Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Biochimie (1992) 74, 357–362

Recognition of correct reading frame by the ribosome

EN Trifonov

Department of Structural Biology, The Weizmann Institute of Science, Rehovot 76100, Israel

(Received 17 October 1991; accepted 23 January 1992)

Summary — The translation frame-monitoring mechanism has been suggested earlier, based on transient complementary contacts, between mRNA and rRNA. Recent studies related to the frame-monitoring mechanism are reviewed. The mechanism is well supported by both new experimental and sequence analysis data. Experiments are suggested for further elucidation of the structural details of the mRNA-rRNA interaction in the ribosome.

frame monitoring / mRNA-rRNA interaction / ribosome

Introduction

Mechanistically, mRNA translation into a protein is a process of sequential translocation of the ribosome towards the 3'-end of the mRNA in standard steps of three bases. The initial frame of the translocation is provided by an initiation triplet, usually AUG, while all other coding triplets follow downstream at distances multiple of three bases. This canonical scheme is well illustrated by many known cases of the frameshift mutations, when small deletions or insertions of sizes indivisible by three change the frame of the 3-base translocations. The downstream sequence is read in wrong frame and translated into the wrong amino acid sequence. It also leads to premature termination of the process, with an encounter of a stop-codon. There are cases, however, when the system somehow becomes unstable in that frame, the ribosome counts one base more or one base less, finding itself in the correct frame again [1]. In other words, the ribosome corrects the effects of the textual change, frameshift mutation, by physically shifting on the mRNA molecule, making an unusual step of one or two bases – a translational frameshift.

The question is: what is there in the mRNA sequence, that is recognized by the ribosome as the framing signal, synchronized with correct reading frame, so that the ribosome keeps that frame despite the textual frameshifting?

Several years ago one particular frame-monitoring mechanism was suggested based on the analysis of mRNA and rRNA sequences [2]. The mRNA sequences have been shown to carry well pronounced 3-base periodical motif $(GHN)_n$ where H stands for non-G and N for any base. Ribosomal RNA of the small subunit, on the other hand, namely those parts of it that are believed to interact with mRNA, carry the complementary periodical pattern $(NNC)_n$. There are at least three such sites in the small subunit rRNA: 5'-gcCagCagCcgCgguaau, 5'-guacaCacCgcCcguca and 5'-gauCacCucCuua centered at positions 525, 1400 and at the 3'-end (position 1535), respectively (the numbers correspond to coordinates along 16S rRNA of *E coli*). The $(NNC)_n$ sites have been assumed to neighbor one another in the mRNA-rRNA complex in the ribosome, making together a kind of rack-gear of repeating NNC units, with which the mRNA makes imperfectly complementary contact involving the frequently repeating motif $(GHN)_n$ by keeping the first position G's of the coding triplets always opposite to C's of the frame-monitoring rack-gear.

This model of involvement of both mRNA and rRNA in the frame keeping gained substantial support recently. In what follows some new experimental and theoretical evidence is reviewed, which provides also new details of the frame-monitoring mechanism.

Interactions of the rRNA framing sites with mRNA

The 1400 site was first found in close vicinity of mRNA passing through the ribosome [3,4]. Its inter-

action with mRNA is well documented [5,6]. Involvement of two other sites, however, has been firmly established only recently.

The 525 site was found to be exposed to interaction with complementary probes [7,8] and mapped by DNA hybridization electron microscopy on the back side of the neck of the small ribosomal subunit [8]. A 51-base long mRNA fragment has been cross-linked to the 525 site, as well as to the 1400 site [6]. Recent observation on the tertiary interaction of the bases 524–526 of the 525 loop with the bases 505–507 of the nearby bulge [9] indicates that, perhaps, a conformational 'switch' occurs in this region of the 16S rRNA structure [9]. Obviously, only the open loop alternative, without the tertiary contact, can take part in the interaction with the mRNA.

The 3'-end of the 16S rRNA is known to be involved at the translation initiation stage, by making complementary contact with the Shine-Dalgarno sequence upstream from AUG initiation triplet [10,11]. Its involvement in elongation was first indicated by mRNA sequence analysis [12]. It was found that the 1535 site shows clear above average complementarity to mRNA in one of its three frames. That such complementary contact, indeed, exists during elongation was proven by introducing point mutations into the 1535 sequence and mRNA sequence which resulted in well predictable changes in efficiency of translational frameshifting in one particular case [13].

