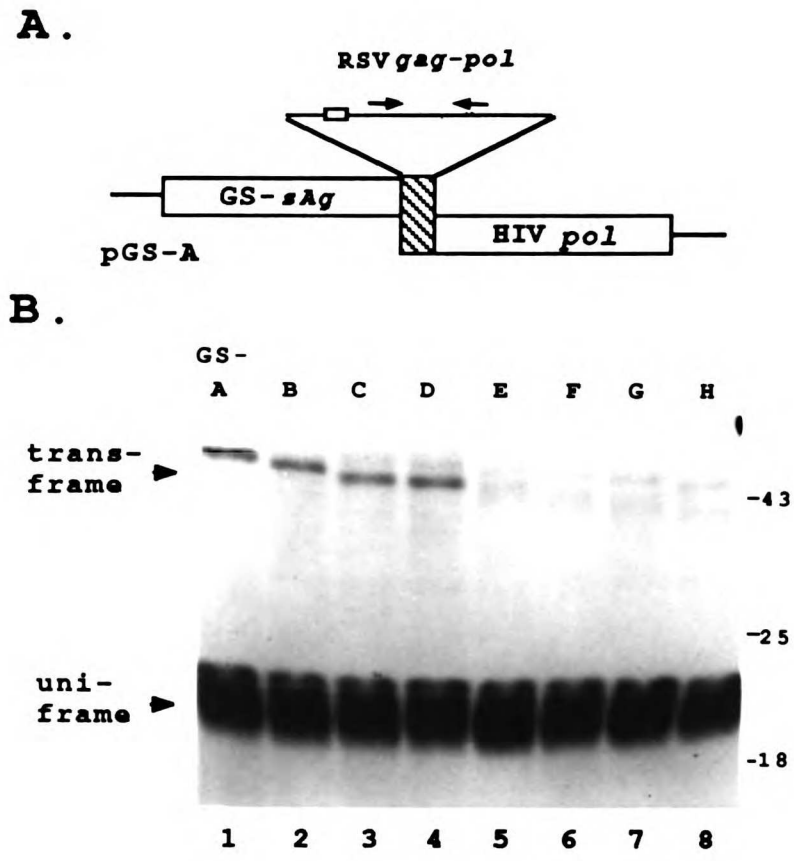
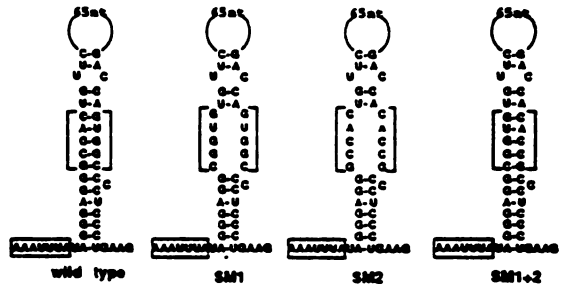

A.
RSV gag-pol
GS-sAg
HIV pol
pGS-A
B.
GS-
A B C D E F G H
trans-
frame
uni-
frame
-43
-25
-18
1 2 3 4 5 6 7 8

Figure 3.

A.



B.

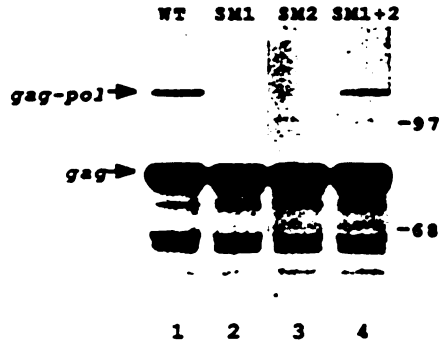


Figure 4.

