



The ribosome as an exon detector

Luis Mendoza, Jaime Lagunez-Otero*

Instituto de Química, UNAM, Ciudad Universitaria, Coyoacán, C.P. 04510, México, D.F., México

Received 31 May 1995; revision received 17 July 1995; accepted 24 July 1995

Abstract

We present an analysis where different regions of small subunit rRNA are tested for their capacity to discriminate between coding and non-coding regions. We find that the 530 loop is one of the best sequences to be used in an algorithm for detecting exonic sites.

Keywords: Ribosome; Exon detector; Exonic sites

1. Introduction

The protein translation mechanism is a complex one, the ribosome alone is a structure made up of more than 50 protein and three RNA molecules. Even though it is not clear how every ribosomal component is involved in translation, some understanding of the role of rRNA is arising (Raué et al., 1990). For example, it is recognised that rRNA is central in the alignment of mRNAs (Shine and Dalgarno, 1974), and it is believed that base pairing is crucial in determining mRNA orientation within a ribosome (McKuskie et al., 1988).

Ribosomal RNA has some phylogenetic conserved regions. We are particularly interested in the 530 loop (3'-uggcgccgac-5') of the SS rRNA, which makes contact with mRNAs (Dontsova et

al., 1991, 1992) and has been implicated in proofreading and frame maintaining during protein translation (Trifonov, 1987; McKuskie et al., 1988; Lagunez-Otero, 1993; Powers and Noller, 1994). Moreover, that region has the periodical sequence 3' (C-G-N)_n 5', which is very suitable for base pairing with the periodical consensus pattern 5'(G-C-U)_n 3' found in mRNAs coding regions (Trifonov 1987; Lagunez-Otero and Trifonov 1992).

We have presented a profile of interactions between region 530 and mRNAs along pre-translation, translation and post-translation zones (Mendoza et al., 1995), showing that the 530 loop is a sequence which contains characteristics that allow for discrimination between coding and non-coding regions, start and stop signals, as well as distinguishing the correct reading frame. Our present interest is to determine if it is feasible to design an exon detection algorithm based on the scanning of nucleotidic sequences using the 530 loop.

Abbreviations: PSI, potentially stable interaction; SS rRNA, small subunit ribosomal RNA

* Corresponding author.

2. Methods

2.1. Nearest-neighbour method for determining ΔG°_{37} for duplex formation

Freier et al. (1986) devised a method that provides standard free-energy changes needed for oligonucleotide helix initiation and propagation. Briefly, ΔG°_{37} of an oligonucleotide interaction is obtained by adding up: (1) the free energy change associated with formation of the first base pair in an oligonucleotide duplex; (2) a sum of propagation free energies caused by the formation of each subsequent base pair; and (3) applying a symmetry correction value when paired sequences are fully complementary.

2.2. Sequences

We used a set of 200 mRNAs, representing the biological kingdoms extracted from the GenBank (Bilofsky et al., 1986); sequences were taken from base -59 to base +60 with respect to the origin of translation. Three fragments of the human SS rRNA were used: 3'-uggcgccgac-5' (530 loop), 3'-cagcagcggc-5', which will be referred to as the best (discriminating) sequence from here on, and 3'-ccugcggcuc-5' named as the worst (discriminating) sequence here. We are interested in the 530 region because it has a biological role in protein translation (see introduction), whilst the other two are sequences that have contrasting interaction values, in the sense that they give extreme scores when tested for interactions with mRNAs (Mendoza et al., 1995); hence, we make use of them for comparison with the 530 sequence.

2.3. mRNA-rRNA interactions

First, ΔG°_{37} for helix formation between each of our three rRNA sequences and 200 mRNAs were calculated. Secondly, all three sets of results were smoothed out, substituting each value for the average of the previous four and following four scores in the same reading frame. Finally, ΔG°_{37} values that represent stable interactions (negative numbers) were quantified and presented in the form of histograms (see Figs. 1–3); different

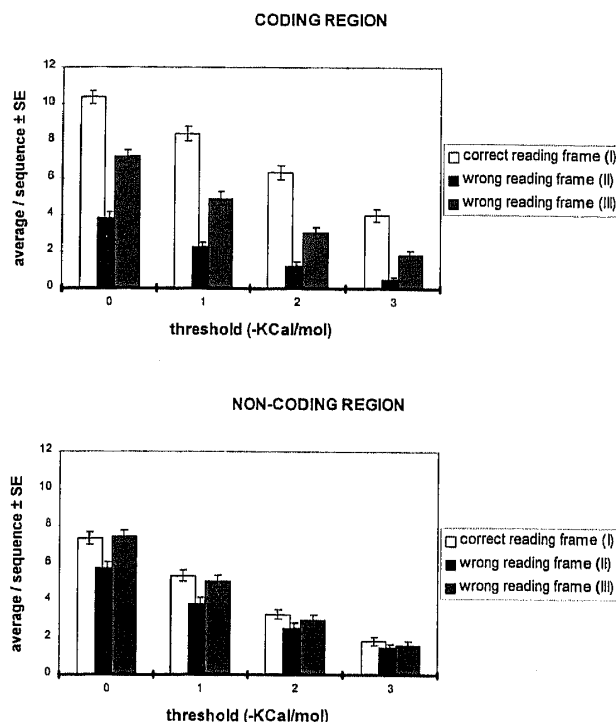


Fig. 1. ΔG°_{37} for helix formation between sequence 3'-cagcagcggc-5' of the SS rRNA, and 200 mRNAs were calculated. Then values were smoothed out, substituting each original value for its average with the previous and following four scores in the same reading frame. Finally, for each messenger, those ΔG°_{37} 's that represent stable interactions (negative numbers) were quantified. Average of those values and its standard error measured are presented.

thresholds were used to see if interaction patterns changed or not.

