By 1966 the general nature of readout of the genetic code and codon identity had been established. What was not appreciated then was that decoding is dynamic. Decoding can be altered in an mRNA-specific manner and in a remarkable variety of ways.

The specific meaning of individual codons can be redefined in response to signals in an mRNA. Or a proportion of translating ribosomes can be diverted to a different reading frame at a specific site. And ribosomes can be directed to bypass a block of nucleotides or even to resume on a different mRNA. This book chronicles and analyzes these “recoding” phenomena both to understand the contribution they make to the complexity of gene expression and to understand the mechanisms involved, illuminating the features of ribosomes and mRNA.

These unusual genetic decoding events tell us that the readout of the code itself has been subject to the wiliness of selection, increasing the repertoire of ways to utilize the richness of information encoded in DNA or RNA. A coding sequence in mRNA can specify additional protein products not predicted from standard readout of the classical open reading frame. In some cases the recoding event is a control point for a regulatory circuit. In certain other cases, the key feature is specification of the “special” amino acids selenocysteine and pyrolysine. Not surprisingly the world of viruses and small mobile chromosomal elements is rich with examples of recoding since their genomes are compact and every mechanism is used to maximize gene density. But, with one viral exception, the cases known so far of specification of the “special” amino acids are for cellular gene decoding.

Deciphering recoding has led to the realization that there is an extra layer of information in messenger RNA that can change the program for its own individual readout. These instructions include a site where the nonstandard decoding event occurs and an assortment of types of signals that greatly stimulate the proportion of ribosomes that perform the recoding event. These stimulatory signals can be 3’ or 5’ of the recoding site or both. The recoding signals located 3’ can be nearby, or distant from the recoding site, and are often in the form of intra-mRNA structures (e.g., single stem-loops or pseudoknots) that somehow influence the ribosome. There are even translation factors that are specialized to specifically interact with some of these signals. Another set of signals involves mRNA pairing with the rRNA of translating ribosomes; in the established cases, the mRNA segment involved is 5’
of the recoding site. Yet another signal can be a particular sequence of amino acids in the growing nascent peptide acting within the peptide exit tunnel of the translating ribosome. How the ribosome senses and responds to this variety of signals is still quite unclear but is now becoming amenable to study due to the major advances in knowledge of ribosome structure and an emerging understanding of ribosome conformational changes during the translation cycle.

**Redefinition.** Carboxy terminal extensions of proteins can be programmed when the meaning of a UAG or UGA stop codon is redefined so that a proportion of ribosomes accepts a near-cognate aminoacyl-tRNA, such as that charged with glutamine (for UAG) or tryptophan (for UGA) instead of a release factor. Translation then continues in the zero frame to synthesize a “readthrough” protein which often contains an additional domain or two. UGA within an open reading frame can also be redefined in a different way, to specify the non-universal, 21st amino acid, selenocysteine, often located at the crucial active site of the enzyme product. Dramatically, multiple UGAs are redefined in selenoprotein P mRNA (10 in human and apparently 28 in sea urchin) for the purpose of transporting selenium. Redefinition of the UGAs in these mRNAs is clearly programmed because it is messenger specific; other UGAs in the same cell specify termination. However, in methanogens when UAG specifies the 22nd amino acid, pyrrolysine, there may be an ambiguous reassignment of the meaning of UAG. But, the specific context of an mRNA may enhance the specification of pyrrolysine.

In the inverse of stop codon redefinition, a sense codon in a specific context can mediate termination. In the case of the StopGo (also called “Stop-Carry on”) phenomenon the specific sense codon specifies an amino acid, the protein chain is terminated, and translation continues on to make a second protein from the single ORF.

So far there is no known case of a simple programmed change in the meaning of a standard sense codon – switching one amino acid for another (though there is dynamic redefinition of an exceptional codon for tryptophan at some, but not other, positions in a particular mRNA in the ciliate *Euplotes*).

**Redirection of linear readout.** Ribosomal frameshifting links two overlapping ORFs, with a variety of mechanisms, a mix of functional results, and with a variety of mRNA-specific signals.