Proximity of the 525, 1400 and 1535 sites to one another

One would expect the synchronizing sites in the rRNA to be in a juxtaposition to be able to make one continuous or almost continuous C-periodical pattern. Their close proximity to one another is also suggested by the overall size of the mRNA-rRNA contact. Indeed, the minimal estimate of this size is given by total length of the $(NNC)_n$ rack with its $10 \times 3 = 30$ bases. This coincides with the estimated length of mRNA involved in the contact with the ribosome [14,15]. The physical length of the contact is, minimally, about $30 \times 3.4\ A \approx 100\ A$. Since the diameter of the small subunit neck is about 65 A [16], this would mean that the mRNA passes largely around the neck and, perhaps, cuts through it. This is consistent with the observation that the 3'- and 5'-ends of the mRNA fragment engaged in the ribosome meet on the same side of the neck [17]. The experimental evidence on the spatial disposition of the sites in the body or on the surface of the 30S subunit is rather controversial and usually of low accuracy since much of the evidence is obtained by electron microscopy [8].

Analysis of the sites in the small ribosomal subunit that are protected by tRNA, in combination with cross-linking data, indicated that the sites 525 and 1400 are located on opposite sides of the neck, separated by at least 65 A [16]. Modelling of 3-D-folding of the 16S rRNA in the ribosome, by taking into account all protein-protein, RNA-RNA and RNA-protein distances known or, rather, estimated, also leads to the same conclusion on the substantial separation between the 525 and 1400 sites [18]. Worth noting is, however, that the total length of the 525 and 1400 sites put together end-to-end in form of the double-helical imperfect complementary structure made of contacting mRNA and rRNA, is minimally about $21 \times 3.4\ A \approx 71\ A$. This spans the estimated large distance between the sites.

By using a complementary probe to the 525 site, its position on the surface of the 30S subunit has been determined by immune electron microscopy [19]. The site was found on the right side of the neck, which is close enough to the decoding 1400 site, in the base of the cleft, to be considered as its immediate neighbor. DNA hybridization electron microscopy of the small ribosomal subunits, with biotinylated DNA probes, locates the two sites close to one another as well [8,20], but in different positions: 525 site on the back of the neck and 1400 site on the left side of it [21], rather than in the cleft [22,23].

A recent cross-linking study using mRNA analogues with photoreactive thio-U residues in various positions demonstrated conclusively that the 525 and 1400 sites are in close vicinity to one another, being cross-linked to the same short piece of model mRNA in the elongation 70S complex [24]. The 525 site is found in contact with mRNA, several bases downstream from the triplet interacting with the decoding site. By the same technique but with several thio-U residues in one mRNA fragment, the cross-linking was detected of the 3'-end of rRNA to mRNA few bases upstream from AUG triplet in the initiation 30S complex [25]. This new technique of cross-bridging by pieces of RNA [24,25] appears to be the most appropriate for determination of the spatial relationships among the rRNA-mRNA binding sites. It not only shows that both the 525 site and 1535 site are within a few bases of the 1400 site, but it also potentially allows to determine exact distances and the order of the sites along the mRNA, both at the initiation and at the elongation stages. Whatever the order is, the experiments provide conclusive support for one of the main suppositions of our frame-monitoring scheme: the close proximity of the framing sites 525, 1400 and 1535 to one another within the ribosome.

Involvement of the 1535 site in the translational frameshifting

The idea that the site 1535, the anti-Shine-Dalgarno sequence, would be involved in the frame-keeping

process [2] was the most vulnerable part of the hypothesized frame-monitoring scheme. The site is known as one of the key elements of the translation initiation process. It does bind complementarily to the Shine-Dalgarno sequence immediately upstream of the initiation triplet, but nothing was known about its possible binding to the coding part of the mRNA, except for one indication from the mRNA sequence analysis [12].

