10
The Mechanism of Recoding in Pro- and Eukaryotes

Elizabeth S. Poole, Louise L. Major, Andrew G. Cridge, and Warren P. Tate

10.1
Introduction

During protein synthesis, the ribosome uses mechanisms that maintain the translational frame and the nature of the interactions between the RNA participants are critical to this process. Of the RNA visitors to the ribosome, the tRNA occupies two of three possible sites [1] at the active center depending on whether the ribosome is in the pre-translocational or post-translocational state of the polypeptide-chain-elongation cycle [2]. In the pre-translocation state, the tRNAs are at the A-site (as the reaction substrate) and the P-site (as the reaction intermediate). In the post-translocation state, the tRNAs occupy the P-site and the E-site (as the reaction product). These pairs of tRNAs make unique interactions with the host structural rRNAs in their particular environments, and the tRNAs themselves undergo some conformational flexing during these interactions which are important for maintaining canonical events. The other key RNA visitor to the ribosome during protein synthesis is the mRNA that occupies a specific channel in the neck of the small ribosomal subunit as it threads through the decoding site during triplet decoding [3]. This threading does not impose a stress on the triplet code reading frame unless there is a 'tangle' of some kind in the downstream region of the mRNA. Such ordered tangles, commonly in the form of stem-loops or pseudoknots, can impose sufficient pressure to facilitate a non-canonical or 'recoding' event in the form of a change in frame in the mRNA or a change in the interpretation of a particular codon. Generally, this will occur only when additional particular sequence motifs occur in the mRNA itself. Under these conditions, there is a recoding event, since the expected translational event does not occur because of the new meaning of the codon or because of a subtle or more significant reading frame change [4].

Genetic recoding initially only seemed to happen in the test tube under special conditions, or, at most, to be the domain of viral RNAs that use extraordinary means to subvert the host ribosomes' fidelity for their own purposes. Clearly, viruses are major users of recoding mechanisms on host-cell ribosomes, but, in addition, it is now well established that cells themselves use recoding strategies as another layer of regulating the expression of a small subset of their genes. Interest in genetic recoding has grown as its importance as a mainstream mechanism of

Protein Synthesis and Ribosome Structure. Edited by K. H. Nierhaus and D. N. Wilson

ISBN 3-527-30638-2

regulating gene expression has become apparent and as its absolute importance in the biology of some pathogenic viruses, like HIV-1, has become obvious. Speculation is now occurring as to whether recoding is widespread as a mechanism to produce minor products in addition to standard proteins from a particular mRNA. This could contribute to the proteome in ways that are beyond comprehension at present [5]. However, genetic recoding is not just restricted to changes in the translational frame. It is a term used to encompass other non-canonical events that give unexpected results from the translation of a signal in the mRNA. For example, included in this definition are the following:

1. Selenocysteine incorporation at a stop codon in a small subset of UGA stop codons in pro- and eukaryotic genes [6].
2. Incorporation of amino acids by the decoding of stop codons by near-cognate tRNAs (commonly referred to as readthrough) [7].
3. Disengaging and slipping through a section of the mRNA (bypassing) [8].
4. Frameshifting on the mRNA, either + or – (slipping forward or slipping backward) [9].

10.2 Maintaining Decoding Accuracy and the Reading Frame

Ribosomes are not absolute in their avoidance of error during translation of an mRNA; they incorporate an incorrect amino acid only occasionally, perhaps 1 in $10^3 – 10^4$ times and they lose the reading frame of the mRNA maybe 1 in $10^4 – 10^5$ times. Fortunately, this level of error does not compromise the ability of the ribosome to make a protein of 1000 amino acids. This means that if recoding is to occur, then there must be active mechanisms to promote recoding rather than relying on natural error, especially if it is to occur at a specific site. We know now from the work of Ramakrishnan and co-workers [10] that the rRNA uses a sensing mechanism of the codon–anticodon interaction in each of the three nucleotide positions when the substrate tRNA is in the A-site. For example, interaction of a 16S rRNA base (A1493) with the 2′-hydroxyl in each ribose of the nucleotides in the first base pair precludes non-Watson–Crick interaction at this position. In contrast, the wobble base pair is not as constrained as the other two base pairs and allows for more variety in the kinds of tRNA:mRNA interactions at this position. Similarly, the structures of the 70S ribosome with a P-site tRNA and the 30S subunit, where part of a second subunit molecule was found to mimic the decoding stem of the P-site tRNA, indicate that there are constraints on interactions at the P-site. At this site, the mRNA is forced to adopt a kinked formation allowing the anticodon stems of the A- and P-site tRNAs to be relatively far apart. The ribosome structures have revealed much detail and allowed insight into how decoding accuracy can occur, resulting in speculation of constraints that prevent a shift in reading frame.

Clearly, there are mechanisms and constraints through structural interactions of the tRNAs with the rRNA and the mRNA that are strong determinants for the canonical events of protein synthesis and ensure that non-canonical events are an exception rather than the rule. For a non-canonical event to occur there have to be

extraordinary circumstances that overcome the normal restraints on such events. These can be primary sequence *cis* signals and/or specific secondary structures in the mRNA spaced at particular distances from a primary signal. These mRNA elements somehow perturb the normal kinetics of protein synthesis, often causing pauses at decoding sites that allow competition between canonical and non-canonical events. The non-canonical result will occur in a proportion of the ribosomal passages through the signal with the frequency dependent on competition strength between the canonical and non-canonical events.

10.3 The Use of a Stop Signal for both Elongation and Termination of Protein Synthesis

Although the stop codons (UAA, UAG, and UGA) were once thought to be used universally as stop signals in protein synthesis, there are now many specific examples where they have been captured to encode amino acids. However, in most of these instances they only signal stop or sense. For example, in mitochondria and in mycobacterium species, UGA is frequently used to code for tryptophan but in these cases does not signal stop. Also, unicellular eukaryotic organisms such as Tetrahymena use UGA for stop whereas UAA and UAG code for glutamine. These are all examples where there has been codon 'takeover' or, perhaps, 'reassignment' in different organisms although the events that they signal are still canonical processes of protein synthesis.

An exception to this kind of promiscuity for stop codons is when a UGA signals selenocysteine (Sec) in a small number of genes but still signals stop in the same organism in the vast majority of occurrences. This implies that elongation and termination must be in competition at the signal. This competition occurs in a wide range of organisms in the eubacteria, archaea and eukarya kingdoms. The mechanisms for incorporation of Sec at UGA sites are distinct in prokaryotic and eukaryotic organisms, although there are some similarities. For example, there is a secondary structural element that is critical for the signal to function in both but in prokaryotes it is within the coding region and close to the primary signal UGA, whereas in eukaryotes it is quite distant and found in the 3′-hychroxyl untranslated region. In both cases, there is a special tRNA for Sec that is a minor isoacceptor of a serine tRNA where the serine has been modified by specific proteins.

10.4 The Mechanism for Sec Incorporation at UGA Sites in Bacterial mRNAs

Elegant studies from Böck and co-workers [11, 12] through the 1980s and 1990s defined the genes that were responsible for incorporation of Sec into proteins and largely defined the mechanism of how this occurred at the ribosome. Four genes controlling this mechanism were defined *selA–D*.

10.4.1 The Gene Products

1. The *selD* gene product is a 37 kDa monomeric protein, selenophosphate synthetase [13]. Although the identity of the physiological selenium substrate that is phosphorylated by ATP is still uncertain, there is evidence for an enzyme-bound phosphoryl intermediate [14], which is then attacked *in vitro* by selenide at the active site releasing the selenophosphate product. Selenide is a highly reactive molecule and may not be the physiological substrate or, if so, may be sequestered by another molecule for this reaction.
2. *SelC* is the gene encoding a minor serine tRNA that is first aminoacylated with serine by the normal synthetase [15] and after conversion of Ser to Sec is subsequently used during translation to deliver Sec into the elongating polypeptide. Interestingly, the tRNA is not well recognized by the typical elongation factor, EF-Tu, that delivers aminoacyl-tRNAs to the ribosome. This implies that the Ser-tRNA^{Sec} would have structural features unlike all other tRNAs and, indeed, it does have subtle differences in its structure. At a length of 95 nucleotides, it is one of the longest tRNAs known and this is largely because it has a large variable arm. Moreover, several invariant residues in other tRNAs are different in the *selC* gene product and, significantly, the amino acceptor arm has eight rather than the seven base pairs of other tRNAs [16].
3. *SelA* encodes a 50 kDa subunit of an oligomeric protein comprising 10 subunits. Each subunit has a pyridoxyl phosphate moiety. Conversion of Ser-tRNA^{Sec} to the Sec derivative is catalyzed by this enzyme using the selenophosphate as a substrate donor of selenium. The whole conversion takes place on this enzyme [17]. The serine is converted into amino acrylyl derivative by elimination of water from the seryl moiety first and then the activated selenium derivative is added to this intermediate to complete the conversion. The enzyme has a high degree of specificity for the Ser-tRNA, with one tRNA bound per two subunits [18]. Using electron microscopy, it was determined that the enzyme comprises a double ring of five subunits each, consistent with the stoichiometry of tRNA binding (five per enzyme) [19]. The extra long amino acceptor stem of the tRNA and its large variable loop are both important for this binding.
4. *SelB* encodes the specific elongation factor that recognizes the Sec-tRNA^{Sec} and is clearly important for the delivery of Sec to the elongating polypeptide. It is a protein of ~69 kDa and exhibits a high degree of sequence similarity to both EF-Tu and the initiation factor, IF-2, within its N-terminal region (244 amino acids). SELB is much bigger than EF-Tu (69 kDa versus 43 kDa) and the C-terminal extension on SELB not shared with EF-Tu may have some other function such as recognizing the mRNA context of the UGA recoding site [20]. The protein cannot bind Ser-tRNA^{Sec} in contrast with the Sec derivative and this explains why Ser is not incorporated at a UGA recoding site. The major determinant on the tRNA for binding to SELB is the eight base amino acceptor stem and conversion to the typical seven base pair stem abolishes binding [21].

10.4.2
The Mechanism of Sec Incorporation

Sec incorporation involves an intriguing mechanism in which Sec-tRNA^{Sec} and SELB are major players. As well, SECIS (selenocysteine insertion sequence) elements are critical in a small number of specific mRNAs such as that for formate dehydrogenase F (FDHF) [22]. These mRNAs contain UGA and a stem-loop (the SECIS element). In *fdhF,* it was originally predicted that the stem-loop closely followed the UGA codon but further studies have suggested that the UGA is within the stem-loop structure that forms the SECIS element before it approaches the ribosomal decoding site [23]. SELB binds to the stem at a specific site in the apical loop and upper helical region. The structure of the loop rather than its primary sequence seems to be the important determinant. A bulged region on the upper 5′-arm of the stem and nucleotides in the apical loop are protected by SELB from hydroxyl radical cleavage in footprint experiments.

