Available online at www.sciencedirect.com
SCIENCE DIRECT
Gene 361 (2005) 13–37
GENE
SECTION
FUNCTIONAL GENOMICS
www.elsevier.com/locate/gene

Review

Regulation of translation via mRNA structure in prokaryotes and eukaryotes

Marilyn Kozak *

Department of Biochemistry, Robert Wood Johnson Medical School, 675 Hoes Lane, Piscataway, NJ 08854, USA

Received 25 April 2005; received in revised form 31 May 2005; accepted 27 June 2005
Available online 5 October 2005
Received by A.J. van Wijnen

Abstract

The mechanism of initiation of translation differs between prokaryotes and eukaryotes, and the strategies used for regulation differ accordingly. Translation in prokaryotes is usually regulated by blocking access to the initiation site. This is accomplished via base-paired structures (within the mRNA itself, or between the mRNA and a small *trans*-acting RNA) or via mRNA-binding proteins. Classic examples of each mechanism are described. The polycistronic structure of mRNAs is an important aspect of translational control in prokaryotes, but polycistronic mRNAs are not usable (and usually not produced) in eukaryotes. Four structural elements in eukaryotic mRNAs are important for regulating translation: (i) the m7G cap; (ii) sequences flanking the AUG start codon; (iii) the position of the AUG codon relative to the 5′ end of the mRNA; and (iv) secondary structure within the mRNA leader sequence. The scanning model provides a framework for understanding these effects. The scanning mechanism also explains how small open reading frames near the 5′ end of the mRNA can down-regulate translation. This constraint is sometimes abrogated by changing the structure of the mRNA, sometimes with clinical consequences. Examples are described. Some mistaken ideas about regulation of translation that have found their way into textbooks are pointed out and corrected.

Keywords: Ribosome binding site; Protein synthesis; Scanning model; AUG; Repressor protein; Reinitiation

1. Introduction

This review focuses on the initiation phase of protein synthesis—in particular, on regulatory mechanisms built into the structure of the mRNA.

Initiation of translation in prokaryotes is mediated by three protein factors, designated IF1, IF2 and IF3. Eukaryotic initiation factors are more numerous (eIF1, eIF1A, eIF2, eIF2B, eIF3, eIF4A, eIF4E, eIF4G, eIF5, eIF5B) and some of these play important regulatory roles (Harding et al., 2001; van der Knaap et al., 2002). One essay cannot cover everything, however, and the initiation factors will be discussed herein only incidentally. Other reviews do an adequate job of explaining the functions of these proteins (Dever, 2002; Kapp and Lorsch, 2004; Laursen et al., 2005; von der Haar et al., 2004).

Other reviews might be consulted also regarding some important emerging stories, such as temporal control of translation during embryonic development (Kuersten and Goodwin, 2003), translation-linked degradation of defective mRNAs (Baker and Parker, 2004), and regulation of translation by microRNAs (Bartel, 2004; Yang et al., 2005). Here, I have focused on mechanisms that are more fully defined.

Regulation of translation is not limited to the initiation step, of course. At the level of elongation, the most common regulatory device involves frameshifting (Márquez et al., 2004; Matsufuji et al., 1995; Namy et al., 2004). Other interesting regulatory mechanisms are built around the pausing of ribosomes at a particular point in elongation (Chartrand et al., 2002; Mason et al., 2000; Murakami et al., 2004; Rüegsegger et al., 2001; Snyder et al., 2003).

With those acknowledgments concerning what the review omits, here is a preview of what it includes. Section 2 discusses aspects of prokaryotic mRNA structure that are important for initiation in general. Section 3 describes specific examples of translational regulation in bacteria and bacteriophage. The unit

Abbreviations: Csr, carbon storage regulator; IF, initiation factor; eIF, eukaryotic IF; IRE, iron-response element; IRES, internal ribosome entry site; IRP, iron-response protein; LEF-1, lymphoid enhancer factor-1; ORF, open reading frame; RBS, ribosome binding site; SD, Shine–Dalgarno sequence; TMV, tobacco mosaic virus; TPO, thrombopoietin; TRAP, *trp* RNA-binding attenuation protein; upORF, upstream regulatory ORF; UTR, untranslated region.

* Tel.: +1 732 235 5355; fax: +1 732 235 5356.
E-mail address: kozakma@umdnj.edu.

0378-1119/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2005.06.037

on eukaryotes begins with an overview of mRNA structures relevant to initiation (Section 4), followed by examples of regulation via reinitiation (Section 5) and mRNA binding proteins (Section 6). Section 7 addresses some common questions and misunderstandings about initiation of translation in eukaryotes. Section 8 traces some of the misunderstandings to recurrent problems in the design and execution of experiments.

2. Structural elements in prokaryotic mRNAs that control initiation of translation

mRNA sequences are numbered by designating the A of the AUG codon as +1. The preceding base is position −1 and negative numbering proceeds upstream.

2.1. AUG (or other) start codon

Selection of the AUG or alternative start codon by the 30S ribosomal subunit sets the reading frame for the rest of the translation process. AUG is recognized via pairing with the anticodon (3′-UAC-5′) in fMet-tRNA (Mayer et al., 2003). Structural analyses of initiation complexes help to explain why other tRNAs cannot be used in this step (Allen et al., 2005).