Most programmed frameshifting involves single nucleotide, −1 or +1 shifts (some −2 shifts are known). At least most of these cases involve a dissociation of anticodon:codon pairing, followed by tRNA:mRNA realignment and anticodon re-pairing to mRNA in a new frame (but the situation of Ty3 frameshifting in yeast appears different and in several cases of +1 frameshifting the initial pairing of the tRNA involved is not as stringent as generally occurs).

The known cases of programmed +1 frameshifting involve a slow-to-decode codon in the ribosomal A-site, either a stop codon or a sense codon for which the relevant aminoacyl-tRNA is limiting (a “hungry” codon). There is competition between the peptidyl-tRNA realigning forward and the tRNA or release factor for the zero frame A-site codon. Thus the first nucleotide of the A-site codon can be pivotal for frameshifting-mediated regulatory circuits.
Programmed $-1$ frameshifting generally yields a fixed ratio of shift to non-shift products: the product whose synthesis involved a frameshift event and the product of standard decoding. The most common type of $-1$ frameshifting involves tandem dissociation of the anticodon:mRNA pairing of tRNAs in both the P- and A-sites, followed by realignment and re-pairing of both mRNAs in the $-1$ frame, although re-pairing of only the A-site tRNA is likely to be involved in some cases.

A greatly exaggerated version of dissociation and re-pairing occurs when re-pairing of peptidyl-tRNA to mRNA occurs not at an overlapping codon but at a downstream triplet on the same mRNA, thus bypassing the mRNA sequence in-between. In the best characterized case, 50 nucleotides are bypassed by about half the ribosomes reading the message apparently due to the formation of mRNA structure within the bypassing ribosomes.

In an even more extreme case of redirection, coding resumption occurs on a specific, unique “mRNA,” tmRNA. In this case a protein, SmpB, is crucial for resume site selection. tmRNA function was initially thought to be just an elegant mechanism for rescuing ribosomes stuck at the 3’ end of aberrant mRNAs that lacked a terminator and for facilitating the destruction of the associated incomplete proteins. However, it is now apparent that tmRNA’s role is more extensive as in some cases it is involved in regulation. Also there is emerging evidence of distant 5’ nucleotide sequence in several mRNAs that influence tmRNA action.

**Examples of Function.** Many of the viruses that utilize recoding are of great medical or economic importance, and their mobile chromosomal gene counterparts have had a significant evolutionary impact. The panoply of decoding versatility and sophistication by compact genomes is common and accomplishes diverse goals. For instance, in some plant RNA viruses, frameshifting may be part of the strategy for preventing a logjam of opposing ribosomes and RNA dependent, RNA polymerase acting on the same RNA. In another example, recoding generates the retroviral GagPol polyprotein that results in the precursor form of reverse transcriptase being included in the virion by virtue of its linkage to a small proportion of Gag. This crucial linkage of Gag and Pol could also be accomplished by RNA splicing. But, this would be deleterious because the location of the RNA packaging site would result in virion packaging of subgenomic RNA yielding defective viruses.

Interestingly, the type of recoding utilized by murine leukemia virus for this purpose is programmed readthrough whereas that utilized by HIV is programmed frameshifting — two recoding solutions to the same problem.

Another case of using different types of nonstandard mechanisms to accomplish the same result is the expression of two DNA polymerase subunits from a single bacterial chromosomal dnaX gene. In *Escherichia coli*, decoding the standard ORF yields a product containing two carboxy terminal domains that are lacking in the product resulting from a ribosomal frameshift event two-thirds of the way through the ORF. This foreshortened protein likely has a role in translesion polymerase that helps deal with transition through lesions or obstacles on template DNA. Its synthesis is mediated by 50% efficient ribosomal frameshifting with ribosomes in the new frame quickly encountering a stop codon. In contrast, in *Thermus thermophilus*,
foreshortened products are derived from translation of the transcripts that result from transcriptional slippage at a run of A residues in the DNA. The population of mRNAs with varying numbers of extra nucleotides at the slippage site result in ribosome termination at now in-frame stop codons.

Evolution of recoding involves selection for both the position and the nature of the recoding site with its requisite stimulatory signals. In the absence of stimulatory signals, sites at which frameshifting or readthrough occur at low levels are, of course present. The current evidence suggests that, at least in bacteria, the most shift-prone sites that are not utilized for recoding are largely confined to poorly expressed mRNAs. For the sites whose “shifty” nature is dependent on scarcity of a particular tRNA, overexpression of an mRNA can lead to an increase in frameshifting raising a cautionary note for expression of high levels of proteins, often in nonhomologous systems, for biotechnological applications.