Over-expression of SELB and SELC does not lead to a misincorporation of Sec at typical UGA stop codons. This indicates that delivery of SELB.GTP.Sec-tRNA^{Sec} to the ribosome is different than that for other tRNAs in ternary complexes. Hüttenhofer and Böck [24] have obtained evidence that suggests the ternary complex may be in a 'pre-competent state' before binding to the mRNA stem-loop. A variety of approaches indicate that SELB must be complexed with the SECIS element for a productive interaction with the ribosome to occur. These studies suggest that binding to the SECIS element induces a conformational switch in SELB that facilitates the formation of an anticodon:codon interaction between the Sec-tRNA^{Sec} and the UGA codon as it reaches the ribosomal A-site. The SECIS element would then act like a safety switch, preventing normal UGA termination codons being decoded as Sec by the SELB.GTP.Sec-tRNA^{Sec} ternary complexes [24, 25]. Once the switch converts SELB into a 'competent state', it would give SELB a strong selective advantage when it reaches the ribosomal A-site and is in competition with the decoding release factor, RF2, to decode the now A-site UGA. In this way, the SECIS element acts not only as a functional switch for protection, but also as a facilitator to send the ternary complex along a kinetic path whereby Sec is incorporated into the polypeptide chain.

10.4.3
The Competition between Sec Incorporation and Canonical Decoding of UGA by RF2

Factors affecting the competition between the RF2 and Sec incorporation *in vivo* during translation of the *fdhF* Sec (UGA) recoding site have been defined with wild-type and modified *fdhF* sequences [26]. Altering sequences surrounding the UGA codon to create more or less efficient UGA-containing stop signals without affecting the secondary structure of the SECIS element, have indicated that the kinetics of stop signal decoding have a significant influence on Sec incorporation efficiency. The UGA codon in the specific *fdhF* sequence remains 'visible' to the decoding RF2 that *in vitro* can form a site-directed 'zero-length' crosslink to it when the secondary

structure of the mRNA created by the stem-loop is absent [27]. Increasing the cellular concentration of either the RF2 decoding molecule for termination, or the $tRNA^{Sec}$ decoding molecule for elongation (for Sec incorporation), showed that these molecules are able to compete for the UGA by a kinetic competition that is dynamic and dependent on the growth rate of *Escherichia coli*. The $tRNA^{Sec}$-mediated decoding can compete more effectively for the recoding site UGA at lower growth rates, consistent with the well-established anaerobic induction of *fdhF* expression, when, presumably, Sec-containing enzymes are in an environment protected from oxidation.

How is the competition between the RF2 and $tRNA^{Sec}$-decoding molecules mediated? There is a reciprocal relationship between termination and Sec incorporation efficiencies at the *fdhF*-recoding site UGA, providing compelling evidence for competition between the canonical and non-canonical decoding events. Mansell et al. [27] have proposed a 'helical approach' mechanism for how this competition might be mediated. The competitiveness of either decoding molecule at the UGA can change according to the relative concentrations of the participating molecules. The SELB complex carrying $tRNA^{Sec}$ is bound to the apical loop of the *fdhF* stem-loop as the sequence approaches the ribosomal decoding site. If the complex is to remain bound, it must rotate about the axis of the helical stem as the secondary-structure unwinds. There is a likelihood of a ribosomal pause or translational slowing because of the increased torsional load imposed by the unwinding hairpin. The SELB complex is ideally positioned to deliver the $tRNA^{Sec}$ just as the UGA reaches the A-site and this would apparently give the $tRNA^{Sec}$ a significant advantage over the decoding RF2 to reach the inner cavern of the ribosomal active center. Indeed, this may be why the relatively efficient termination context of the UGA performs poorly against this competition. However, the creation of a translational slowing [28] by the 'helical approach' of the SELB complex may also provide the window of opportunity for the RF2 to remain relatively competitive for decoding the UGA (Fig. 10.1A).

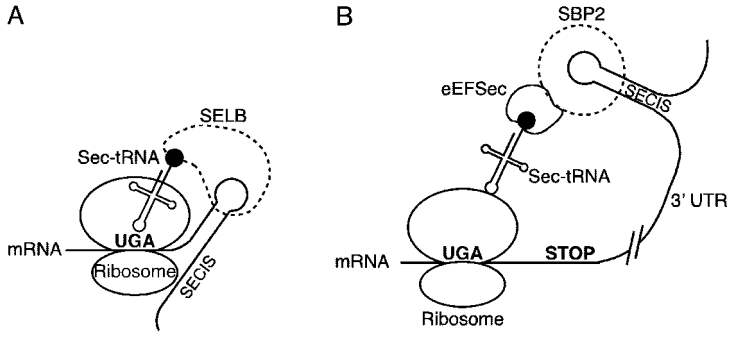

Figure 10.1 The *cis* elements and *trans* factors critical for Sec insertion during selenoprotein synthesis. The mechanism for Sec incorporation is shown for prokaryotes in (A) and for eukaryotes in (B).

10.5 Mechanism for Sec Incorporation at UGA Sites in Eukaryotic and Archaeal mRNAs

Less is known about the mechanism of Sec incorporation into eukaryotes and archaea. The SECIS elements in eukaryotes are also stem-loops but differ in structure and location from their bacterial counterparts. They are distant from the UGA site where the incorporation of Sec is to occur. Indeed, in eukaryotes, there is a minimal spacing of 60 nucleotides between the two for the SECIS element to function, but it can be as far away as several thousand bases. Although these elements are downstream of the UGA in most cases, there is one report of a SECIS element in the archaeon, *Methanococcus jannaschii* that is upstream of the UGA recoding site [29].

10.5.1 The Gene Products

Are there equivalent genes to those found in bacteria that mediate Sec incorporation into archaea and eukaryotes?

1. **SELA**, the oligomeric selenocysteine synthase has been found in archaea but not, as yet, in eukaryotes although putative homologs have been suggested [6].
2. **SELB.** The search for this specific elongation factor protein has been protracted and has resulted in several false leads. Initially, a protein, selenocysteine-binding protein 2 (SBP2), essential for Sec incorporation into rabbit reticulocyte lysate was thought to be the eukaryotic equivalent of SELB but it lacked elongation factor function [30]. Eventually, after database searches of increasing complexity to look for Sec-specific elongation factor homology from the archaea through to the eukaryotes, specific candidates for the SELB protein were identified [31]. One of these candidates, eEFSec, expressed as a recombinant protein, exhibited all the necessary and expected characteristics of the required factor being highly specific in its binding of Sec-tRNA^{Sec} and associating with SPB2 to form a complex with the SECIS element during selenoprotein synthesis.
3. **SELC** was identified first as a minor serine-specific tRNA in mammals [32] and later was shown to be the tRNA^{Sec} in eukaryotes [33]. It is present in all eukaryotic species examined and, almost exclusively, the gene is present as a single copy. The human tRNA can substitute for the bacterial SELC in Ser to Sec conversion. An important recognition determinant is a 13 base-pair coaxial helix involving an extended acceptor stem of 9 base pairs and a shortened T stem of 4 base pairs probably present in both bacterial and human tRNAs. Although the 9/4 arrangement rather than a 7/5 structure is somewhat controversial, the archaeal SELC can fold only into a 9/4 structure and provides an evolutionary reason for the presence of this 9/4 coaxial helix.
4. **SELD** was identified in humans by Berry and co-workers [34] and, although having a low similarity to the bacterial SELD, could complement a bacterial *selD* mutation. The bacterial protein is also functional in mammalian cells

suggesting a strong commonality of mechanism. The mouse and human enzymes themselves are selenoproteins with selenium at the active site and, therefore, may be involved in autoregulation of selenocysteine metabolism.

10.5.2
The Mechanism of Sec Incorporation at Specific UGA Stop Codons

Incorporation of Sec into a mammalian or archaeal protein during UGA decoding is dependent on the presence of specific structures in the mRNAs encoding these proteins. They can have different structures and certainly different primary sequences. Two consensus classes have been defined [35]. The first consists of a 9–11 base-pair stem separating a conserved SECIS element core at the base of the stem, from a 10–14 nucleotide loop with three adenosines at the 5′-side of it. In the second class, there are three adenosines comprising an internal bulge in the stem before it continues to a smaller 3–6 nucleotide loop at the tip.

How does the location of this SECIS element affect the mechanism, given that the decoding complex in bacteria is placed immediately following the UGA codon at the decoding site and the same mechanism is not feasible in mammals or archaea? Unlike the situation in prokaryotes, there are two proteins involved rather than simply a specific elongation factor to carry the Sec-tRNA^{Sec}. The first to be discovered was a SECIS element binding protein, SBP2 [30]. This protein binds selenoprotein mRNAs specifically, and Sec incorporation depends on its presence. It is speculated that the protein may play a role in excluding the eukaryotic release factor from the UGA site since it has homology to a yeast omnipotent termination suppressor of protein biosynthesis, SUP1. The eEFSec binds both isoforms of Sec-tRNA^{Sec} but not its serylated precursor or other tRNAs and, as well, binds GTP to show the classic characteristics of an elongation factor. Indeed, this protein also interacts with SBP2 and the two proteins function together for Sec incorporation into selenoproteins. This implies that the delivery complex consists of the SECIS element, SBP2, and the Sec-tRNA^{Sec} bound to eEFSec (Fig. 10.1B). How this spans the distance to the upstream UGA is not clear, but if there were a kinetic exclusion mechanism to prevent eRF1 from decoding the UGA as stop, the complex could position itself optimally for the decoding event. For example, if SBP2 were to make an association with the UGA before it entered the ribosomal A-site, then the complex already would be positioned for decoding and the eRF1 would be compromised. As both eEFSec and SBP2 have nuclear localization sites, it is speculated that the complex might be assembled on the mRNA in the nucleus ensuring the first round of translation is primed for Sec incorporation.

10.6
Why does Recoding Occur at Stop Signals?

Incorporation of selenocysteine at UGA stop codons could be explained simply by the presence of specific SECIS elements. However, the immediate context of the

UGA stop signal at most Sec incorporation sites influences competition by the Sec incorporation machinery. For example, the nucleotide immediately following the UGA recoding site in eukaryotic selenoprotein mRNAs is usually either pyrimidine U or C. Changing this nucleotide from C to a purine in the type I iodothyronine 5′-deiodinase mRNA decreases the amount of complete product and significantly increases the premature chain-termination product. This suggests that there is kinetic competition between termination and Sec incorporation at the site [36]. The context of the nucleotides surrounding stop codons clearly has a major influence on whether a particular stop codon is efficient and whether competing non-canonical events can occur (Fig. 10.2). If stop codons were decoded at different kinetic rates according to the context of the surrounding nucleotides, then there is opportunity for near-cognate tRNAs to be more competitive in some circumstances. The stop codon could be decoded as sense, or even facilitate a frameshift event to occur during a translational pause when the kinetics of stop signal decoding are particularly slow.

What is the evidence that upstream and downstream contexts can affect stop codon efficiency? As early as 1981, Kohli and Grosjean [37] highlighted an apparent bias in both the codon immediately prior to stop codons and in the nucleotide immediately following the stop codon in the very limited data set of gene sequences available at the time. Analysis of nucleotide bias surrounding stop codons became possible as more sequencing data emerged and, more recently, as the sequences of whole genomes have been completed. An algorithm was created to extract sequences around stop codons [38] and a TransTerm database constructed as a resource for translational signal analysis [39]. It was concluded from a study of nearly 1000 *E. coli* genes and lesser numbers of genes from other bacteria that the stop signal was actually a tetranucleotide rather than a triplet codon and, as well, there were clear upstream and downstream contextual biases [40]. This was most apparent for stop signals used in the most highly expressed genes (the top 10%) where, in addition to a limited subset of sense codons known to be used in these mRNAs, the stop signals had U following the stop codon almost exclusively. More recent analyses of larger data sets of both prokaryotic and eukaryotic genes indicate that there is a clear 'signature' comprising a sequence element that starts two codons before the stop codon itself and extends beyond for another six nucleotides or so [41]. This strongly hinted that there may be a hierarchy of stop signals of varying decoding efficiencies and that this may exert a significant influence on whether a recoding event could occur.

