Gene 432 (2009) 1–6

Contents lists available at ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Review

# Evolution and the universality of the mechanism of initiation of protein synthesis

Tokumasa Nakamoto

*Department of Biochemistry and Molecular Biology, The University of Chicago, 5801 S. Ellis Ave., Chicago, IL 60637, USA*

ARTICLE INFO

*Article history:*
Received 4 August 2008
Received in revised form 30 October 2008
Accepted 2 November 2008
Available online 8 November 2008

Received by A.J. van Wijnen

*Keywords:*
Cumulative specificity mechanism in the initiation of protein synthesis
Multisubstrate enzyme system in the initiation of protein synthesis
Non-unique initiation signals in protein synthesis
Initiation mechanism for eukaryotic protein synthesis
Evolutionarily conserved initiation signals in protein synthesis
Universal initiation mechanism for protein synthesis

ABSTRACT

The main mechanisms advanced to account for the specificity of the initiation of protein synthesis are reviewed. A mechanism proposed by Shine and Dalgarno (SD), focused on the base pairing of a unique leader sequence in the initiation site—the SD sequence—with the 3′ end of the 30S ribosomal RNA as the only step necessary for selecting the initiation site in prokaryotes. Studies showed, however, that the SD interaction is not obligatory and protein synthesis can occur even in its absence. In fact, comparison of a large number of initiation site sequences revealed that the sites are composed of diverse combinations of preferred bases, and, thus, the apparatus that is able to recognize all these sites is *de facto* a multisubstrate enzyme system. As such, it has the hallmarks of the cumulative specificity mechanism, and the SD interaction, when present, is only one of a number of contributing factors in the selection of the initiation site.

The cumulative specificity mechanism proposed that secondary structure selectively interdicts access to most of the non-initiator methionine codons while leaving open the true initiation site and that the final recognition of the initiation site occurs by cooperativity and cumulative specificity of the several ligand recognition sites of the ribosomes, which confer broad substrate specificity to the system. This mechanism appears to be universal; it can encompass the initiation of all protein syntheses since it is consistent with all the salient observations on the initiation of both eukaryotic and prokaryotic protein syntheses. Studies of eukaryotic/prokaryotic hybrid systems further strengthen this conclusion: They show that the prokaryotic initiation signals are evolutionarily conserved in the eukaryotic mRNAs, since prokaryotic ribosomes are able to translate eukaryotic mRNAs. Conversely, eukaryotic ribosomes also recognize prokaryotic initiation signals and initiate synthesis, indicating that the eukaryotic ribosomes may have also conserved the prokaryotic initiation mechanism. The universality of a single process of protein synthesis in all kingdoms is also manifest in the conservation of a complex apparatus, consisting of ribosomes, mRNA's, tRNA's including an initiator methionyl-tRNA, aminoacyl tRNA synthetases, and other protein factors. Thus, the mechanism of initiation of protein synthesis is conserved, and it is universal.

The third initiation mechanism is the scanning mechanism for eukaryotes. It proposes that the 40S ribosome-methionyl-tRNA complex recognizes and binds to the 5′-end of the mRNA and the complex then scans the messenger for the initiator codon. Once it is located, the 80S ribosome initiation complex is formed with the 60S subunit and initiation is completed when a second aminoacyl-tRNA is bound and a peptide bond is formed. Exceptions to this mechanism were observed, where the ribosome bound directly to internal mRNA sites and initiated synthesis. Consideration of the conflicting observations in this review, however, has led to the conclusion that the primary eukaryotic mechanism is a conserved prokaryotic mechanism and that the "scanning process" involves two steps. The first step is an interaction of the initiation factors with the cap, which makes the IS accessible, and the second, initiation of translation by the conserved prokaryotic mechanism.



## 1. Introduction

Two characteristics of the protein synthesizing system and evolution provide major insights into the nature of the mechanism of the initiation of protein synthesis. The first characteristic is the fact that the initiator codon is not unique, which suggests that the mRNA should contain some initiation signal or a way to differentiate the initiator methionine codon from those, otherwise identical codons, that code for methionine located internally in the protein. Such an initiation signal could be, for example, a unique base sequence in the proximity of the initiator codon. It has actually been proposed that the prokaryotic initiation signal is a special leader sequence preceding the initiator codon, which is referred to as the Shine–Dalgarno (SD) sequence (Shine and Dalgarno, 1974). An alternative is that the signal is not a unique base sequence but, rather, consists of an ensemble of

*Abbreviations:* SD, Shine–Dalgarno; IS, initiation site; mRNA, messenger RNA; IRES, internal ribosome entry site; IRESs, IRES RNA structure or site.

*E-mail address:* tokumasa@sbcglobal.net.

0378-1119/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2008.11.001

preferred bases. Such a signal would then be recognized by a mechanism appropriate to handle multiple substrates, that is, one that has broad substrate specificity. If this is the case, then the protein synthesizing system is *de facto* a multisubstrate enzyme system (Nakamoto, 2007). Another alternative is the recognition of the 5′ end of the mRNA instead of the recognition of an initiation signal (Kozak, 1978). The second enlightening characteristic of the protein synthesizing system is that the mRNA interacts extensively with itself to form secondary structures. These secondary structures must be controlled sufficiently to keep the initiator codon accessible to the ribosome since the accessibility of the initiator codon is the first, *sine qua non* condition for initiation. In evolution, the conservation in all kingdoms of the major components of the apparatus for synthesizing proteins suggests that the initiation mechanism may also be conserved and is universal.