This second involvement of the site had been recently established by elegant experiments where concerted point changes were introduced in mRNA and rRNA sequences [13]. Complementarity of the 1535 site to the sequence 8 bases upstream from the translation shift site in a certain mRNA construct was found crucial for the efficiency of the translational frameshifting. Remarkably, in that complementary position the framing cytosines of the 1535 site (gauCacCucCuua) were found in register with the first triplet positions of the new frame. Apparently, the translational frameshifting in this case was caused by the complementarity of the mRNA sequence (in the new frame) to the C-periodical 1535 site, in accordance with its role in the hypothetical frame-monitoring mechanism.

On the distributional nature of the translational frameshifting

Like in many other cases of biomolecular recognition and unlike the unequivocally specific interaction between DNA and restriction enzymes, the choice of the right reading frame by the ribosome is of a distributional recognition nature [26]. That is, there are many structural elements in the ribosome and in mRNA as well as additional factors securing all together the reading frame or its change when required. Individual contributions of these 'signal' components are different in every particular case. Sometimes only one or two of them are sufficient.

In particular, translational frameshifting can be caused by 'shifty' tRNA [27], by runs of the same base [28], by downstream mRNA hairpins [29] or tertiary mRNA folding [30], by the 3-base periodical pattern in the mRNA [2], or by combinations thereof. In every case one or another factor prevails, depending on particular sequence requirements unrelated to the frame-monitoring function. Thus, there is no necessity for the simultaneous presence of all of these 'signal' components together in every case of the frameshifting. The distributional nature of the framekeeping function explains why, for example, the G-periodicity sometimes is not present [2]. Notwithstanding these exceptions, the 3-base periodicity appears to be the factor most frequently involved in the translational frameshifting.

Components of the 3-base periodicity

The periodical signal carried by the protein-coding sequences is itself of distributional nature. Indeed, the stability of the presumed imperfect complementary contact between mRNA and rRNA depends on at least three different sequence features corresponding to basic physical components of the interaction. First, these are isolated complementary base oppositions G·C, A·U and G·U flanked by the non-complementary ones. Second, the stability of the contact is substantially higher when neighboring bases in mRNA are complementary to the respective two bases of rRNA. In this case strong stacking interaction between the neighboring base pairs is the stabilizing factor. Last, but also essential, is the destabilizing contribution of non-complementary base oppositions, especially of the similar-named ones: A·A, C·C, G·G and U·U [31]. For the mRNA-rRNA contacts to be sufficiently stable in the correct frame at every stage of the 3-base translocations, the number of the single complementary contacts should be higher and/or the number of complementary dinucleotides should be higher and/or the number of the bad mismatches should be lower than in the other two frames. The role of the first factor is reflected, in particular, by the frequent occurrence of G in the first positions of the triplets [2]. Special analysis with the emphasis on the possible role of the dinucleotides, excluding the mononucleotide background, resulted in detection of the excessive dinucleotides UG in the third positions of the mRNA triplets and CU in the second positions (Lagunez-Otero, Trifonov, submitted). This allows to express original degenerate pattern $(GHN)_n$ carried by mRNA in more detailed form: $(GCU)_n$. The same pattern is generated when avoidance of the bad base oppositions is taken into account [32]. Thus, the predominant 3-base periodical motif of the mRNA, reconstructed from the components of the presumed distributional framing recognition of mRNA by rRNA, is 5'-$(GCU)_n$. Remarkably, this pattern, extracted from mRNA sequences, without any *a priori* information on ribosomal RNA structure, turned out to be almost perfectly complementary to the proofreading 525 site (Lagunez-Otero, Trifonov, submitted) as shown in figure 1. This is the only such site in the 16S rRNA, complementary to the $(GCU)_n$ mRNA pattern. The mRNA sequence, therefore, carries in it not just the framing 3-base periodicity, but information about the rRNA structure as well. This also points to an exceptional role of the universally conserved 525 site in the ribosome function(s). The mRNA sequence analysis potentially could lead, perhaps, to a complete description of the sequence structure of the mRNA-rRNA contact, including all the contact sites and their relative positions and distances along mRNA, as soon as they are reflected in the mRNA sequences .

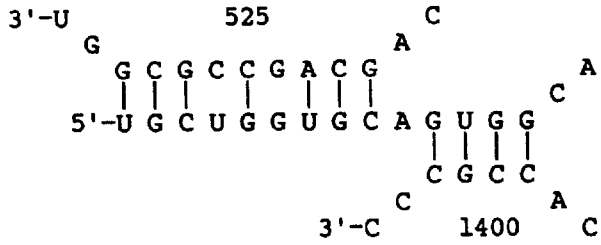

Fig 1. Complementarity of mRNA periodical motif $(GCU)_n$ to the 525 site. The framing cytosines are indicated by dots.