Scarcity of charged tRNAs can also be caused by amino acid starvation, a not uncommon state for bacteria. Starvation-induced frameshifting might be utilized to retune metabolism in response to the new growth state, so far this has not been shown.

Another consequence of recoding that needs further investigation is a possible under-appreciated role for frameshift-, bypassing-, and readthrough-derived events that do not exist to produce functional products. Ribosomes entering a region of mRNA not accessible by standard translation could have significant consequences on mRNA structure perhaps altering mRNA half-life. Alternatively, frameshifting within a coding sequence that yields early termination in a new frame could also affect mRNA half-life.

**Recoding and Human Disease.** Much remains unknown about the possible role of nonstandard translation in aging, viral infection, and certain autoimmune diseases. But the beginnings are there.

The stability of some of the proteins derived from ORFs not accessed by standard decoding is of particular interest from an immunological perspective. Preferential display on MHC class I molecules of peptides derived from short-lived proteins for activation of CD8+ T lymphocytes, this is important for the rapid CD8+ T-cell response to viral infection. Though the exact pathway for creating the array of peptides for display is not clear, models invoke rapidly degraded translation products. Some of these could be created by release of short nascent peptides due to ribosomal frameshifting.

Also, frameshifting may influence the severity of some of the triplet repeat diseases. The expanded string of repeats induces frameshifting leading to some product with poly-alanine in place of poly-glutamine.

Other genetic diseases involve frameshift mutations or substitutions that generate premature stop codons. If these new in-frame stop codons happen to be in a favorable context, small molecule drugs that alter translational fidelity can be used to phenotypically partially correct the mutations by stimulating synthesis of even a small portion of full-length product. This could alleviate the symptoms. Clinical trials in cystic fibrosis and Duchenne’s muscular dystrophy are in an advanced stage.
It may also be possible to phenotypically correct certain frameshift mutants. Compensatory frameshifting can be stimulated by supplying a small RNA molecule to create a stimulatory signal in the mutant mRNA. Additions to tissue culture cells of such an RNA to create a signal just downstream of a frameshift mutant have yielded some positive results in optimal circumstances, but delivery problems remain.

Recoding events themselves may be targets for beneficial intervention. Since the ratio of Gag to GagPol is critical for HIV propagation, the efficiency of the frameshift event required for GagPol synthesis is a target for drug development. However, success depends on the host not having crucial similar targets. This is just one of the reasons for curiosity about the number of chromosomal genes that utilize the different types of frameshifting.

Foot and mouth disease virus appears to be a case in hand where it appears that the host cell does not use the unique StopGo recoding mechanism that the virus needs for propagation. This StopGo mechanism could be a target for antiviral development.

**The path to recoding studies.** The origin of knowledge about recoding has several different threads. In the mid-1960s, it was thought that decoding was so rigidly triplet that deviations from it would not be found, i.e., compensatory leakiness of frameshift mutations would not be detectable. And it was thought that mutants of translation components which would violate triplet decoding could not be found, i.e., external suppressors for frameshift mutants would not be isolatable. By 1972, both propositions were known to be incorrect.

Later that decade, an RNA phage-encoded product whose synthesis involved a frameshift event was detected. Also the balance of WT tRNAs was shown to be important for one type of frameshifting, and the relevance of noncognate codon:anticodon interaction was recognized. Nevertheless, the impact of these studies and of the discovery of a DNA phage frameshift product in 1983 was limited.

It was not until 1985–1987 that there were big breakthroughs in the detection of the utilization of specific frameshifting for gene expression. These cases are described in this book.

Redefinition of the meaning of one of the stop codons, UGA, was first discovered in the decoding of the coat protein gene of the RNA phage Qβ in the early 1970s. A proportion of translating ribosomes read through the stop codon by inserting an amino acid at the corresponding position in the protein. Not long afterward, essential readthrough was also shown for some plant viruses to make their RNA polymerase and for murine leukemia virus to make the GagPol precursor protein. This was accepted only slowly since the discovery of RNA splicing in 1977 provided a convenient explanation for accessing alternate open reading frames.

