

# This Week's Citation Classic®

Kozak M. Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell* 44:283-92, 1986; and Kozak M. An analysis of 5' noncoding sequences from 699 vertebrate mRNAs. *Nucl. Acid. Res.* 15:8125-48, 1987.

[Department of Biological Sciences, University of Pittsburgh, PA]

A survey of 5' noncoding sequences from vertebrate mRNAs revealed a consensus motif around AUG initiator codons. Within the GCCACCAUGG motif, the most highly conserved nucleotides were a purine, most often A, in position -3 (three nucleotides upstream from the AUG codon) and G in position +4, following the AUG codon. Systematic mutagenesis around the translational start site in preproinsulin mRNA revealed that recognition of the AUG codon was strongly dependent on A<sub>3</sub> and G<sub>+4</sub>. Translation was moderately affected by mutations in the remainder of the consensus sequence. [The *SCF*<sup>®</sup> indicates that these papers have been cited in more than 1,570 and 1,060 publications, respectively.]

## Identifying AUG Initiator Codons

Marilyn Kozak  
Department of Biochemistry  
University of Medicine and  
Dentistry of New Jersey  
Piscataway, NJ 08854

During my first job interview, my announced intention to study how eukaryotic ribosomes recognize the AUG initiator codon met with the response that "Joan Steitz has already done it." What Steitz had already done brilliantly, of course, was to figure out how prokaryotic ribosomes identify authentic translational start sites.<sup>1</sup> However, my early work with Aaron Shatkin<sup>2</sup> had suggested that the mechanism of initiation of translation in eukaryotes might be fundamentally different from prokaryotes. The studies described in my two *Citation Classic*<sup>®</sup> papers defined and tested a consensus sequence for translational start sites in vertebrate mRNAs. Other studies, summarized in recent reviews,<sup>3</sup> revealed that neither the AUG codon nor the consensus sequence constitutes a direct entry site for ribosomes. Rather, in striking contrast with prokaryotic systems, eukaryotic ribosomes apparently enter at the 5' end of the mRNA and then migrate down to the AUG codon, which is most effective as a "stop signal" when it is flanked by GCCACC...G.

These papers are widely cited because they offer guidance for predicting translational start sites in newly sequenced genes. Knowledge of the consensus sequence has also been helpful in constructing vectors for the efficient expres-

sion of cloned cDNAs. cDNA sequences in which the context around the first AUG codon is extremely weak (e.g., pyrimidines in positions -3 and +4) often turn out to be incomplete; the real initiator codon lies upstream from the artificially truncated 5' end of the cDNA. So, knowledge of the translation-initiation context rules has encouraged some to search for the missing N-terminal coding portion of the gene in question.

In rare mRNAs, mostly of viral origin, the first AUG triplet occurs in a very weak context. In such cases ribosomes initiate at the first and second AUG codons, producing two independently initiated proteins.<sup>4</sup> The postulated dependence of this "leaky scanning" on context has been confirmed in a few labs by showing that mutations that improve the context around the first AUG codon suppress production of the second protein. This is one way in which the context rules have been verified.

Occasionally the rules have been misapplied. Because eukaryotic ribosomes reach the AUG initiator codon by scanning from the 5' end of the mRNA, initiation occurs not at the AUG codon that best matches the full consensus sequence but at the first AUG codon in an adequate context. For example, if the context around the first AUG codon includes A or G in position -3 and G in position +4, the first AUG triplet is likely to be the unique site of initiation even if the rest of the sequence differs from the consensus and even if there is a perfect consensus sequence farther downstream.

The primary sequence around the AUG codon is a major, but not the sole, determinant of initiation site selection. Other structural features in vertebrate mRNAs that affect translation initiation are reviewed elsewhere.<sup>5</sup>

Getting these *Citation Classics* into print was not easy. One referee's judgment was that the detailed study of point mutations that affect translation "may not be of compelling interest to the broad audience of *Cell*." The editor of *Nucleic Acids Research* wondered who, if anyone, would use the results of the survey. If my rebuttals to *Cell* and *Nucleic Acids Research* carried the day, it's probably because I had been practicing rebuttals since that first job interview with Jacques Fresco at Princeton. The job I actually landed was at the University of Pittsburgh, where these studies on context effects were conducted.

1. Steitz J A & Jakes K. How ribosomes select initiator regions in mRNA: base pair formation between the 3' terminus of 16S rRNA and the mRNA during initiation of protein synthesis in *Escherichia coli*. *Proc. Nat. Acad. Sci. USA* 72:4734-8, 1975. (Cited 530 times.)

2. Kozak M & Shatkin A J. Sequences of two 5'-terminal ribosome-protected fragments from reovirus messenger RNAs. *J. Mol. Biol.* 112:75-96, 1977.

3. Kozak M. A consideration of alternative models for the initiation of translation in eukaryotes. *Crit. Rev. Biochem. Molec. Biol.* 27:385-402, 1992.

4. ----- . An analysis of vertebrate mRNA sequences: intimations of translational control. *J. Cell Biol.* 115:887-903, 1991.

5. ----- . Structural features in eukaryotic mRNAs that modulate the initiation of translation.

*J. Biol. Chem.* 266:19867-70, 1991.

Received August 10, 1993