Compensation effects

The decomposition of the 3-base periodical pattern of mRNA in the three components of the distributional recognition mRNA-rRNA as described above implies also their compensatory relationships. In particular, if the first position G of the originally suggested pattern $(GHN)_n$, and second position H (non-G) are considered as two distributional signal elements, then in the mRNA sequences lacking GNN griplets, the triplets nHn should be overrepresented, to compensate the lack of another component of the pattern. That is, if there are not enough guanines in the first triplet positions, their avoidance in the second positions should become even stronger. This, indeed, is found to be the case [32]. The avoidance of G in the second positions is explained by the predominant occurrence of G in the second positions of 3'-$(CNN)_n$ sequence of the main 525 framing site, since the opposition G·G appears to be one of the most destabilizing base oppositions in RNA structure [31].

Another interesting example of such compensation is the human alpha-fibrinogen gene that lacks the canonical G in the first positions of the triplets of a large repeat region of the gene [2, 33]. It turned out that the dinucleotide CU of the second positions of the triplets is overrepresented in this case, apparently to keep the mRNA pattern $(gCU)_n$ sufficiently complementary to the 525 site [32].

The compensation effects of the distributional recognition can be generally utilized for consecutive extraction of unknown signal elements from the sequence, as a signal detection and purification technique. Indeed, if in a given set of recognition sites the normally present signal features are underrepresented, it should be compensated by overrepresentation of other signal elements, undetected earlier.

Suggested experiments

Informative as they are, the computational approaches to the study of the mRNA-rRNA interactions can and should be complemented by direct experiments. A straightforward experiment would be to use RNA probes complementary to the framing sites, as competitors for mRNA. Initiation of translation and, perhaps, elongation are expected to be inhibited by the probes. One known observation can be interpreted this way. A short RNA sequence was discovered that strongly inhibits translation *in vivo* [34]. The sequence carries a strong periodical pattern $(GNN)_n$ and, thus, should bind well to the framing sites. In figure 2, two possibilities are shown. The small RNA possesses significant complementarity to both sites, 525 and 1400. Perhaps, one of these contacts, or both, are actually responsible for the inhibitory activity of the small RNA.

This example also suggests the way to experimentally determine relative positions and distances between the rRNA contact sites along the mRNA. For this purpose probes could be synthesized carrying simultaneously sequences complementary to two different framing sites . By varying positions of the complementary sequences along the probe one could, thus, determine the spatial relationships between the sites in the ribosome.

Similarly, by taking the sequences with no complementarity to the rRNA framing sites, and introducing the sequences in mRNA, one can design the especially 'shifty' loci in the designed mRNA. The frameshifting at these loci is expected to be even more efficient if the sequences immediately downstream would contain a strong $(GCU)_n$ motif in a new frame. The best designs, presumably, would be the ones that carry several such non-complementary sequences reflecting the spatial relationships between the contact sites. Perhaps, the most 'shifty' sequences would be the ones that carry the segment 5'-UAAUGGCGCCGACGACCG anticomplementary to the main 525 framing site. Here the similar-named bases are chosen to oppose the bases of the 525 site all along the antiparralel contact, to make it maximally unstable [31].

Concluding remarks

The frame-monitoring mechanism, as proposed earlier, is well supported by both recent experiments and

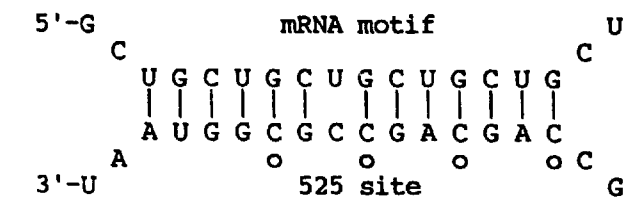

Fig 2. Possible contacts of small cytoplasmic RNA [35] with the framing sites 525 and 1400.

nucleotide sequence analysis. Its main proposition, on the transient complementary contacts between mRNA and rRNA, is confirmed for one of three framing sites, site 1535 [13]. It remains to be demonstrated directly that the two other sites make as well the complementary contacts with mRNA. Analysis of mRNA sequences and detection of the universal pattern complementary to the 525 site makes this site a good candidate as well. All three sites are found in close vicinity to one another [24,25] as suggested by the framing mechanism. The relative positions and actual distances among the sites along mRNA remain to be determined, and straightforward experiments are suggested to work this out.