Weaker pairing (two rather than three base pairs) with fMet-tRNA is part of the reason that translation is less efficient when an alternative start codon replaces AUG. In one study, translation was reduced ~8-fold when AUG was replaced by GUG or UUG (Sussman et al., 1996, Table 3). Notwithstanding this reduction in efficiency, 14% of *Escherichia coli* genes use GUG as the start codon and another 3% use UUG (Blattner et al., 1997). Use of UUG as a start codon is more common in Gram-positive bacteria and some bacteriophage (Kunst et al., 1997;Łobocka et al., 2004).

AUU functions even less efficiently than UUG in experimental tests (Sussman et al., 1996), and AUU is the natural start codon in only two *E. coli* genes. One of these encodes a potentially toxic protein, which explains why translation must be restrained (Binns and Masters, 2002). The other encodes initiation factor IF3. This factor has the interesting function of proofreading initiation complexes; i.e. IF3 disfavors initiation at nonstandard start codons, as evidenced by increased initiation at AUU, AUA and CUG codons when IF3 is mutated (O'Connor et al., 2001; Sussman et al., 1996). This leads to the prediction that IF3 mRNA should be translated better when IF3 protein levels are low, which is indeed the case (Butler et al., 1987). Evolutionary conservation of this autoregulatory mechanism underscores its importance (Hu et al., 1993). In addition to functioning as a fidelity factor for selection of the start codon, IF3 also promotes dissociation of 70S ribosomes, generating the pool of free 30S subunits required for initiation.

2.2. SD element and nearby sequences

The RNA component (16S rRNA) in the 30S ribosomal subunit plays a major role in selecting the translational start site. Authentic start codons are preceded by a purine-rich sequence which is complementary to, and base pairs with, a sequence near the 3′ end of 16S rRNA (Jacob et al., 1987; Steitz and Jakes, 1975). This so-called Shine–Dalgarno (SD) sequence in mRNA is typically 4 or 5 nt in length. (It can be as long as 8 nt or as short as 3 nt, if two of the three base pairs are G–C. The mRNA/rRNA complementarity must not be interrupted by unpaired bases.) An exhaustive analysis of *E. coli* genes documents the existence of an SD sequence in all but a few exceptional cases (Shultzaberger et al., 2001).

The SD is usually positioned some 5–8 nt upstream from the start codon.[1] The optimal spacing depends on exactly which bases at the 3′ end of 16S rRNA (3′-AUUCCUCCAC…5′) participate in the interaction (Chen et al., 1994a). Spacing is clearly important, as evidenced by cases in which unused AUG codons occur between the SD and the actual start codon. The spacing requirement can be rationalized by structural models of the ribosome which show the P site, where AUG binds, on the interface side of the 30S subunit while the "anti-SD sequence" in 16S rRNA is around the corner, on the solvent side (Yusupova et al., 2001).

In most mRNAs, the standard 4 or 5 base pair SD interaction is strong enough to mediate efficient translation. Thus, experimentally lengthening the SD sometimes produces no increase in translation (Munson et al., 1984) or only a modest increase (Chen et al., 1994b, construct IF6) or even a diminishment (de Smit and van Duin, 1994a; Komarova et al., 2002). A stronger-than-normal SD interaction does help, however, when the start codon is not AUG (Weyens et al., 1988) or when the initiation site is masked by secondary structure (de Smit and van Duin, 1994a; Munson et al., 1984). On the latter point, the clearest evidence comes from an evolutionary study with coliphage MS2 in which expansion or abbreviation of the SD provoked compensatory changes in the strength of a hairpin structure that encompasses the ribosome binding site (Olsthoorn et al., 1995).

Whereas the presence of secondary structure within the initiation region can be offset by a stronger-than-normal SD sequence, an A/U-rich initiation site that forms no stable secondary structure might require no SD interaction at all (Fargo et al., 1998). Thus, an A-rich, G-poor leader sequence derived from tobacco mosaic virus (TMV) which augments initiation when transposed to bacterial mRNAs (Gallie and Kado, 1989) might do nothing more than preclude secondary structure. The unusually weak SD in ribosomal protein S1

[1] Feltens et al. (2003) describe an unusual case in which a single SD (GGAGG) precedes two functional AUG codons. The sequence is cagGGAGGgagAUGgAUG, wherein the first AUG initiates RNase P and the second AUG initiates ribosomal protein L34. The postulated dual use of an SD is not certain, however, as an upstream AGG sequence (underlined) is better positioned to function as the SD for the first AUG. Thus, the hypothesis requires testing. Another deviation-from-the-norm was postulated for ribosomal protein S1 mRNA (Boni et al., 2001). Here, an SD located far upstream is supposedly brought close to the AUG codon by an array of hairpin structures. The authors invoke phylogenetic conservation as evidence for the model, but in some species the predicted hairpins are very weak (mostly A–U and G–U base pairs). The model was actually tested only with *E. coli* S1 mRNA, where some but not all mutations produced the expected effects.