These two characteristics of the protein synthesizing system that illuminate the mechanism of initiation and evolution should be taken into consideration in formulating the initiation mechanism. This will lead one to focus on the identity of the initiation signal and the manner of its recognition, on the regulation of secondary structure to keep the initiator codon accessible to the ribosome, and to consider the possibility of a universal mechanism. In this review, the three leading mechanisms of the initiation of protein synthesis, the SD, the cumulative specificity, and the scanning mechanisms are reviewed. The way in which each of the three leading mechanisms interprets the two characteristics of the protein synthesizing system and evolution in formulating its respective mechanism is examined.

## 2. Initiation signal

In order to delineate the structural properties of the initiation signal, a model IS was computer generated from 68 *Escherichia coli* ribosome binding site sequences (Scherer et al., 1980). The model IS contains 47 nucleotides and shows a preference, rather than a specific base in any given position. Surprisingly, nearly half of the site includes amino acid codons. The authors suggested that this explains how leaderless mRNAs may be recognized by the ribosomes, the amino acid codons in the initiation site also having recognition features. Perhaps the most important revelation of the IS model is that there is no unique initiation sequence, rather, the ISs consist of a collection of varied sequences of preferred nucleotides. In other words, the prokaryotic initiation sites constitute a large multiplicity of loosely related substrates and the protein synthesizing system, which recognizes all of them, is a multisubstrate enzyme system.

In eukaryotes, the initiation sites of protein synthesis exhibit similar properties: A careful analysis of 211 leader sequences of eukaryotic ISs yielded only a consensus sequence of eight bases without any base in a specific position (Kozak, 1984). Thus, eukaryotes do not appear to have any unique initiation sequence either and the results suggest again, that the eukaryotic protein synthesizing system is also a multisubstrate enzyme system with broad substrate specificity.

## 3. Secondary structure and accessibility

In typical enzymatic reactions, the selective recognition of the structural features of the enzyme's specific substrate(s) is a result of the direct stereospecific interaction of the subsites of the enzyme with the respective subsites of the reactants—substrates, cofactors, inhibitors, or activators. In the case of protein synthesis, however, the macromolecular reactant, the mRNA, interacts so extensively with itself to form secondary structures that a new element is introduced into the recognition of molecular features in the reaction. The new element is the accessibility of the molecular features of the reactants. Specifically, in the case of protein synthesis, it is the accessibility of the initiator codon and the IS of the mRNA to the ribosomes. Accessibility may be generally an important factor in subsite recognition in enzymatic reactions where the reactant is a macromolecule, e.g. a nucleic acid or a protein that may have a secondary, tertiary, or quaternary structure.

If the neighboring secondary structural interactions are extensive, accessibility to a methionine codon in RNA could be virtually eliminated. Because of the prevalence of secondary structures in mRNA, accessibility to the initiator methionine codon could become an important—and perhaps even the only—determining factor in selection from among several potential reaction sites. Based on our unpublished observation that synthetic RNA with four bases in equal proportions and in random sequence did not have any significant messenger function for polypeptide synthesis, we proposed that all non-initiator methionine codons in natural mRNA were sequestered by secondary structure and only initiator codons were accessible to the ribosomes (Kolakofsky and Nakamoto, 1966; Nakamoto and Vogel, 1978). This inference was strengthened by subsequent reports, which showed that as much as 50–60% of the nucleotides of synthetic RNAs containing all four bases in random sequence are base-paired (Gralla and Delisi, 1974; Ricard and Salser, 1975). Natural mRNA is even more highly ordered with 60–70% base pairing and it even has well-defined tertiary structure (Ricard and Salser, 1976).

A later observation, however, indicated the need for a modification of the proposal that accessibility is the sole determinant factor for the selection of the IS. In that study, a model mRNA without the SD sequence was created (Calogero et al., 1988). The model mRNA was designed to minimize secondary structure; it had an accessible AUG triplet but no other apparent IS recognition signal. The model mRNA was able to act as an effective messenger for polypeptide synthesis: It initiated synthesis by starting with the AUG codon, in agreement with the proposed accessibility criterion, but, surprisingly, a second, equally unhindered and thus supposedly accessible AUG failed to act as an initiator. It was thus necessary to conclude that the bases surrounding the unreactive AUG interfered with the functional accessibility through other local, negative interactions. The mechanistic model incorporating this additional feature was named the unique accessibility hypothesis (Nakamoto, 2006). The model states that, in addition to those non-initiator methionine codons that are clearly sequestered by secondary structure, the other, yet accessible non-initiator methionine codons are also made functionally inactive by unfavorable local interactions of their surrounding nucleotides with the ribosomal binding site. For example, such interactions may consist of steric hindrance, hydrophobic/hydrophilic mismatch, or electrostatic repulsion. These factors thus constitute an important facet of the selection mechanism; they all contribute to a negative specificity.

## 4. Evolution and universal initiation mechanism

Among the major cellular processes of transferring information, which are replication, transcription and translation, only the latter process has significant evolutionary conservation of its components and mechanisms. This is not surprising since the process of translation is a very complex biochemical pathway in which a multitude of reactions are so intricately coordinated and interconnected that even a small change may result in a severe disruption. It is then not surprising that the essential components and mechanisms of the process would be evolutionarily conserved. The protein synthesizing apparatus in all kingdoms has similar basic components, consisting of ribosomes, mRNAs, tRNAs (including an initiator methionyl-tRNA), aminoacyl tRNA synthetases, and other protein factors. Universally conserved are: The ribosomes as the central synthesizing organelle with two subunits and with binding sites for peptidyl-tRNA, aminoacyl-tRNA, and for mRNA. The mRNA is the working copy of the genetic information that is to be translated. All the information in the mRNA is encoded in a universal code. The important tRNAs that bridge the recognition gap between the nucleotides and the amino acids and the equally essential aminoacyl tRNA synthetases are also conserved.