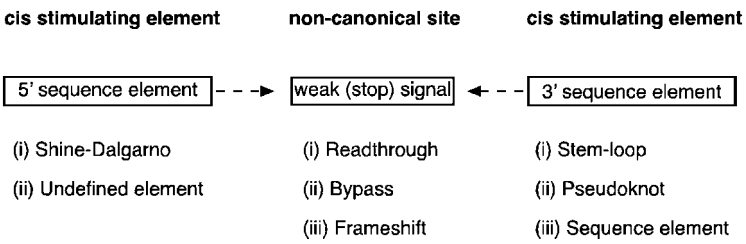

Figure 10.2 Sequence elements that influence recoding sites.

10.6.1 The Stop Signal of Prokaryotic Genomes – Engineered for High Efficiency Decoding?

Important to study this question, has been the availability of the complete genome sequence of two *E. coli* strains that had been separated by approximately 4.5 million years of evolution. The 0157:H7 pathogenic strain has acquired or retained 25% more genes than strain K12 and each has acquired or retained unique genes not found in the other strain, with 1387 genes unique to 0157:H7 and 528 unique to K12 [42]. Nevertheless, the bias of the molecular signature at the stop codons in the genes from these two strains was highly similar for TAA and TGA. In addition, the signatures were somewhat different for each of the three codons. Moreover, when this bias was analyzed to determine which individual nucleotides contributed, there was a high degree of concordance between the genes from the two strains despite there being greater then 75 000 polymorphisms in the 'homologous backbone' of the DNA sequences between the two strains. These data provide further evidence that there is a preferred sequence element for high efficiency stop signal decoding at least for UAA and UGA, and, by inference, the occurrence of non-canonical error events is minimized.

Sequence biases upstream and downstream from stop codons must reflect different mechanisms. The upstream sequences in the mRNA (the last two codons) are already involved in decoding events through ribosomal P- and E-site interactions with tRNAs and, therefore, their influence on RF stop-codon decoding must be indirect. On the other hand, the downstream sequences are not involved in mRNA–rRNA interactions according to current understanding and would be available to make direct interactions with the decoding RF protein itself. The fact that the stop signal is decoded by a protein rather than a tRNA as for sense codons, means that there is no intrinsic reason why more nucleotides than just the triplet codon might make contact with the RF. Indeed, with this in mind, zero-length site-directed crosslinking from specific positions in the mRNA to the *E. coli* RF protein using ^{4}thio-U instead of U in the RNA sequence has been carried out. Crosslinks were obtained from the first position of the stop codon (+1) and the three positions following the stop codon (+4 to +6) but not beyond, suggesting close physical contact between these nucleotide positions and the protein factor [43]. These data support the concept of direct interaction between the RF and stop signal that extends beyond just the three nucleotides from the stop codon.

What might be occurring upstream of the stop codon? Here, there are two features that are created by the tRNA-decoding events. First, the two tRNAs in the P- and E-site positions have specified the ultimate and penultimate amino acids of the completed polypeptide positioned at the peptidyltransferase center and at the beginning of the ribosomal exit tunnel, respectively. Secondly, the tRNAs themselves have a common three-dimensional shape but have micro stereochemical detail in the bases and modifications in each position of their sequence. Therefore, a particular tRNA in the P-site can create a three-dimensional environment against which the decoding RF, spanning between the decoding center and the peptidyltransferase center in the

ribosomal A-site, has the potential to make contact. In this way, both the amino acids specified and the tRNAs themselves may influence the stop signal decoding rate by the RF. This protein is occupying the binding site (A-site) that has been carefully crafted for an aminoacyl-tRNA and, by analogy, is like a cuckoo in the nest of another bird. Just as the cuckoo can be too big for the nest, so the RF may be constrained in the ribosomal A-site. If this were the case, then the upstream sequences coding for specific tRNAs and amino acids may contribute to a three-dimensional binding site for the RF that results in altered kinetics during stop codon decoding.

To determine whether the C-terminal amino acid of a protein might be having an effect on the efficiency of stop signal decoding, it is possible to examine whether a particular amino acid is abundant and whether its occurrence is highly biased. It is interesting that the two *E. coli* strains have similar proportions of each amino acid at the C-terminal positions of their proteins. The abundance and bias in the use of amino acids at this position was highly similar over all genes within the two strains and particularly within the TAA and TGA terminating genes. There were global trends in bias both for and against amino acids with certain characteristics that were still evident when the abundance of the amino acids at the C-terminal position were analyzed. What was significant from this study was that the biases and abundances of the amino acids were not identical for each of the specific stop codons, suggesting that the biases were related to the termination phase of protein synthesis and not simply some other unrelated translational process.

The conclusion from these analyses is that the last amino acid of the protein may have some stereochemical or charge-related effect on the efficiency of RF-mediated stop signal decoding when this amino acid is positioned at the ribosomal peptidyl-transferase center through the P-site tRNA. Although the stop codon is decoded by the RF positioned at the A-site in the decoding centre, for successful termination to occur a signal must be transmitted through the RF structure to the peptidyltransferase center so that release of the polypeptide can occur [44]. Further studies are needed to investigate this possibility.

What is the situation with the penultimate amino acid that was brought to the ribosome by the tRNA now positioned at the E-site? It is difficult to understand a direct effect of this amino acid on termination efficiency as it is likely to be entering the exit tunnel and less likely, therefore, to affect the RF stereochemically. However, there may be an indirect influence on the stereochemical position of the C-terminal residue by the preceding amino acid through its interactions with the surrounding ribosomal architecture and this could explain an influence on stop signal decoding efficiency. The analysis has shown that fewer residues show bias at this position than at the ultimate position but there were still preferences for amino acids with certain characteristics [42]. The picture that has emerged from these studies is that there is a much weaker influence on termination efficiency from the penultimate amino acid than the last amino acid, but still with a suggestion of some indirect influence yet to be understood.

Bias in the two codons upstream of the stop codon may not necessarily reflect the amino acid at all or solely, but may be a more direct effect of the tRNAs that are bound into the two ribosomal sites and their interactions with other RNAs. As a

particular amino acid can be carried by different tRNA isoacceptors often recognizing different codons (e.g., Leu_1 CUG; Leu_2 CUC/U; Leu_3 CUA/G; Leu_4 UUG; and Leu_5 UUA), the codon abundance in the last position before the stop codon has been analyzed in detail. This has allowed a determination of whether there is a subset of tRNAs in the termination complex during stop signal decoding. As for analysis of amino acid frequency, both abundance and bias of codons are important criteria and may not necessarily correlate. There were both specific codons preferred and specific codons that were rare in the last position. These were different from the biases found in the penultimate codon but, as with the amino acid analysis, effects in this position were much less marked and gave a relatively weak signature.

What do these trends in codon use at the last amino acid position mean for the selection of specific tRNAs? There were several striking consequences. All but one of the codons common in this position were decoded by only one isoacceptor tRNA species, whereas codons selected against were often decoded by several isoacceptor tRNAs. There were specific sequence characteristics of the abundant tRNAs and some of these related to particular stop codons. For example, before UAA, the most common four codons were decoded by only two tRNAs that were two (of only three) tRNAs with a modified mnm^5s^2U at the anticodon wobble position 34. These modified tRNAs were also used abundantly before the other stop codons as well. This suggests that the ultimate tRNA is contributing to the efficiency of stop signal decoding and may reflect, as indicated above, its contribution to the binding site architecture for the decoding RF in the A-site or the maintenance of stable peptidyl–tRNA interactions. Structurally, the tRNA would line one side of the space that the RF occupies during decoding and there is a potential for interaction between these two macromolecules. Indeed, site-directed crosslinks from the elbow of the tRNA (position 8) and the anticodon loop (position 32) to the RF has been achieved with zero-length crosslink moieties in the tRNA suggesting that there is very close contact at several positions [45].

Is there any evidence that bioinformatic analyses have revealed important features of the stop signal that are physiologically important for both non-canonical recoding and canonical decoding in protein synthesis? First, classic recoding sites where stop signals are involved do have contexts that are rarely found at natural termination sites of genes. The best example is the frameshift site in the bacterial RF2 gene where the downstream context UGACUA is found in only three other genes. This context is the weakest of all 64 UGANNN sequences tested at this site *in vivo* allowing the non-canonical recoding event to occur in approximately 90% of ribosomal passages. In contrast, the strongest context of the 64 possibilities was UGAUUA with substitution of U for C in the +4 position and consistent with the predictions from bioinformatic analysis. Similarly, the downstream context UGACAC at the selenocysteine incorporation site in the *fdhF* gene was also shown to contribute to a relatively weak termination signal. At each of the three stop codons, the data from the 64 contexts of the +4 to +6 bases gave a hierarchy of signal efficiencies with the +4 base highly influential. This supports the original suggestion that the termination 'codon' should be thought of as comprising four and not three bases [39].

Indeed, in eukaryotic *in vitro* experiments, the RF has been shown to respond to a minimum of a four-base stop signal [46].

Does the termination signal efficiency of particular downstream contexts determined experimentally correlate with signal abundance in *E. coli* genes? For TGA contexts, signal strength correlated well with signal abundance and six-base bias, implying the most efficient signals are those that are most frequently used and that use of inefficient signals is avoided except at recoding sites. However, the minor set of TAG signals in *E. coli* continue to be an enigma as there was a range of decoding efficiencies (Fig. 10.3) but with a negative correlation between abundance and bias with efficiency.

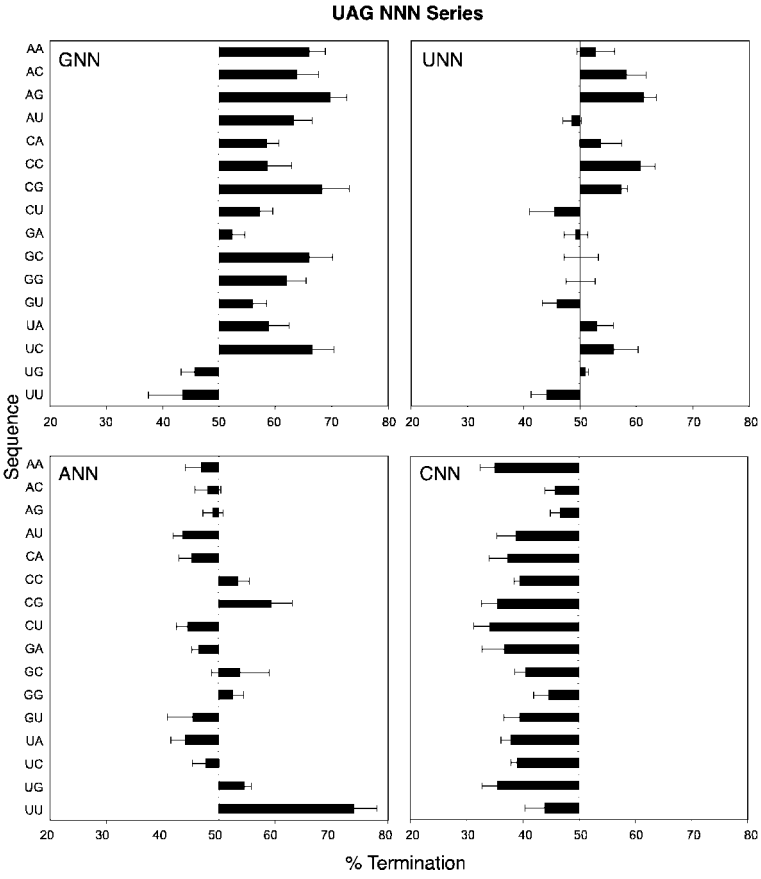

Figure 10.3 Termination efficiencies of UAG-NNN signals in competition with frameshifting. The termination efficiencies were measured as described in Ref. [78]. Efficiencies are shown relative to a point of equal competition (50% termination midline) with signals divided into graphs according to the identity of the +4 base immediately following the UAG codon. The majority of signals with a +4 G support > 50% termination, signals with a +4 C all gave < 50% termination and signals with a +4 U or A fall on either side of this midline. The error bars are the S.E.M. for at least six determinations of each signal.