That selenocysteine was directly encoded by specific UGA stop codons, was discovered in 1986 at approximately the same time as the discovery of the initial cases of programmed frameshifting. The common features of reprogramming led to coining of the term “recoding” in 1992.
Recoding versus Reassignment. There seems to be a clear distinction between mRNA, site-specific, reassignment of codon meaning, and the complete reassignment, as for example in certain mitochondria. However, it is usual in biology for boundaries not to be sharp. Ambiguity arises where reassignment has not been fully refined as suggested above in the case of encoding pyrrolysine by UAG codons. For instance, a codon may be especially slow-to-decode, as with AGU and AGA in certain mitochondria. Perhaps surprisingly, the effects of such a codon in a fortuitous context may make a shift-prone site. Such a case may be evident in the common ancestor of the mitochondria of birds and turtles some 200 million years ago. It is thought that an extra nucleotide was present at an internal site in the coding sequence with frameshifting at a fortuitous “shifty” site restoring essential in-frame decoding. The extra nucleotide, and its associated compensatory frameshifting, is inferred to have been lost in many of the descendents of this common ancestor except in the mitochondrial decoding of the majority of extant birds and tortoises.

A parallel situation with an extra nucleotide occurs in a proportion of tracts of nine or more as in certain AT-rich endosymbionts such as *Buchnera aphidicola* which is associated with Aphids. However, in this case, the reading frame is restored by compensatory transcriptional slippage.

In the ciliate, *Euplotes*, UGA is reassigned so that it does not specify termination. It has been proposed that coincident changes in the release factor cause UAA, especially with a 3′A, to become unusually slow-to-decode. There is efficient frameshifting at AAA UAA A in *Euplotes* and required frameshifting occurs at this “terminator” sequence in a remarkable proportion of identified genes. Together with the mitochondrial frameshifting, *Euplotes* decoding illustrates more overlap between recoding and reassignment than encountered in other organisms.

Ancient decoding. Are there any cases of redefined meaning of a codon that are actually ancestral in an evolutionary sense? Consider UGA. Since special signals are required to change the meaning of UGA to specify selenocysteine, it is easiest to consider the standard termination meaning as ancestral. However, in early decoding there may not have been discrimination between cysteine and selenocysteine and perhaps at a stage before divergence of the common ancestor of bacteria, archaea, and eukaryotes, both amino acids were specified by UGN codons. In one version of this scenario, a next step was limitation of cysteine decoding to UGU and UGC, with UGA encoding selenocysteine. As the original anaerobic atmosphere changed to an aerobic one with the advent of an oxygen-rich atmosphere some 2.4 billion years ago, there could have been selection against oxygen-labile selenocysteine except where it was especially advantageous. Perhaps this “restriction stage” is when selenocysteine-recoding signals started to arise, and non-tagged UGA codons later acquired the termination meaning. Such a model is in marked contrast to the obvious one in which the termination meaning was ancestral.

In modern bacteria UGA specifies selenocysteine only if it is followed by a specific stem-loop structure in the mRNA. It is a reasonable supposition, although no more than that, that a 3′ nearby stem-loop structure became important for selenocysteine specification in the common ancestor of bacteria, archaea, and eukaryotes.
In modern eukaryotes a specific structure in the 3’ untranslated region is required. However, some eukaryotic mRNAs that encode selenocysteine-containing proteins also have some “remnant” of a stimulatory structure just 3’ adjacent to the UGA. This element likely preceded the emergence of specific structures in the 3’ UTR.

At a much earlier time than selenocysteine specification, during the evolution of decoding itself, it seems likely that primitive readout was incapable of being anything other than slipshod. At this time polyamines may have been playing a protein-like role in primitive ribosomes. The result likely was a plethora of products serving as food for selection. As triplet decoding and codon assignment became locked in, was there a parallel refinement of alternative decoding? Or did the currently observed alternative decoding evolve later as a sophisticated refinement after a period of tediously standard decoding?

Frameshifting for expression of bacterial release factor 2 decoding also has an ancient origin. Its hallmark is stimulation of the frameshift event by pairing between mRNA and rRNA during translation. We can wonder whether this interaction between mRNA and rRNA in ribosomes in the act of translating might not itself have an ancient origin. Could interactions of this type have helped to grip the message?