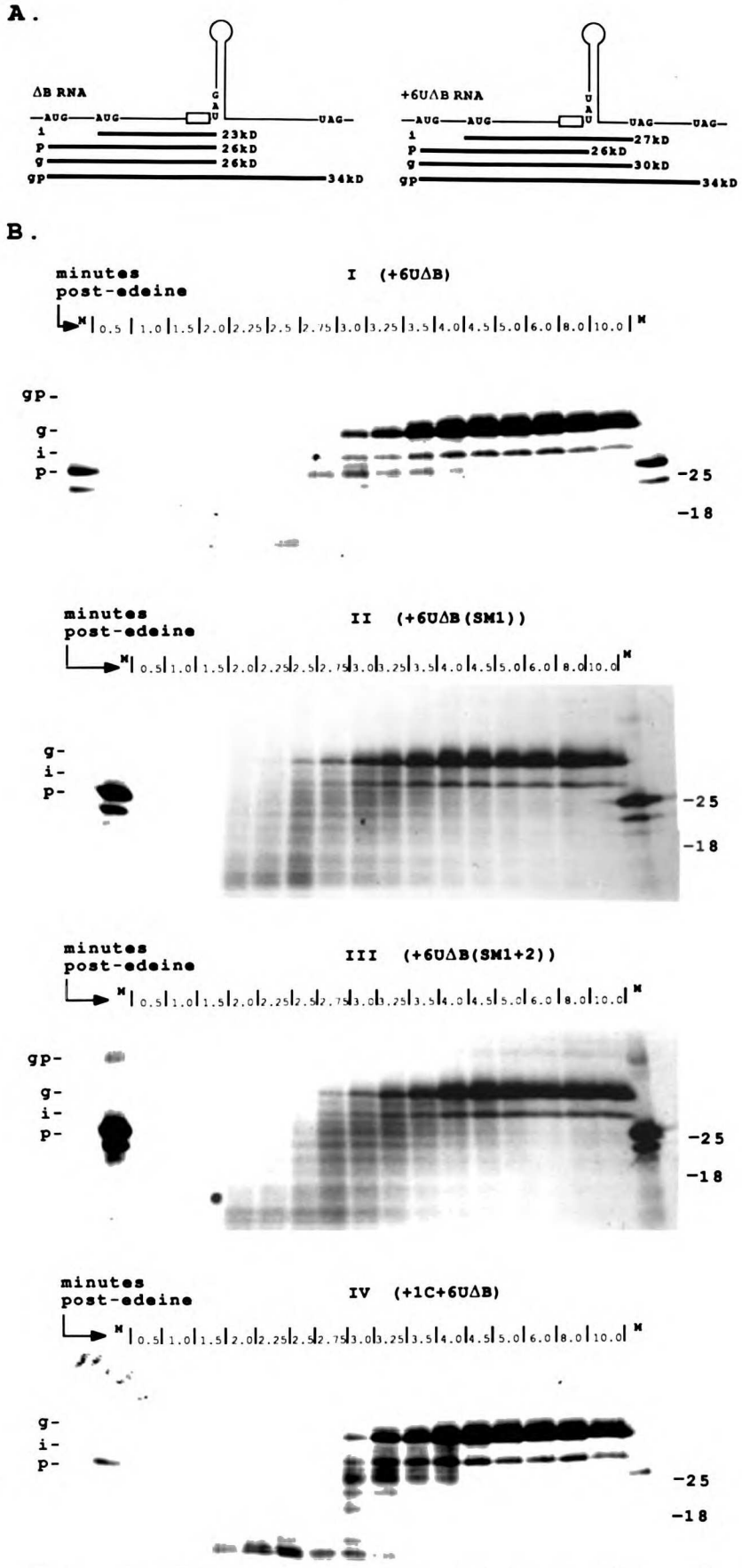

A.
ΔB RNA
—AUG—AUG—
GAU
—UAG—
i 23kD
p 26kD
g 26kD
gp 34kD
+6UΔB RNA
—AUG—AUG—
UAU
—UAG—UAG—
i 27kD
p 26kD
g 30kD
gp 34kD
B.
minutes post-edeine
I (+6UΔB)
M 0.5 1.0 1.5 2.0 2.25 2.5 2.75 3.0 3.25 3.5 4.0 4.5 5.0 6.0 8.0 10.0 M
gp-
g-
i-
p-
-25
-18
minutes post-edeine
II (+6UΔB(SM1))
M 0.5 1.0 1.5 2.0 2.25 2.5 2.75 3.0 3.25 3.5 4.0 4.5 5.0 6.0 8.0 10.0 M
g-
i-
p-
-25
-18
minutes post-edeine
III (+6UΔB(SM1+2))
M 0.5 1.0 1.5 2.0 2.25 2.5 2.75 3.0 3.25 3.5 4.0 4.5 5.0 6.0 8.0 10.0 M
gp-
g-
i-
p-
-25
-18
minutes post-edeine
IV (+1C+6UΔB)
M 0.5 1.0 1.5 2.0 2.25 2.5 2.75 3.0 3.25 3.5 4.0 4.5 5.0 6.0 8.0 10.0 M
g-
i-
p-
-25
-18

Figure 5.

Conclusions

In the Introduction to this thesis, I described the *gag-pol* problem, a longstanding enigma in the field of retrovirology. The work that we and others have done over the last four years has, at least on one level, solved this problem. Retroviruses utilize one of three strategies to express their *pol* genes: stop codon suppression, and single and double ribosomal frameshifting. Before returning to the more molecular aspects of this solution (and some preliminary and planned experiments), I will first offer a rationale for the use of translational control instead of its alternatives.

Teleology. There are many questions to ask at the teleological level, and none of the proposed answers can be definitively substantiated. For example, why do retroviruses synthesize a *gag-pol* (or *gag-pro-pol*) fusion protein instead of an independent *pol* protein? There are three potential explanations here. First, it has long been believed that the *gag-pol* fusion protein is not an active polymerase, and that it requires protolytic release from this precursor to become active. Thus, the N-terminal extension on reverse transcriptase may be a type of molecular silencer designed to keep the enzyme from copying its own mRNA too early or cellular mRNA at all. The second potential reason for synthesizing a *gag-pol* fusion protein is that in this form the *pol* proteins are readily packaged into the maturing virus core particle. Presumably a protein:protein interaction (perhaps coupled with RNA binding activity) between *gag* moieties initiates capsid formation. Since the *pol* proteins are adorned with a *gag* leader, they should be similarly packaged into the capsid. The final speculation is