Apart from suggestions from the bioinformatic analyses described above, earlier experimental work had shown effects of the last two sense codons on termination signal strength at selected contexts, as measured by the failure of the signal to specify stop but instead allow incorporation of an amino acid through near-cognate decoding [47–49]. More recently, termination signals spanning 12 nucleotides comprising the last two sense codons, the stop codon, and the three nucleotides following, were constructed to reflect predicted weak, strong and hybrid (strong upstream and weak downstream, and vice versa) signals (Fig. 10.4) [42]. Again, for UGA signals it was possible to predict correctly which would be strong and which would be weak (those that allow significant stop codon readthrough) with upstream and downstream sequences acting co-operatively. This became more obvious when tested in *E. coli* strains carrying suppressor tRNAs where competition was stronger with a cognate tRNA present. With UAA signals, competition from suppressor tRNAs was sufficient to reveal relative strengths of the stop signals only in the suppressor strain. With the enigmatic UAG signals, the 5′ contexts behaved according to prediction, but the effects of 3′ contexts did not correlate with bias or abundance.

These studies clearly indicate that context both upstream and downstream of the stop codon in bacteria has effects on the efficiency of stop codon decoding. Certain contexts not frequently found with UAA and UGA stop codons are assumed to increase the translational pause at the A-site stop and allow for competing events such as near-cognate decoding or translational frameshifting to occur. They provide the capacity for a recoding site to evolve at a stop codon allowing non-canonical events to occur during protein synthesis, thus recruiting additional complexity to the regulation of gene expression. This is clearly the case at the frameshift site of the *prfB* gene encoding RF2 [50–52] and at the Sec incorporation site of the *fdhF* gene [27]. In both cases, the balance of the canonical and non-canonical events can be significantly altered *in vivo* by changing the stop codon sequence context despite not altering the specific *cis* elements that favor the non-canonical event. Physiologically, it is events *in trans* that alter competition such as the concentration of the decoding RF in the case of the frameshift event [52] and the relative concentrations of the two decoding molecules RF and Sec-tRNA^{Sec} in the case of Sec incorporation [27]. On the other hand, such competitions are precluded at most stop signals found at the ends of the coding

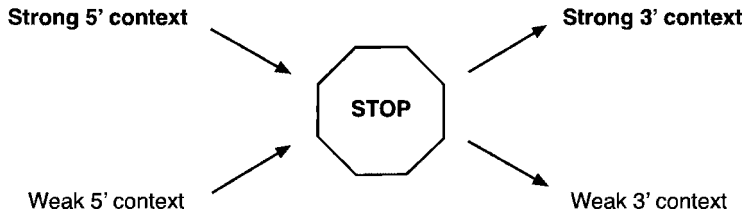

Figure 10.4 The strategy for testing the strength of termination signals spanning 12 nucleotides. Constructs were designed that contained TAA, TGA, or TAG with a predicted strong or weak sequence element 5' or 3' to the stop codon to give 'strong', 'weak' and hybrid signals.

regions of mRNAs because the context makes the stop signal so competitive for termination against non-canonical events that recoding is insignificant.

10.6.2
The Stop Signal of Eukaryotic Genomes – Diversity Contributes to Recoding

When nucleotide bias is examined in the non-redundent cDNA sequences from a number of eukaryotic genomes such as *Saccharomyces cerevisiae*, *Drosophila melanogaster, Caenorhabditis elegans*, or *Homo sapiens*, a characteristic pattern is seen [53]. A scan towards the stop signal reveals a characteristic increase in bias a few nucleotides before the codon as is seen for prokaryotic genomes and following the stop codon there is a gradual decrease of bias for up to nine nucleotides downstream (Figure 10.5A). This is the classic signature of a sequence element, with nucleotides upstream and downstream from the stop codon having the potential to contribute to the strength of the signal. When the nucleotides contributing to the bias are analysed, the pattern is similar amongst the various eukaryotic genomes but different from those of the *E. coli* strains. For example, where U (especially) and G in the +4 position are highly abundant and contribute to highly efficient signals in bacteria, in eukaryotes, the purines G (predominately) and A are favored (Figure 10.5B).

Highly expressed genes in these eukaryotic genomes have been classified by analysis of two-dimensional protein gels and through the use of the codon adaption index (CAI; a measure of the codon subset used within the gene and strongly correlated with level of expression) together with a number of other criteria. The genes identified as highly expressed have a more significant bias in the nucleotides

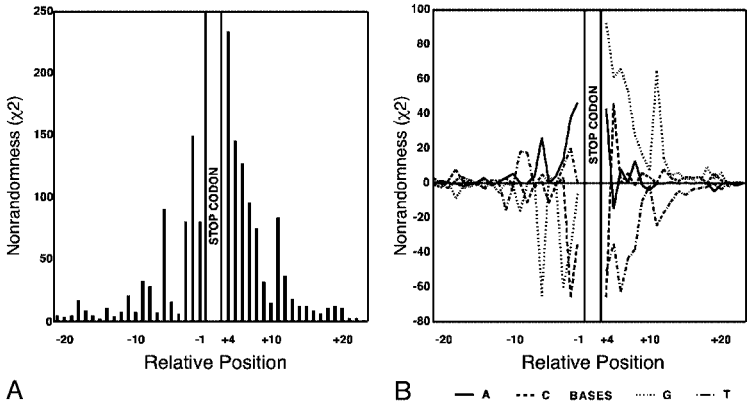

Figure 10.5 A statistical analysis of the nucleotide bias surrounding stop codons in the *Saccharomyces cerevisiae* genome. The bias (nonrandomness) was determined (A, B) by calculating the χ^2 value [(observed-expected)^2/expected]. Bias is shown for nucleotide position (A) and in the individual bases contributing to bias at each position (B).

around the stop codon compared with those genes that are expressed at lower levels. This indicates that there is some translational advantage to the bias in the sequence element specifying termination. The bias and abundance of the amino acid found in the C-terminal position and the particular codon used is less marked than in prokaryotic genomes although there are residues that are still over-represented (His, Lys) and under-represented (Gly, Pro). There are also codon biases, for example, the lysine codon AAA is over-represented with respect to AAG in the yeast genome. This may relate to the fact that the tRNA for AAA is the hypermodified $tRNA_2^{Lys}$, which may improve stability of peptidyl–tRNA interaction at the P-site during stop signal decoding with a corresponding decrease in the rate of readthrough at the following stop codon. Various features of a tRNA contribute to an ability to suppress a stop codon and read it as sense. Suppressor tRNAs, apart from $tRNA^{Sec}$, are normal cellular tRNAs with a primary role to decode cognate sense codons but also they can decode stop codons that are near-cognate. There must be some enhanced codon:anticodon stability in the near-cognate interaction perhaps through unconventional base-pairings. Supporting this, is the extent of tRNA modification in or 3′ to the anticodon important to enhance or depress readthrough rate [54]. Clearly, an unmodified tRNA coupled with a weak termination context would most probably enhance stop-codon suppression and allow readthrough to compete with the termination event.

Bioinformatic analysis of the eukaryotic genomes has enabled prediction of strong and weak stop signal elements. Strong elements favor penultimate and ultimate codons that use tRNAs modified in their anticodons, coupled with G in the +1 position immediately 3′ to the stop codon. In contrast, weak elements favor penultimate and ultimate tRNAs that are unmodified in their anticodons and a stop signal with T or C in the +1 position. These predicted weak signals are similar to the sequences found in viral recoding sites [53]. Experiments in mammalian *in vitro* and *in vivo* systems have confirmed that sequences over-represented upstream and downstream from the stop codon support only low levels of readthrough compared with that supported by under-represented sequences. Sequences found at viral recoding sites supported elevated levels of readthrough in the test system and the patterns were similar for all three stop codons and in contrast with that found at prokaryotic stop codons. This can be explained by the fact that the eukaryotic RF, eRF1, recognizes all three stop codons in contrast with the two bacterial factors that each recognize UAA but are then specific for either UAG or UGA.

Another recent study using mammalian cells *in vivo* showed that the –1 base (the base immediately prior to the stop codon) of a UAG stop signal element influenced recoding at the site mediated either through an effect from the characteristics of the P-site tRNA or interactions of the anticodon with the nucleotide [55]. In yeast, termination efficiency can vary significantly according to context and specific sequence motifs supporting readthrough have been identified [56]. Rousset and co-workers [57] looking for 'weak termination contexts', found readthrough motifs in eight yeast open reading frames within the genome. Studies with one of these, *PDE2* encoding a cAMP phosphodiesterase, found a 20-fold difference in readthrough depending on

whether the strain was [*PSI*$^-$] (an epigenetic element resulting in decreased accuracy of translation termination) or [*psi*$^-$] (normal strain) and the extended protein had lower stability.

Just as is the case with prokaryotes, in eukaryotes, the sequences both upstream and downstream from the stop codon can have a profound influence as to how efficiently the codon signals stop. The nature of the sequence element opens up opportunities for recoding to occur at the stop codons and provides a subtle layer of gene regulation in specific circumstances where defined amounts of a protein are required.

10.7
Readthrough of a Stop Signal: Decoding Stop as Sense

The fact that the sequence context of the stop codon can produce a hierarchy of stop signals of varying efficiencies gives the potential for subtle redefinition of the signal so as to regulate amounts of a protein dependent on a readthrough event or for provision of a balanced ratio of two proteins from the same mRNA sequence. The extreme case is where the codon itself is defined for two purposes as with Sec incorporation at specific UGA codons discussed above. In this case, there is a cognate species for both events rather than a near-cognate competitor. However, it requires extra sequence elements for one of the cognate decoding species, the Sec-tRNA^{Sec}, to be competitive at the site. Indeed, Sec incorporation could be regarded as a canonical event and not as a recoding event where the UGA codon is reclassified in the same genetic system for another purpose. Rather than modifying one of 64 codons to provide a specific codon for Sec with a modified base to provide unique structure, one of the existing codons is utilized. Given that the origin of Sec may have been ancient and its presence particularly relevant as a catalytic residue when oxygen was not such a dominant physiological molecule, then UGA may have originally encoded Sec but subsequently has been captured as a stop codon. It is interesting that in other genetic systems where a stop codon is used to specify an amino acid, for example, UGA for Trp in mammalian mitochondria, there is no duality of meaning for the codon. In mitochondria the definition of the codon as stop has most probably been lost during evolution.