The framing mechanism is more than just a design to maintain correct reading frame. Indeed, the three framing sites, 525, 1400 and 1535 carry, as well, the most important functions of proofreading, decoding and initiation of translation, respectively. Further detailed studies of the translational frameshifting and frame-keeping mechanism, therefore, are of direct relevance to studies on the structural basis and mechanism of translation in general.

References

1 Atkins JF, Weiss RB, Gesteland RF (1990) Ribosome gymnastics – degree of difficulty 9.5, style 10.0. *Cell* 62, 413–423

2 Trifonov EN (1987) Translational framing code and frame-monitoring mechanism as suggested by the analysis of mRNA and 16s rRNA nucleotide sequences. *J Mol Biol* 194, 643–652

3 Prince JB, Taylor BH, Thurlow DL, Ofengand J, Zimmermann RA (1982) Covalent crosslinking of tRNAval1 to 16S RNA at the ribosomal P site: identification of crosslinked residues. *Proc Natl Acad Sci USA* 79, 5450–5454

4 Ciesiolka J, Nurse K, Klein J, Ofengand J (1985) Conversation of RNA sequence and cross-linking ability in ribosomes from a higher eukaryote: photochemical cross-linking of the anticodon of P site bound tRNA to the penultimate cytidine of the UACACACG sequence in *Artemia* saline 18S rRNA. *Biochemistry* 24, 3233–3239

5 Stiege W, Stade D, Schuler D, Brimacombe R (1988) Covalent crosslinking of poly(A) to *Escherichia coli* ribosomes and localization of the cross-link site within the 16S RNA. *Nucleic Acids Res* 16, 2369–2388

6. Wollenzien P, Expert-Bezancon A, Favre A (1991) Sites of contact of mRNA with 16S rRNA and 23S rRNA in the *Escherichia coli* ribosome. *Biochemistry* 30, 1788–1795

7 Lasater LS, McKuskie Olson H, Cann PA, Glitz DG (1988) Complementary oligonucleotide probes of RNA conformation within the *Escherichia coli* small ribosomal subunit. *Biochemistry* 27, 4687–4695

8 Oakes MI, Lake JA (1990) DNA-hybridization electron microscopy. Localization of five regions of 16S rRNA on the surface of 30S ribosomal subunits. *J Mol Biol* 211, 897–906

9 Powers T, Noller HF (1991) A functional pseudoknot in 16S ribosomal RNA. *EMBO J* 10, 2203–2214

10 Shine J, Dalgarno L (1974) The 3'-terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. *Proc Natl Acad Sci USA* 71, 1342–1346

11 Gold L, Pribnow D, Schneider T, Shinedling S, Sincer BS, Stormo G (1981) Translational initiation in prokaryotes. *Annu Rev Microbiol* 35, 365–403

12 Sedlacek J, Fabry JM, Rychlik I, Volny D, Vitek A (1979) The arrangement of nucleotides in the coding regions of natural templates. *Mol Gen Genetics* 172, 31–36

13 Weiss RB, Dunn DM, Dahlberg AE, Atkins JF, Gesteland RF (1988) Reading frame switch caused by base-pair formation between the 3'-end of 16S rRNA and the mRNA during elongation of protein synthesis in *Escherichia coli*. *EMBO J* 7, 1503–1507

14 Takanami M, Zubay G (1964) An estimate of the size of the ribosomal site for messenger RNA binding. *Proc Natl Acad Sci USA* 51, 834–839

15 Kang C, Cantor CR (1985) Structure of ribosome-bound messenger RNA as revealed by enzymatic accessibility studies. *J Mol Biol* 181, 241–251

16 Moazed D, Noller HF (1986) Transfer RNA shields specific nucleotides in 16S ribosomal RNA from attack by chemical probes. *Cell* 47, 985–994