related to the last. The most important functions carried by the *gag-pol* protein are enzymatic: reverse transcriptase, integrase, and protease. As enzymes, these proteins are required in the virus particle in lesser amounts than the structural components contributed by the *gag* protein. By making *gag-pol* expression dependent on a rare translational event in *gag* retroviruses assure a high *gag* to *gag-pol* ratio. The inclusion of these two proteins in the core particle reflects this ratio. Indeed, there are approximately 1000 of each of the *gag* subunits and about 50 reverse transcriptase molecules in virus particles. The 20:1 *gag*:*gag-pol* ratio is also observed in the cytoplasm of infected cells and in our *in vitro* translations of RSV RNA.

One or a combination of these considerations probably explains why a *gag-pol* protein is preferable to a *pol* protein alone. But why is translational suppression the means for synthesizing the fusion when mRNA splicing is so much more common in eukaryotic cells? There are at least two possible reasons for this. If splicing were the mechanism for *gag-pol* expression and a separate, subtly spliced "*gag-pol*" mRNA were formed, it would be difficult to avoid packaging this species into virions. Bringing the "*gag-pol*" mRNA through the lifecycle would lead to viruses that constitutively produced the *gag-pol* protein. Given the presumed need for a fixed *gag*:*gag-pol* ratio for normal core assembly, these "*gag-pol*" viruses would almost certainly be defective. Again, relying on the presumed need for a specific ratio of *gag* to *gag-pol*, the second argument against a splicing mechanism is the variable efficiency of splicing reactions in different cell lines and species. Whereas the

observed frameshifting efficiency is constant in many cell types from many species (as deduced from *gag* to *gag-pol* ratios in retrovirus-infected cells and in many cell-free systems), the level of splicing of the genome-length mRNA to *env* mRNA is very different in chicken cells versus mammalian cells.