Readthrough of the UGA stop codon at the Sec insertion site in the *fdhF* mRNA is prevented even in the absence of selenium [58] and this depends on sequences upstream and downstream from the UGA. This could be interpreted either as a special protective element or as the sequence element of the stop signal now simply out-competing near-cognate events. The natural sequences were better at preventing readthrough than alternatives tested. This context may have evolved so that competition for decoding was only between Sec incorporation and termination, excluding the third possible event of another amino acid being incorporated at the site. This would be potentially possible if the decoding rate by the competing RF was slowed. However, such a protein would be non-functional without the key Sec residue at the active site thereby providing a rationale to exclude this possibility.

Most cases of stop codon readthrough would involve near-cognate tRNAs that become competitive either because of the sequence of the stop signal or because there are other *cis* elements that favor the near-cognate over the cognate event. Given the potential for readthrough to be used as a mechanism of regulating gene expression and creating more diversity in the proteome, it is surprising that more examples have not been found. To date, there are a small number of examples where readthrough seems to be important and these span viruses, bacteria, and eukaryotes. Readthrough occurs at UGA and UAG stop codons with the most common amino acids incorporated, Trp and Gln, respectively. Although the efficiency of the event is relatively low (1–10%) in competition with termination, this is still 100- to 1000-fold above the error rate. The *cis* elements that influence readthrough can be well beyond the boundaries of the stop signal. For example, similar to the SECIS element in eukaryotes for Sec incorporation, an element several hundred nucleotides downstream from the stop codon influences readthrough at a specific stop codon in barley yellow dwarf virus [59]. In addition, secondary-structural elements are also important. A classic example is found in the synthesis of the murine leukemia virus gag-pol precursor protein, where a pseudoknot is an important mediator of the recoding event [60]. On the other hand, other viral examples do not seem to have *cis* elements beyond the immediate environment of the stop codon that is under recoding pressure. For UGA recoding sites in Sindbis virus [61], and in an *E. coli* bacteriophage RNA Qβ [62] only the +4 base of the stop signal seems critical. In both cases, the nucleotide following the stop codon is C contributing significantly to a poor context for termination. These situations would be classic candidates for readthrough based on what we know now about the stop signal.

Atkins and co-workers [63] have examined 91 unique viral sequences where readthrough of stop signals is known to occur. It is of interest that 90% had one of six tri-nucleotide sequences downstream from the site (out of the possible 64). The authors make the point that the identity restriction of six nucleotides following the stop codon is remarkable given they come from RNA viruses where mutation rates are high. In other words, there has been strong pressure to retain the contexts. While readthrough may reflect the strength of the stop signal that these contexts create and the rate of decoding by the RF, in the case of RNA viruses where evolution of optimum sequences is likely to occur quite rapidly, other equally or more important features may also have evolved. These may allow for an enhanced rate of aminoacyl-tRNA binding during near-cognate decoding (perhaps mediated by non-canonical interactions between particular context sequences of the stop-codon mRNA and the rRNA in the environment of the A-site) apart from any secondary structural elements that may be enhancing these effects. Recoding in these circumstances utilizes a stop codon as the marker for the site and in most cases a stop signal that is decoded more slowly by the RF so that there is a favorable site for building a recoding signal with the addition of other sophisticated and specific elements in *cis* or in *trans*.

The examples in prokaryotic and eukaryotic genomes where readthrough appears to be important are generally less well studied but are potentially very interesting. For example, in *D. melanogaster,* at least three genes seem to be regulated via stop codon

readthrough. Readthrough at a UGA in the *kel* gene is regulated both in a tissue and developmentally specific manner with maximal readthrough during metamorphosis [64]. A topoisomerase gene in *Bacillus firmus* is the only documented bacterial gene not derived from bacteriophages that is supposed to use readthrough as a means of regulation [65].

10.8 Bypassing of a Stop Codon: 'Free-wheeling' on the mRNA

In the last decade or so, the more RNA is studied the more remarkable mechanisms are discovered associated with its biology. Many of these could not have been anticipated or dreamt by scientists as possibilities. One such recoding event is stopcodon bypassing, where a section within the mRNA coding region is missed out and recoding starts again further down the mRNA. The classic example is for the bacteriophage T4 *gene 60*, a topoisomerase subunit gene, where 50 nucleotides are omitted from the decoding process for the correct full-length protein to be produced [66]. In this case, the flanking codons are matching GGAs decoded by $tRNA^{Gly2}$ and the next codon to be decoded after the first GGA is a stop signal with a weak downstream context (UAGCCU). In this case, the choice of events is for the peptidyl-tRNA in the P-site to detach from the mRNA before the UAG in the A-site is decoded and synthesis terminated. This detachment (called 'take-off' by Atkins and colleagues) occurs with very high efficiency and under physiological conditions it appears that termination is out-competed. There are additional *cis* and *trans* elements that drive the event that includes in addition to detachment of the peptidyl-tRNA, scanning of the mRNA and re-attachment (or 'landing' according to the Atkins' nomenclature) where canonical protein synthesis resumes.

As has become the familiar pattern at recoding sites, the stop codon forms the basic platform for the event and other sophisticated elements have been put in place to ensure its efficiency. The detachment site and the stop codon are within the stem of a stem-loop secondary structure but, intriguingly, a sequence of charged and hydrophobic amino acids in the nascent peptide synthesized up to this point also acts as mediator of the event [67]. Clearly, to initiate the event there must be competition between detachment and termination. The elements favoring detachment overwhelm stop codon decoding by RF1, the cognate decoder of UAG, as most ribosomes initiate bypassing. A detailed study has been undertaken to try to assign the importance of the various *cis* and *trans* elements in the three stages of the event; detachment, scanning and landing [68]. Structures of the bacterial ribosome suggest the peptidyl-tRNA is held at the P-site with a number of interactions between it and rRNA in the vicinity of the decoding site where codon–anticodon interaction is occurring [69]. This is in contrast with the A-site tRNA, where there is a paucity of apparent interactions near the site of codon–anticodon interaction. Slippage of the peptidyl-tRNA on the mRNA involving both detachment and movement to prevent re-attachment must require some significant disruption to these normal contacts between tRNA and rRNA. It is assumed that the decoding rate of UAG by RF1 is sufficiently slowed so that the force for detachment can predominate. Indeed, even with

a temperature-sensitive RF1 strain it was not possible to demonstrate competition between termination and detachment because of the overwhelming advantage of the recoding event and it was not until functional RF1 was over-expressed that a decrease in detachment of the peptidyl-tRNA was observed [68]. Presumably, this was because the local concentration of RF1 at the recoding site was higher, the rate of decoding the stop codon was enhanced and, therefore, the kinetic pause was shortened.

This study supports a simple pause model at the termination codon for providing the right conditions for 'take off' and modest stability of the peptidyl-tRNA interaction at the P-site. The current model for RF-mediated release of the completed polypeptide from the P-site tRNA invokes a conformational change in the RF with altered ribosomal interactions and positioning in the A-site after successful cognate decoding of the stop codon. This is proposed to trigger a signal to the peptidyltransferase center that initiates hydrolysis of the ester bond between the ultimate tRNA and growing polypeptide [70, 44]. If such a conformational change is prevented or altered following successful decoding at the recoding site, this might also favor detachment of the peptidyl-tRNA before hydrolysis can occur. However, as high concentrations of RF1 can compete successfully, the bypassing elements can, at most, decrease the likelihood of a successful signal being triggered to the peptidyltransferase center by the RF. On the other hand, the initial binding rate of RF1 to the A-site might be decreased significantly leaving the A-site empty and an empty A-site might be required for detachment.

The role of the *cis*-acting stem-loop can be compensated for by the removal of ribosomal protein L9. This suggests L9 may have a role in preventing slippage of the peptidyl-tRNA at the P-site, or a role in A-site decoding. However, recent data suggest defects in L9 may enhance mRNA movement through the ribosome [71]. Therefore, the ability of ribosomes lacking L9 to complement mutations in the stem-loop might be through this mechanism. On the other hand, the loss in bypass efficiency by mutant $tRNA^{Gly2}$ can be compensated only by mutations in the nascent *gene 60* protein before the site. This suggests the two *trans* elements, the tRNA and the nascent peptide, are operating through different mechanisms. The data support the nascent peptide either enhancing the peptidyl-tRNA dissociation and the stem-loop occupancy of the A-site, or indirectly enhancing movement of a dissociated peptidyl-tRNA. Presumably, all of the elements will be impinging on the ribosomal architecture and interactions in different ways to loosen structurally mediated tight controls on frame maintenance that are important for canonical decoding events during translation.

The mechanism of bypassing seems bizarre to contradict all reason as a logical mechanism of making a viable protein but this is characteristic of RNA. That these highly unusual mechanisms are present, is an important example of how a complicated RNA machine involving three types of RNA can evolve even more sophisticated processes beyond the extremely intricate procedures required for normal canonical events.

10.9
Frameshifting Around Stop or Sense Codons

Frameshifting involves a disruption and slippage of the interactions between mRNA and tRNA. There is forward (+) or backward (–) movement of the mRNA with respect to tRNA anticodon:codon binding. However, with frameshifting, the movement is of only one or two bases before re-engagement of the tRNA and disruption of the original reading frame is a consequence of this slippage. This is in contrast with 'bypassing' mRNA, where a greater length of the mRNA is avoided with respect to the tRNA. Whereas disruption of the reading frame can occur here as well, translation still has the potential to resume in the same frame as the prior synthesis. Gallant and Lindsay [72] have shown that ribosomes can slide over 'hungry codons' (described as such where an aminoacyl-tRNA is limiting) and then resume translation at a cognate codon many nucleotides downstream similar to the classical bypass event shown with bacteriophage T4 *gene 60*. Slippage over 'hungry codons' is not 'programed' but occurs in the mRNA under a special set of physiological circumstances [72]. In both frameshifting and bypassing, the existing interactions between tRNA and mRNA and possibly also rRNA are perturbed in the initial event. The difference is in the re-engagement process. It occurs almost immediately in the case of frameshifting and is mostly dictated by the particular mRNA sequence and the opportunity for the tRNA to re-engage in the new frame through anticodon:codon interactions. In this way, frameshifting is an example of a programed translational event.

Thus far, discussion of recoding in this chapter has focussed on sites that have stop codons as an essential framework (readthrough, Sec incorporation, bypassing). In the case of frameshifting, + or – slippage events are found not only at stop codons but also at regions of the mRNA where a stop codon is not present. Clearly, at these sites the advantage of having a slowly decoded stop codon to allow for kinetic competition at the translational pause is not needed because there are alternative ways in which the pause is created at a sense codon or kinetic competition sufficiently favors the non-canonical event. Non-programed frameshifting can occur naturally at sense codons but at a very low frequency (less than 5×10^5 per codon) [73].