In modern day ribosomes, it is anticodon pairing that holds the mRNA in place. Detachment and realignment lead to frameshifting, at least in most cases. There is an appealing if somewhat controversial suggestion that standard frame maintenance is maintained by pairing two tRNAs at all times. In this scenario, anticodon pairing by E-site tRNA does not dissociate until A-site aminoacyl-tRNA pairing is established. So strong ribosomal gripping of tRNA would lead to the in-frame grip of the mRNA. However, the E-site appears to be a late addition in ribosome evolutionary history since it is protein-rich. Therefore, before it existed, what served to clasp mRNA? One candidate is the rRNA:mRNA Shine–Dalgarno pairing which was discovered because of its role in initiation of protein synthesis in bacteria. Programmed frameshifting studies have revealed that this interaction is not unique to initiation in that the anti-Shine–Dalgarno sequence of translating ribosomes can scan the mRNA being decoded for potential complimentarity. After such a rRNA:mRNA hybrid forms, the ribosome continues translation for up to 10 nucleotides before the hybrid ruptures. Whether interactions of this type played a role in primordial protein synthesis is of course unknown. But, if so, rather than the primordial coding sequences having been G-rich, perhaps there could have been blocks of coding sequences spanned by G-rich noncoding “anchors” that decoding could bypass. Setting aside such speculative “excesses,” recoding studies are clearly contributing to our knowledge of standard decoding and scanning by the anti-Shine–Dalgarno sequences of translating ribosomes is one of several cases in point.

**Transcription slippage (also called pseudo-templated transcription or stuttering)**

Realignment during transcription parallels translational realignment. A few examples are mentioned above where transcription slippage substitutes for cases of programmed frameshifting. In these cases there has been selection for high-level
transcription slippage at specific sites. Such slippage yields mRNAs with inserts of one or more nucleotides – in a bacterial case a diminishing series of mRNAs with up to 15 additional nucleotides and a small minority with deletions of one or a few nucleotides. Standard translation of these mRNAs yields unique products. Instead of the detachment of triplet anticodon pairing, dissociation of the nascent RNA hybrid with template DNA in the transcription bubble is involved. The identity of flanking sequence can delimit the number of extra nucleotides inserted to 1. But whether the flanking sequence can also enhance the frequency, possibly even by the ability of the nascent RNA chain to form a short stem, remains to be seen.

Editing of preformed transcripts can also have consequences similar to several types of recoding. For instance, mRNA editing that changes a stop codon to a sense codon can give the equivalent of stop codon readthrough. Similarities even extend to variable efficiencies of the process and to the importance of mRNA structure. Editing to change the identity of one sense codon to another in a proportion of the mRNAs, constitutes a type of diversity for which there is only one specialized recoding counterpart. It will be fascinating to discover to what extent nonstandard transcription and RNA editing parallel and substitute for their translational counterparts.

**Future.** As this book attests, our knowledge of recoding has a firm basis but much remains to be done. Together with studies of mutants of ribosomal components, advances in structural information about translation components now are offering the prospect of an understanding of how ribosomes sense and respond to recoding signals. The deluge of sequence information is providing exciting bioinformatic opportunities for comparative analyses to reveal the extent of recoding and transcription slippage. And a dramatic recent advance in determining ribosome location en masse at sub-codon resolution by sequencing vast numbers of mRNA segments protected within ribosomes at a specific time, has great potential in this regard.

Knowledge of the “dark matter” of the genome, those transcribed regions that do not encode mRNA, tRNA, or rRNA, is rapidly showing the complex roles of small RNAs in gene expression. Are some cases of recoding influenced by them?

We look forward to discovering the answers to these and questions not yet asked.

**Acknowledgment:** We thank Ken Keiler for instigating the American Society of Microbiologists’ session that inspired Andrea Macaluso (Springer) to propose this book, our past colleagues, especially Bob (R.B.) Weiss and Alan Herr, and our current colleagues. We also thank Marshal Nirenberg, the pioneer of codon identification, for his generous contribution.

John Atkins and Ray Gesteland
Recoding: Expansion of Decoding Rules Enriches Gene Expression
Atkins, J.F.; Gesteland, R.F. (Eds.)
2010, XX, 466 p. 156 illus., 63 illus. in color., Hardcover