Mechanism. Given that a translational mechanism for *gag-pol* expression is preferable, how does it work? I will only summarize our current conception of frameshifting in retroviral gene expression here, since it is covered quite thoroughly in Chapters 4 and 5. For RSV (and I suspect all retroviruses that utilize this mechanism) frameshifting can be explained in terms of two important features of the mRNA: a site and a structure. The site, the position on the mRNA where ribosomes change reading frame, is relatively simple and, I think, reasonably well understood. The best characterized of these sites, that of RSV, is composed of two short homopolymeric runs of nucleotides (AAA and UUU) that appear to allow two ribosome-bound 0-frame tRNAs to "slip" into the -1 frame. Although they have not been examined experimentally, the demonstrated or suspected frameshift sites of other retroviruses are similarly structured to that of RSV and probably function analogously.

The structure is somewhat more elusive. Again, in the best understood example, RSV, there is little doubt that a stem-loop (or more complicated) structure is necessary for efficient frameshifting to occur. Moreover, in this case, we have observed a biochemical phenomenon associated with the stem-loop, namely, the slowing of translating ribosomes. In light of the proposed mechanism of

frameshifting, the translational pause might allow ribosomes (or their bound tRNAs) increased time to sample the -1 frame. The combination of slippery codons and translational pauses is very reminiscent of slippery stops in E. coli and the mechanism of frameshifting on the *RFII* gene (see Introduction).

While we initially believed (and, in fact, still believe) that frameshifting in other retroviral genes would be dependent on downstream structures, this position has come under some attack of late. Two sets of experiments, one performed by Hiten Madhani, the other by myself, have called into question the simple site-structure model. As described in the discussion to Chapter 5, Hiten has successfully dispensed with the HIV-1 stem-loop without an obvious drop in frameshifting efficiency in vitro. Whether the role of the HIV-1 stem-loop would be more obvious in vivo or in a more natural RNA context remains to be seen.

In some recent unpublished experiments, I attempted to test the site-structure model for the MMTV *gag-pro* frameshift. A small (approximately 50 nucleotide) oligonucleotide cassette carrying the MMTV *gag-pro* frameshift site and stem-loop structure (plus five additional nucleotides downstream of the stem) was inserted between the RSV *gag* and HIV-1 *pol* genes and assayed for frameshifting. (This is, in fact, a direct test of the favored explanation of the experiment described in Figure 5 of Chapter 2.) Much to my surprise, the level of frameshifting in this context was, at most, 1%, approximately 5% of the level observed in its native setting. In a last ditch effort to rescue high-level frameshifting in the artificial construct, I added on first 75 then 150

additional nucleotides of MMTV sequence 3′ to the stem-loop. (Unfortunately, due to the nature of the constructions, both of these plasmids carry a point mutation [A to G] six nucleotides downstream of the stem.) Neither of the mRNAs synthesized from these plasmids showed augmented frameshifting efficiency, despite the fact that in a mRNA derived from the original MMTV test plasmid (described in Chapter 2) having 150 nucleotides of sequence downstream of the *gag-pro* stem-loop is sufficient to direct high-level frameshifting in vitro. We do not yet know whether the base substitution downstream of the stem or the general RNA context is responsible for the low activity of the artificial construct, but we can say with confidence that not every retroviral site and simple stem-loop are sufficient to direct frameshifting in any given context.