Farabaugh and co-workers [74] have described how frameshifting can disrupt the reading frame by a number of possible mechanisms such as:

1. Expansion or contraction of codon size to four or two bases instead of three (this would not require disengagement of existing interactions).
2. Orientation of the incoming tRNA to allow recognition of three bases but not the next three that are in-frame.
3. Translocation of the mRNA after peptide-bond formation to move the mRNA two or four bases forward instead of three.
4. Disruption of existing RNA–RNA interactions to facilitate slippage after translocation.

Debate on which of these mechanisms operate physiologically has been lively, but there is no necessity to explain all frameshift events by a universal mechanism.

For example, a mutant $tRNA^{Gly}$ is able to correct frameshift mutations [75] and a possible mechanism is the recognition of GGGN rather than GGN. However, it is hard to accommodate this model with the recent structural information on how the rRNA elegantly 'senses' each of the three bases of the correct codon in the ribosomal A-site [10].

10.9.1 Forward Frameshifting: the +1 Event

The first discovery of frameshifting in a cellular gene initially appeared to be a 'one-off' discovery because it was an exquisite example of how a specific gene could regulate its own synthesis by a unique mechanism. The gene was *prfB* encoding the bacterial RF2 that recognizes the UGA stop codon. The frameshift site within the RF2 mRNA contained a UGA stop codon, the very codon recognized and decoded by RF2 as stop. This immediately suggested that here was a unique mechanism by which RF2 could control its own synthesis. The stop codon is in-frame at position 26 and it was discovered from sequencing of the first 44 amino acids of the protein that there had to be a +1 frameshift event at the stop codon during translation to obtain the derived sequence of the full-length protein [50]. This was a programed event in that while the stop codon provided the framework of the site, there were other *cis* elements that facilitated the event. These elements included the codon immediately before the stop codon, CUU, that can detach from the anticodon of the peptidyl-tRNA in the P-site allowing it to re-pair in the +1 frame with UUU, comprising the last two bases and the first base of the stop codon UGA. When this occurs, the codon in the new +1 frame A-site is GAC allowing for Asp to be incorporated as the next amino acid.

Not only is translational frameshifting conserved in the *prfB* genes from a wide range of bacteria (approximately 70% of those sequenced so far) but the CUUUGA motif is also retained implying a conservation of mechanism [76]. Just 5' of the CUUUGA motif, an internal Shine–Dalgarno sequence normally found 5–6 bases in front of start codons in bacterial genes, base-pairs with a complementary region of the 3′-terminus of the 16S rRNA to facilitate the frameshift event [77]. As well, the stop signal contains the least efficient downstream sequence, UGACUA, of any found in bacteria [78, 42]. This implies that stress is placed on interactions between the mRNA, tRNA and rRNA during decoding of the CUU. At the same time as the 'misplaced' Shine–Dalgarno interaction is occurring, a weak stop signal is present in the A-site giving a decoding pause and allowing time for the tRNA to disengage from the peptidyl-tRNA and re-engage with the next base in the mRNA.

The *trans* element for frameshifting in the RF2 mRNA is, of course, the RF2 protein itself (Fig. 10.6). Modulation of its concentration can shift frameshift efficiency from 0–100% [51, 52, 79], illustrating how effective the mechanism can be to regulate the cellular concentration of RF2. Under normal physiological conditions of bacteria growing in rich media in log phase, the frameshift efficiency is approximately 30–50%. Frameshifting will also occur if the stop codon is replaced by a sense codon but then the efficiency is inversely related to the rate of amino-

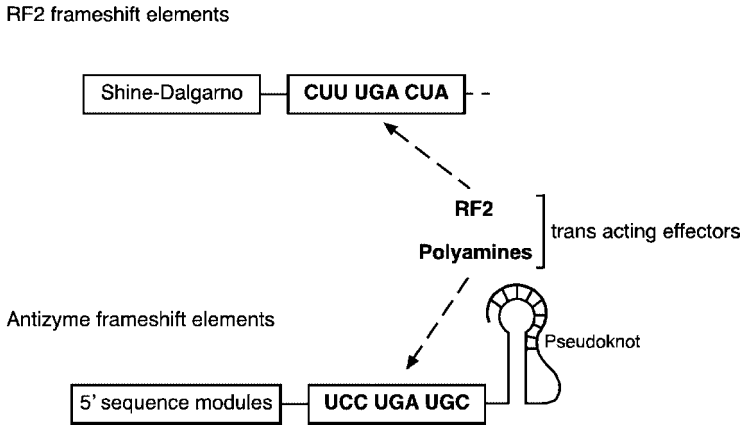

Figure 10.6 The cis elements and trans-acting effectors critical for +1 frameshifting at the RF2 and antizyme recoding sites.

acyl-tRNA selection and supports the importance of the translational pause [80]. Perhaps a U:G wobble base pair in the pre-shift codon:anticodon that has been associated with high-frequency frameshifting facilitates the event at the RF2 site [81].

A similar +1 programed frameshifting event with a novel autoregulatory mechanism was discovered some 10 years later in the mammalian ornithine decarboxylase antizyme gene [82]. Antizyme synthesis is induced by polyamines (see Fig. 10.6). Antizyme binds to and induces a conformational change in the key enzyme in polyamine synthesis, ornithine decarboxylase, targeting it for turnover. This was another example of a remarkable autoregulatory circuit where the key molecule was not a protein this time but the polyamine metabolites, spermidine, spermine and putrescine. Polyamines are known to interact with RNA and probably influence key interactions at the decoding site that facilitate the recoding event. The recoding site in the antizyme mRNA is also at the codon preceding a stop codon (UCCUGAU). Key *cis* elements are responsible for enhancing frameshift efficiency; at least 50 nucleotides in modules just 5′ of the frameshift site are responsible for 2- to 3-fold stimulation of frameshift efficiency, the stop codon increases efficiency 15- to 20-fold and there is a downstream pseudoknot just three nucleotides distant from the motif that contributes 2.5- to 5-fold. The upstream sequences may be equivalent in some way to the Shine–Dalgarno element of the *prfB* site in function. In contrast with this site, whereas findings from site-directed mutagenesis studies suggested that re-pairing between the tRNA and the +1 shift CCU was unlikely, phylogenetic analysis and observations of –2 frameshifting at the site in yeast are supportive of a re-pairing mechanism [83].

Two mammalian paralogues of what is now called antizyme 1 have been found [84] each with a frameshift site, with antizyme 3 being tissue and cell-type specific. Atkins and co-workers [83] have studied these antizyme sites in a wide range of organisms and this has enabled them to make several conclusions about recoding

events. First, recoding sites can be selected over very long evolutionary periods. Secondly, some plasticity is characteristic of the site and, thirdly, once an initial site is established a range of stimulators acting in *cis* can evolve to give specific characteristics and optimize recoding efficiencies.

An example of a mechanism for + 1 frameshifting, which is different from that of tRNA slippage coupled with slow decoding of the codon in the A-site, seems to occur with the yeast retrotransposon Ty3 [73]. In this case, frameshifting occurs at the sequence GCGAGUU as a result of the incoming tRNA pairing with the out-of-frame GUU without slippage of the peptidyl-tRNA. Interestingly, over-expressed near-cognate P-site tRNAs were able to induce frameshifting generally at the Ty3 site and is in contrast with cognate tRNAs that decreased frameshift efficiency [84]. This suggests that near-cognate tRNAs may be decoding GCG in the Ty3 site and, when coupled with a slowly decoded codon like AGU, allow take-over by the tRNA recognizing the +1 codon, GUU.

Yet another example of +1 frameshifting has been found in the yeast telomerase gene, *EST3*. Yeast use telomerase to maintain the ends of their chromosomes and the +1 translational frameshift event required to produce fully functional telomerase occurs at a motif, CUUAGUUGAG [85]. This motif is similar to the Ty1, Ty2 and Ty4 frameshift site, CUU AGG C, in the underlined portion but, as well, contains a stop codon. These additional examples indicate there may be more sites yet to be found where + 1 frameshifting plays an important part in gene regulation.

10.9.2
Programed –1 Frameshifting: A Common Mechanism used by Many Viruses During Gene Expression

Diverse virus groups have evolved a recoding site to express coat proteins and enzymes in a carefully balanced ratio. The mechanism involves switching frame during the synthesis of a polyprotein (gag) so that the extended product (gag:pol) from the new frame contains the enzyme sequences that can be excised from the protein. For most passages (approximately 90%), the ribosome does not frameshift with the result that the coat protein subunits are produced in 10-fold higher amounts than the enzymes. This subunit : enzyme ratio is important for productive infection of the virus. For such viruses, the recoding site has clear motifs and stimulatory elements consisting of two *cis* acting sequences, a heptanucleotide motif XXXYYYZ (a slippery sequence) and a secondary structural element, usually a pseudoknot.

It was a seminal paper in 1985 that indicated quite clearly that frameshifting was not going to be restricted to the example crafted so elegantly for RF2 gene expression [86]. The Rous Sarcoma Virus used a recoding site that included a stop codon (AAAUUUAUAG) similar to that used in expression of RF2 but here the shift was backwards rather than forwards. The new codon in the A-site became AUA rather than UAG and the stop codon was avoided. Jacks and Varmus [86] noticed that the AAAUUUA motif allowed slippage of the two tRNAs (Asn and Leu) simultaneously to re-pair in a near-cognate manner with AAA UUU. They proposed a simultaneous-

slippage model from the A- and P-sites meaning that slippage would occur before peptide transfer and translocation. This mechanism became the generally accepted model for –1 frameshifting.

In hindsight, there were two aspects of this mechanism that required further investigation. First, the proposed mechanism would not involve the stop codon in the recoding event within the decoding center and yet the importance of the stop codon as a part of recoding sites as discussed here is clearly established. While the codon following the slippery motif in viruses is often not a stop codon, for example, it is GGG in HIV-1, there is still a restricted subset of codons used in this position. This is highly suggestive that the UAG codon in the Rous Sarcoma Virus frameshift site (or the GGG codon in the HIV-1 site) might be involved in the recoding event and may have a role at the decoding site (Figure 10.7). How can this occur? A simultaneous-slippage model where slippage occurs at the P- and E-sites rather than the A- and P-sites, would allow the next codon to be occupying the A-site. Then the stop or sense codon could contribute to a translational pause and facilitate the slippage event. Secondly, the unusual simultaneous-slippage mechanism proposed had slippage occurring in competition with peptide-bond formation that is normally a kinetically rapid event in protein synthesis. It is more likely that frameshifting is in competition with a kinetically slower event. While others have favored this to be a step after peptide-bond formation but prior to translocation so that slippage could still occur from the A- and P-sites, we believed an equally similar step occurred after translocation, i.e., slippage from the P- and E-sites in competition with the new event in the now empty A-site. This could be the decoding of the stop or sense codon depending on the sequence composition of the particular viral site.