17 Evstafieva A, Shatsky IN, Bogdanov AA, Semenkov Y, Vasiliev VD (1983) Localization of 5' and 3' ends of the ribosome-bound segment of template polynucleotides by immune electron microscopy. *EMBO J* 2, 799–804

18 Stern S, Weiser B, Noller HF (1988) Model for the three-dimensional folding of 16S ribosomal RNA. *J Mol Biol* 204, 447–481

19 Lasater LS, Montesano-Roditis L, Cann PA, Glitz DG (1990) Localization of an oligodeoxynucleotide complementing 16S ribosomal RNA residues 520-531 on the small subunit of *Escherichia coli* ribosomes: electron microscopy of ribosome-cDNA-antibody complexes. *Nucleic Acids Res* 18, 477–485

20 Oakes ML, Kahan L, Lake JA (1990) DNA-hybridization electron microscopy. Tertiary structure of 16S rRNA *Mol Biol* 211, 907–918

21 Oakes ML, Clark MW, Hendreson E, Lake JA (1986) DNA hybridization electron microscopy: ribosomal RNA nucleotides 1392–1407 are exposed in the cleft of the small subunit. *Proc Natl Acad Sci USA* 83, 275–279

22 Keren-Zur M Boublik M, Ofengand J (1979) Localization of the decoding region on the 30S *Escherichia coli* ribosomal subunit by affinity immunoelectron microscopy. *Proc Natl Acad Sci USA* 76, 1054–1058

23 Gornicki P, Nurse K, Hellman W, Boublik M, Ofengand J (1987) High resolution localization of the tRNA anticodon interaction site on the *Escherichia coli* 30S ribosomal subunit. *J Biol Chem* 259, 10493–10498

24 Rinke-Appel J, Junke N, Stade K, Brimacombe R (1991) The path of mRNA through the *Escherichia coli* ribosome: site-directed cross-linking of mRNA analogues carrying a photo-reactive label at various points 3' to the decoding site. *EMBO J* 10, 2195–2202

25 Dontsova O, Kopylov A, Brimacombe R (1991) The location of mRNA in the ribosomal 30S intiation complex: site-directed corsslinking of mRNA analogues carrying several photoreactive labels simultaneously on either side of the AUG start codon. *EMBO J* 10, 2613–2620

26 Trifonov EN (1983) Sequence-dependent variations of B-DNA structure and protein-DNA recognition. *Cold Spring Harbor Symp Quant Biol* 47, 271–278

27 Atkins JF, Gesteland RF, Reid BR, Anderson CW (1979) Normal tRNAs promote ribosomal frameshifting. *Cell* 18, 1119–1131

28 Weiss RB, Dunn DM, Atkins JF, Gesteland RF (1987) Slippery runs, shifty stops, backward steps, and toward hops: –2, –1, +1, +2, +5 and +6 ribosomal frameshifting. *Cold Spring Harbor Symp Quant Biol* 52, 687–693

29 Jacks T, Madhani HD, Masiarz FR, Varmus HE (1988) Signals for ribosomal frameshifting in the Rous sarcoma virus gag-pol region. *Cell* 55, 447–458

30 Brierley I, Digard P, Inglis SC (1988) Characterization of an efficient coronavirus ribosomal frameshifting signal: requirement for an RNA pseudoknot. *Cell* 57, 537–547

31 Trifonov EN (1985) Imperfect complementarity and RNA structure. *J Biosci* Suppl 8, 781–790

32 Lagunez-Otero J, Trifonov EN (1992) Translation framing pattern in mRNA–compensation effects. *In: Modelling and Computer Methods in Molecular Biology and Genetics* (Ratner V, ed) Nova Science Publishers, New York, in press

33 Kant JA, Lord ST, Crabtree GR (1983) Partial mRNA sequences for human A-alpha, B-beta, and gamma fibrinogen chains: evolutionary and functional implications. *Proc Natl Acad Sci USA* 80, 3953–3957

34 Lorberboum H, Digweed M, Erdmann VA, Servadio Y, Weinstein D, De Groot N, Hochberg AA (1986) Small cytoplasmic RNAs from human placental free mRNPs. Structure and their effect on *in vivo* protein synthesis. *Eur J Biochem* 155, 279–287