Shifty tRNAs. In collaboration with Bernie Dudock of the State University of New York, Stony Brook, we are attempting to isolate and characterize the tRNAs responsible for slippage at certain retroviral frameshift sites. The exact nature of the assay system has not been worked out at this point, but, generally, it will take advantage of an in vitro translation system that lacks the necessary shifty tRNA and consequently fails to produce a transframe protein. The shifty tRNAs will be purified by virtue of their ability to reconstitute efficient frameshifting and transframe protein production in this system. We are excited about the prospects for this experiment for two reasons. First, I was able to show some time ago (unpublished) that a rabbit reticulocyte translation system depleted of its endogenous tRNAs and

supplemented with wheat germ tRNA would make plentiful RSV *gag* protein but very little *gag-pol* protein from an RSV *gag-pol* mRNA. This type of assay system is an ideal starting point for isolating the shifty tRNA(s). Secondly, all retroviruses known to utilize frameshifting have chosen one of three A site codons in their frameshift sites (see Chapter 4). This suggests that the corresponding tRNAs may be unusual in some way that could become apparent upon purification and sequencing.

Affecting frameshifting by mutations and drugs. As argued in the beginning of this section, we believe that the level of frameshifting (or stop codon suppression) observed in retroviral genes might be critical for virus replication in that it determines the ratio of *gag* protein to its C-terminal extensions. Experiments being carried out by Peter Pryciak should address this claim. Peter has constructed approximately twelve mutants of RSV that carry mutations in and around the frameshift site and stem-loop structure. He will be characterizing these mutants for growth in tissue culture cells. Already Peter has evidence that mutations that subtly affect frameshifting efficiency in vitro have a profound influence on virus viability. By combining variant hyper- and hypo-frameshifting mutant viruses in the same cell, Peter may be able to rescue replication in trans. He may also be able to select second-site revertants that restore frameshifting to normal levels and inform us about the mechanism of frameshifting in unexpected ways.

We are engaged in a collaboration with Bob Weiss, John Atkins, and Ray Gesteland at the Howard Hughes Medical Institute in Salt Lake City

to try to find drugs that either inhibit or stimulate frameshifting in retroviral genes. As with Peter's experiments described above, we hope that subtle changes in frameshifting efficiency might inhibit retrovirus replication. The obvious target for these agents is the AIDS virus, HIV-1, which, as described in Chapter 3, utilizes frameshifting to produce its *gag-pol* protein.

Frameshifting in other viruses and cells. The use of an anti-frameshifting drug could be limited if some of our own genes require frameshifting for their expression. As yet only one cellular gene, the *RFII* gene of E. coli, has been shown to require frameshifting. With the help of Robin Colgrove, we have used our knowledge of frameshifting in retroviral gene expression to ask whether any known eukaryotic genes are similarly controlled. Robin wrote a computer program to search nucleic acid sequence data bases for heptameric frameshift sites (see Chapter 4) located in the correct reading frame within coding regions. Those genes that satisfied this criterion (and there were several) were then screened for the presence of significant RNA secondary structure downstream of the putative frameshift site. After much effort, Robin discovered just four genes that we felt were worth investigating; none of them were cellular. Three of the candidate genes are in the alpha viruses, Sindbis virus, Semliki-Forest virus, and Ross river virus, the fourth potential frameshift site is in the genome of tobacco etch virus, a plant potyvirus (see Chapter 4, discussion). As discussed in Chapter 4, there is now one non-retroviral example of frameshifting in higher eukaryotic cells. This occurs in the F1/F2 overlap of the coronavirus

avian infectious anemia virus. We believe more viral examples of frameshifting will follow from directed efforts like ours and also by chance discovery from nucleotide sequencing. As is so often true after the discovery of a novel viral mechanism, the cellular examples are probably not far behind.

RETURN TO the circulation desk of any
University of California Library

or to the

NORTHERN REGIONAL LIBRARY FACILITY
Bldg. 400, Richmond Field Station
University of California
Richmond, CA 94804-4698

ALL BOOKS MAY BE RECALLED AFTER 7 DAYS

- 2-month loans may be renewed by calling (510) 642-6753
- 1-year loans may be recharged by bringing books to NRLF
- Renewals and recharges may be made 4 days prior to due date

DUE AS STAMPED BELOW

DEC 1 8 2004

DD20 15M 4-02

FOR REFERENCE

NOT TO BE TAKEN FROM THE ROOM

BRODART CAT. NO. 23 012 PRINTED IN U.S.A.