For this reason, we attempted to determine whether the codon after the slippery site was being decoded in the A-site before slippage had occurred. Initially, because the genes for the bacterial RFs were available within plasmids we used the HIV-1

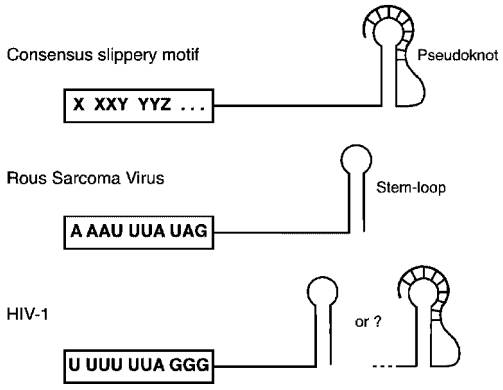

Figure 10.7 Sequence motifs and cis elements at viral –1 frameshift sites.

recoding site with the GGG codon replaced with each of the stop codons and bacterial ribosomes. We then tested whether over-expression *in vivo* of the bacterial RFs would affect frameshifting at the recoding site that occurred readily on these bacterial ribosomes. There were two key observations. First, changing the codon from the natural GGG to each of the stop codons, UAG, UGA and UAA, markedly lowered frameshifting, indicating that the codon was having a major effect on the mechanism. More significantly, over-expression of the factors eliminated frameshifting in a codon-specific manner (RF1 at UAG and UAA, RF2 at UGA and UAA). This was compelling evidence that the codon following the recoding site was being decoded before slippage had occurred and supported a P–E site simultaneous-slippage model [87]. Now that the genes for the eukaryotic RFs are available we have repeated this study *in vivo* using mammalian cells. The results from these studies using mammalian ribosomes concur with our *in vivo* studies that used bacterial ribosomes. First, the stop codon depressed frameshift efficiency. Secondly, over-expression of eRF1 caused a further reduction of efficiency as it did also at the antizyme recoding site used as a control [88]. This is provocative evidence that this –1 frameshift event is occurring in a similar manner to a +1 frameshift where recoding is in competition with a canonical decoding event at the ribosomal A-site.

Recently, Dinman and co-workers [89] have proposed an elegant integrated model to explain programed frameshifting that defines why in some circumstances +1 and in others –1 frameshifting occurs. A key element of the model is that the occupancy states of the ribosome are different for each event. For +1 frameshifting, the A-site is empty and the shift occurs in the post-translocational state (P- and E-sites occupied with tRNAs), but for –1 frameshifting, the A-site is proposed to be already occupied and therefore the ribosome is in the pre-translocational state (A- and P-sites occupied with tRNAs). They have proposed this model as a result of studies that used antibiotics with normal and mutant yeast strains and how under these circumstances frameshifting efficiency was affected. A consistent set of conclusions was drawn from the differential antibiotic sensitivities of the +1 and –1 frameshift events to support the model. For example, translocation inhibitors interfere with +1 but not –1 frameshifting, although neither do they enhance it. As predicted by the model, a peptidyltransferase inhibitor, sparsomysin, that would increase a ribosomal pause when the A- and P-sites are occupied enhances –1 but not + 1 frameshifting.

These conclusions are consistent with what is known about +1 frameshifting from the examples described in this chapter. However, it conflicts with conclusions drawn from our recent evidence that frameshifting at the classic viral –1 site also occurs at the post-translocational state of the ribosome at which time the A-site is empty. In our model, +1 and –1 frameshift events would generally start from the same post-translocational state. It is the specific sequence of the site, the nature of the constraining *cis*-acting elements and how they impinge on interactions between the tRNAs, mRNA and rRNA and, in particular, the stability of the peptidyl-tRNA:mRNA interaction that will determine whether the shift is forwards or backwards. It invokes the common element of a translational pause at the decoding A-site, dislocation in most cases of the codon:anticodon pairing interactions at the P- and E-sites and re-pairing

close by where possible (+1 and –1 events) or bypassing larger tracts of sequence if immediate re-pairing is not favorable (Fig. 10.8).

How can these apparently conflicting positions be accommodated? It could be that if there were a weakening of the already compromised interactions at recoding sites by any effector such as an antibiotic or a mutation creating a new pause in a particular ribosomal state, then this would have the potential to facilitate a shift in frame. Frameshifting may be able to occur in different states of the ribosome under these circumstances. If the partially inhibited process still does not become the rate-limiting step, then it may have no effect on frameshifting. However, if it now makes that process the slowest step, then frameshifting could be enhanced (e.g., inhibiting –1 but not +1 frameshifting by the peptide-bond formation inhibitor, sparsomycin). Moreover, as the detailed new structural information on the ribosome has become available it is revolutionizing our understanding on how antibiotics act and our understanding on the various steps of protein synthesis deduced from biochemical studies both *in vitro* and *in vivo*. Further experiments will be needed to resolve some of these apparent paradoxes and it may not be possible to invoke a completely integrated model to explain all programed frameshift events.

In addition to the classic viral –1 frameshift events of the retroviruses and other viral groups including bacteriophages, there are also examples identified in cellular genes. In the prokaryotic examples, such as that found in the gene *dnaX* encoding a subunit of DNA polymerase III [90–92], there are different features to the recoding site that reflect how elements can be added to a basic recoding framework. The Shine–Dalgarno sequence documented for the *prfB* gene (+1 frameshifting) is found in the *dnaX* gene as an upstream element, but at a different and precise spacing for acting as a stimulator. The slippery heptamer sequence, AAAAAAG, allows two Lys–tRNAs paired with AAAAAG to slip backwards and re-pair with AAAAAA so that the following G becomes the first base of the next codon. Before slippage, this G would be paired with a modified U (mnm^5S^2U) in the anticodon of the tRNA, the same modification found often in the P-site tRNA at termination sites. However, the modification prevents stable base pairing with G and thereby weakens the codon:anti-

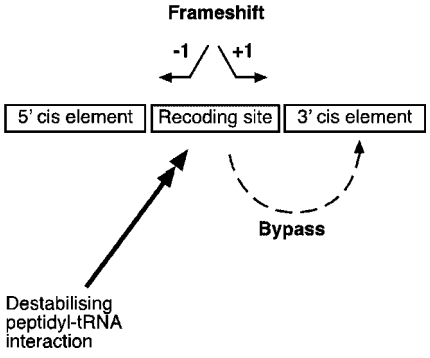

Figure 10.8 Alternative recoding events that could result from destabilization of the peptidyl-tRNA interaction.

codon interaction. Downstream from the slippery site there is a hairpin loop that can also act as a stimulator independent of the upstream element.

Does –1 frameshifting occur in cellular genes in eukaryotes? As the features of a –1 viral recoding site are quite clear-cut, it has been possible to create an algorithm to search genomes for genes that might use this strategy. While a number of possible genes have been flagged, there is, as yet, neither any evidence that they use this strategy nor a rationale as to why they might do so [93]. Nevertheless, recently, an EST was discovered that represented a gene, *edr*, that appeared to have a retrovirus as its ancestor. The recoding site was GGGAAAC with a pseudoknot located downstream. It was expressed temporally and spatially during development although it is not known whether expression is essential for development. However, it seems to use a –1 frameshift mechanism for expression [94]. As programed –1 frameshifting is used in the viral biology of pathogenic viruses like HIV-1 and is a potential target for an antiviral agent, it is important to identify putative human genes that might use this mechanism as this might preclude the viral recoding site as a place of attack in an antiviral strategy. So far, there is no definitive data to preclude this approach although further study is required on the importance of expression of the *edr* gene.

10.10 Conclusion

When RNA molecules interact there can be amazing consequences. We have known for some time that protein synthesis is likely to be the consequence of a highly accurate and fast RNA machine that incorporates specific features to ensure that speed and accuracy can occur together. Recent developments have heightened this appreciation now that the RNA and protein components can be seen in atomic detail. For the first time, it is possible to visualize how some of these special features function. The sensing by the rRNA that the correct codon:anticodon interactions between the tRNA and the mRNA are occurring at the decoding A-site is an exquisite example. The 'enzyme' (the rRNA) holds the 'substrates' (the tRNAs) with unique interactions to maintain the canonical three-base reading frame to ensure they are read with an acceptable level of accuracy. It is not surprising then that this high degree of precision can be overridden. Sometimes, this is a result of a low-frequency error. Sometimes, it is as a result of an atypical physiological perturbation such as starvation for an amino acid but as we know now, sometimes, it is a result of a non-canonical recoding event that has evolved for highly specialized physiological reasons.

As each interaction between the mRNA, rRNA and the tRNAs is so critical for maintenance of precision, there is potential for elements acting in *cis* or *trans* to perturb one or more of these specific pairings. Sequences in the mRNA, or structures that can form within it, have the capacity to form new interactions with the rRNA and put strain on normal interactions. As a consequence, the highly ordered co-ordination between structure and function can be disrupted and kinetic parameters can be slowed giving a chance for alternative events that normally would not be competitive to be significant. *Trans* activators of non-canonical recoding can be of a diverse nature such as the protein RF2 acting at the recoding site within its own mRNA, or

metabolites like polyamines in the antizyme frameshift site or a unique tRNA like $tRNA^{Sec}$ in the Sec incorporation site. Sites where recoding events have evolved often have a stop codon as a central core with upstream and downstream elements that help to ensure that stop codon is more slowly decoded. These sequence elements exert influence in two ways. First, sequences contribute to the decoding efficiency of the stop signal itself. Upstream elements likely to be contributing to the three-dimensional architecture of the RF binding site and downstream elements are probably facilitating stable interactions between the factor and mRNA. Secondly, the upstream and downstream elements can independently perturb other parts of the decoding center and affect parameters independent of the stop codon. While having a stop codon at the recoding site is not universal, it does provide an ideal environment for the creation of translational pauses and provides a vulnerable 'Achilles heel' of canonical protein synthesis where recoding events can evolve.

It is not surprising that most examples of recoding occur in the highly specialized genetic system of the virus. Both bacteriophages and eukaryotic viruses have evolved and maintained highly successful recoding mechanisms. The classic –1 frameshift site found in many eukaryotic viruses has allowed multiple proteins to be synthesized from a single RNA with economy and in balanced amounts. It is surprising, now that we know recoding can be an important translational control mechanism for gene expression, that more examples have not been found in prokaryotic and eukaryotic cellular genes. Those that are known and relatively well characterized such as +1 frameshifting in both the *prfB* gene for RF2 and in the gene for ornithine decarboxylase antizyme and Sec incorporation at selected UGA stop codons, are highly elegant mechanisms for gene expression regulation. These events add a layer of fine control and subtlety to the more typical control mechanisms. Now that new genome and proteome information is appearing rapidly, will the opportunity to examine gene function in detail reveal more examples of the recoding types that are already known and, perhaps, even new events? It is likely that new events will be discovered but these probably will represent rare exclusive examples where some niche advantage is gained. RNA provides 'flexible moulding clay' from which new mechanisms and functions can emerge [95]. There are sure to be more surprises.

References

1 H. J. Rheinberger, K. H. Nierhaus, *Proc. Natl. Acad. Sci. USA* **1981**, *78*, 5310–5304.
2 K. H. Nierhaus, *Mol. Microbiol.* **1993**, *9*, 661–669.
3 J. Frank, J. Zhu, P. Penczek et al., *Nature* **1995**, *376*, 441–444.
4 R. F. Gesteland, J. F. Atkins, *Annu. Rev. Biochem.* **1996**, *65*, 741–768.
5 S. L. Alam, J. F. Atkins, R. F. Gesteland, *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 14177–14179.
6 M. J. Berry: in *Translational Control of Gene Expression, eds* N. Sonenberg, J. W. B. Hershey, M. B. Mathews, Cold Spring Harbor Laboratory Press, New York 2000.
7 P. V. Baranov, R. F. Gesteland, J. F. Atkins, *Gene* **2002**, *286*, 187–201.
8 A. J. Herr, J. F. Atkins, R. F. Gesteland, *Annu. Rev. Biochem.* **2000**, *69*, 343–372.
9 P. J. Farabaugh, *Microbiol. Rev.* **1996**, *60*, 103–134.

10 J. M. Ogle, D. E. Brodersen, W. M. Jr. Clemons et al., *Science* **2001**, *292*, 897–902.
11 A. Böck, K. Forchhammer, J. Heider et al., *Mol. Microbiol.* **1991**, *5*, 515–520.
12 A. Böck, K. Forchhammer, J. Heider et al., *TIBS* **1991**, *16*, 463–467.
13 W. Leinfelder, K. Forchhammer, B. Veprek et al., *Proc. Natl. Acad. Sci. USA* **1990**, *87* 543–547.
14 L. S. Mullins, S-B. Hong, G. Gibson et al., *J. Am. Chem. Soc.* **1997**, *119*, 6684–6685.
15 W. Leinfelder, E. Zehelein, M. A. Mandrand-Berthelot et al., *Nature* **1988**, *331*, 723–725.
16 C. Baron, E. Westhof, A. Böck etal., *J. Mol. Biol.* **1993**, *231*, 274–292.
17 K. Forchhammer, W. Leinfelder, K. Boesmiller et al., *J. Mol. Biol.* **1991**, *266*, 6318–6323.
18 K. Forchhammer, A. Böck, *J. Biol. Chem.* **1991**, *266*, 6324–6328.
19 H. Engelhardt, K. Forchhammer, S. Muller et al., *Mol. Microbiol.* **1992**, *6*, 3461–3467.
20 K. Forchhammer, W. Leinfelder, A. Böck, *Nature* **1989**, *342*, 453–456.
21 C. Baron, A. Böck, *J. Biol. Chem.* **1991**, *266*, 20375–20379.
22 F. Zinoni, J. Heider, A. Böck, *Proc. Natl. Acad. Sci. USA* **1990**, *87*, 4660–4664.
23 A. Hüttenhofer, E. Westhof, A. Böck, *RNA* **1996**, *2*, 354–366.
24 A. Hüttenhofer, A. Böck, *Biochemistry* **1998**, *37*, 885–890.
25 S. J. Klug, A. Huttenhofer, M. Kromayer et al., *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 6676–6681.
26 J. B. Mansell, *PhD Thesis*, University of Otago, New Zealand, 1999.
27 J. B. Mansell, D. Guévremont, E. S. Poole et al., *EMBO J.* **2001**, *20*, 7284–7293.
28 S. Suppmann, B. C. Persson, A. Böck, *EMBO J.* **1999**, *18*, 2284–2293.
29 R. Wilting, S. Schorling, B. C. Persson et al., *J. Mol. Biol.* **1997**, *266*, 637–641.
30 P. R. Copeland, J. E. Fletcher, B. A. Carlson et al., *EMBO J.* **2000**, *19*, 306–314.
31 R. M. Tujebajeva, P. R. Copeland, X.-M. Xu et al., *EMBO Rep.* **2000**, *1*, 158–163.
32 D. Hatfield, A. Diamond, B. Dudock, *Proc. Natl. Acad. Sci. USA* **1982**, *79*, 6215–6219.
33 B. J. Lee, M. Rajagopalan, Y. S. Kim et al., *Mol. Cell Biol.* **1990**, *10*, 1940–1949.
34 S. C. Low, J. W. Harney, M. J. Berry, *J. Biol. Chem.* **1995**, *270*, 21659–21674.
35 E. Grundner-Culemann, G. W. Martin III, J. W. Harney et al., *RNA* **1999**, *5*, 625–635.
36 K. K. McCaughan, C. M. Brown, M. E. Dalphin et al., *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 5431–5435.
37 J. Kohli, H. Grosjean, *Mol. Gen. Genet.* **1981**, *182*, 430–439.
38 C. M. Brown, P. A. Stockwell, C. N. A. Trotman et al., *Nucleic Acids Res.* **1990**, *18*, 2079–2086.
39 C. M. Brown, M. E. Dalphin, P. A. Stockwell et al., *Nucleic Acids Res.* **1993**, *21*, 3119–3123.
40 C. M. Brown, P. A. Stockwell, C. N. A. Trotman et al., *Nucleic Acids Res.* **1990**, *18*, 6339–6345.
41 W. P. Tate, E. S. Poole, J. A. Horsfield et al., *Biochem. Cell Biol.* **1995**, *73*, 1095–1103.
42 L. L. Major, *PhD Thesis*, University of Otago, New Zealand, 2001.
43 E. S. Poole, L. L. Major, S. A. Mannering et al., *Nucleic Acids Res.* **1998**, *26*, 954–960.
44 D-J. G. Scarlett, K. K. McCaughan, D. N. Wilson et al., *J. Biol. Chem.* **2003**, *278*, 15095–15104.
45 L. L. Major, E. S. Poole, W. P. Tate, in preparation.
46 D. S. Konecki, K. C. Aune, W. Tate et al., *J. Biol. Chem.* **1977**, *252*, 4514–4520.
47 S. Mottagui-Tabar, A. Björnsson, L. A. Isaksson, *EMBO J.* **1994**, *13*, 249–257.
48 S. Zhang, M. RydénAulin, L. A. Isaksson, *J. Mol. Biol.* **1996**, *261*, 98–107.

49 A. Björnsson, S. Mottagui-Tabar, L. A. Isaksson, *EMBO J.* **1996**, *15*, 1696–1704.
50 W. J. Craigen, R. G. Cook, W. P. Tate et al., *Proc. Natl. Acad. Sci. USA* **1985**, *82*, 3616–3620.
51 W. J. Craigen, C. T. Caskey, *Nature* **1986**, *322*, 273–275.
52 B. C. Donly, C. D. Edgar, F. M. Adamski et al., *Nucleic Acids Res.* **1990**, *18*, 6517–6522.
53 A. G. Cridge, E. S. Poole, W. P. Tate, unpublished results.
54 H. Beier, M. Grimm, *Nucleic Acids Res.* **2001**, *29*, 4767–4782.
55 M. Cassan, J.-P. Rousset, *BMC Mol. Biol.* **2001**, *2*, 3.
56 B. Bonetti, L. W. Fu, J. Moon et al., *J. Mol. Biol.* **1995**, *251*, 334–345.
57 O. Namy, G. Duchateau-Nguyen, J.-P. Rousset, *Mol. Microbiol.* **2002**, *43*, 641–652.
58 Z. Liu, M. Reches, H. Engelberg-Kulka, *J. Mol. Biol.* **1999**, *294*, 1073–1086.
59 C. M. Brown, S. P. Dinesh-Kumar, W. A. Miller, *J. Virol.* **1996**, *70*, 5884–5892.
60 N. M. Wills, R. F. Gesteland, J. F. Atkins, *Proc. Natl. Acad. Sci. USA* **1991**, *88*, 6991–6995.
61 G. Li, C. M. Rice, *J. Virol.* **1993**, *67*, 5062–5067.
62 A. M. Weiner, K. Weber, *J. Mol. Biol.* **1973**, *80*, 837–855.
63 L. Harrell, U. Melcher, J. F. Atkins, *Nucleic Acids Res.* **2002**, *30*, 2011–2017.
64 D. N. Robinson, L. Cooley, *Development* **1997**, *124*, 1405–1417.
65 D. M. Ivey, J. Cheng, T. A. Krulwich, *Nucleic Acids Res.* **1992**, *20*, 4928.
66 W. M. Huang, S. Z. Ao, S. Casjens et al., *Science* **1988**, *239*, 1005–1012.
67 R. B. Weiss, W. M. Huang, D. M. Dunn, *Cell* **1990**, *62*, 117–126.
68 A. J. Herr, R. F. Gesteland, J. F. Atkins, *EMBO J.* **2000**, *19*, 2671–2680.
69 J. H. Cate, M. M. Yusupov, G. Zh. Yusupova et al., *Science* **1999**, *285*, 2095–2104.
70 D. N. Wilson, D. Guévremont, W. P. Tate, *RNA* **2000**, *6*, 1704–1713.
71 A. J. Herr, C. C, Nelson, N. M. Wills et al., *J. Mol. Biol.* **2001**, *309*, 1029–1048.
72 J. Gallant, D. Lindsey, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 13771–13776.
73 P. J. Farabaugh, *Annu. Rev. Genet.* **1996**, *30*, 507–528.
74 G. Stahl, G. P. McCarty, P. J. Farabaugh, *Trends Biochem. Sci.* **2002**, *27*, 178–183.
75 D. Riddle, J. Carbon, *Nature New Biol.* **1973**, *242*, 230–234.
76 P. V. Baranov, R. F. Gesteland, J. F. Atkins, *EMBO Rep.* **2002**, *3*, 373–377.
77 R. B. Weiss, D. M. Dunn, A. E. Dahlberg et al., *EMBO J.* **1988**, *7*, 1503–1507.
78 E. S. Poole, C. M. Brown, W. P. Tate, *EMBO J.* **1995**, *14*, 151–158.
79 F. M. Adamski, C. Donly, W. P. Tate, *Nucleic Acids Res.* **1993**, *21*, 5074–5078.
80 J. F. Curran, M. Yarus, *J. Mol. Biol.* **1989**, *209*, 65–77.
81 J. F. Curran, *Nucleic Acids Res.* **1993**, *21*, 1837–1843.
82 S. Matsufuji, T. Matsufuji, Y. Miyazaki et al., *Cell* **1995**, *80*, 51–60.
83 I. P. Ivanov, R. F. Gesteland, J. F. Atkins, *Nucleic Acids Res.* **2000**, *28*, 3185–3196.
84 A. Sundararajan, W. A. Michaud, Q. Qian et al., *Mol. Cell* **1999**, *4*, 1005–1015.
85 D. K. Morris, V. Lundblad, *Current Biol.* **1997**, *7*, 969–976.
86 T. Jacks, H. E. Varmus, *Science* **1985**, *30*, 1237–1242.
87 J. A. Horsfield, D. N. Wilson, S. A Mannering et al., *Nucleic Acids Res.* **1995**, *23*, 1487–1494.
88 R. S. Graves, C. Z. M. McKinney, E. S. Poole et al., unpublished results.
89 J. W. Harger, A. Meskauskas, J. D. Dinman, *Trends Biochem. Sci.* **2002**, *27*, 448–453.

90 A. L. Blinkowa, J. R. Walker, *Nucleic Acids Res.* **1990**, *18*, 1725–1729.
91 A. M. Flower, C. S. McHenry, *Proc. Natl. Acad. Sci. USA* **1990**, *87*, 3713–3717.
92 Z. Tsuchihashi, A. Kornberg, *Proc. Natl. Acad. Sci. USA* **1990**, *87*, 2516–2520.
93 A. B. Hammell, R. C. Taylor, S. W. Peltz et al., *Genome Res.* **1999**, *9*, 417–427.
94 K. Shigemoto, J. Brennan, S. E. Walles et al., *Nucleic Acids Res.* **2001**, *29*, 4079–4088.
95 D. Kennedy, *Science* **2002**, *298*, 2283.