2.4. Exon detection

We wanted to exploit results of mRNA-rRNA contacts to evaluate the ribosome's capacity for exon detection. In order to do this we analysed interactions with messengers one by one. First, interactions of a given messenger was calculated using our three rRNA sequences of interest. Then, as with the previous test, negative values were quantified, and results were combined to obtain a criterion for exon presence (see results).

3. Results

In this section, we deal first with mRNA-rRNA favourable interactions, those with a negative

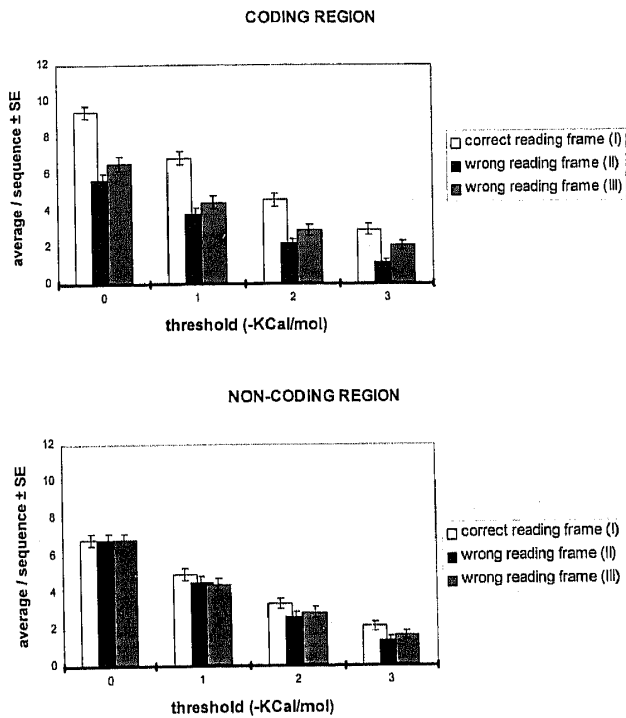


Fig. 2. ΔG_{37}° for helix formation between sequence '3-ugggcccgcac-5' (530 loop) of the SS rRNA, and 200 mRNAs were calculated. Then values were smoothed out, substituting each original value for its average with the previous and following four scores in the same reading frame. Finally, for each messenger, those ΔG_{37}° 's that represent stable interactions (negative numbers) were quantified. Average of those values and its standard error measured are presented.

delta standard free energy. For that, we quantified values below some negative threshold in order to see the distribution of the putative stable interactions. Then, in the second part of the work, we made use of the obtained interaction patterns to analyse the exon detection capacity of our three SS rRNA fragments.

3.1. Interactions

Here, favourable interactions were quantified. This must be said because of the mRNA size of our database and data processing; for each messenger there are 28 possible interactions in each frame, half in the coding region and half in the non-coding; thus the maximum possible value is 14 for any bar in the figures. The average interaction values for the 200 mRNAs using the best SS rRNA sequence are presented in Fig. 1. Looking

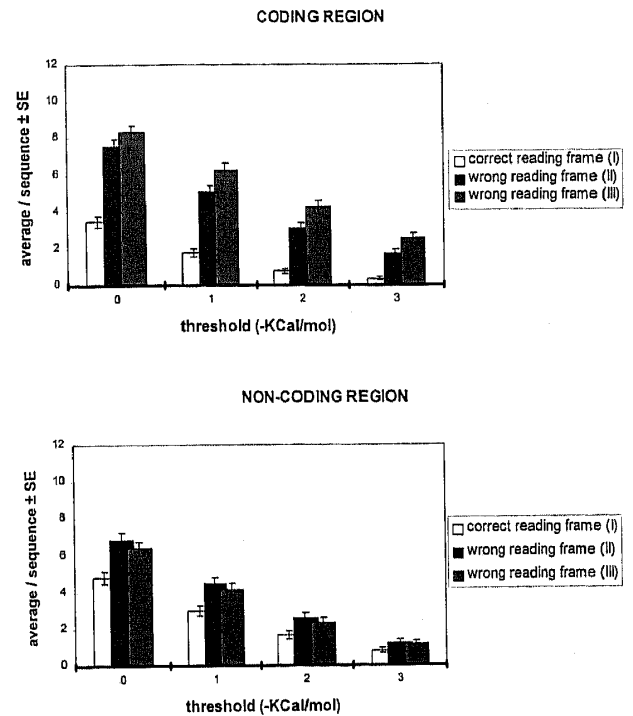


Fig. 3. ΔG_{37}° for helix formation between sequence '3-ccugcggcuc-5' of the SS rRNA, and 200 mRNAs were calculated. Then values were smoothed out, substituting each original value for its average with the previous and following four scores in the same reading frame. Finally, for each messenger, those ΔG_{37}° 's that represent stable interactions (negative numbers) were quantified. Average of those values and its standard error measured are presented.

at the coding region (upper panel), some features are prominent. First, high complementarity scores in the correct reading frame are found much more frequently than in the other two, yielding an average of ten high scores (out of 14 possible) for each messenger. Besides, reading frame II is clearly the one that would be less likely to interact with the ribosomal segment. This trend is almost the same no matter what threshold is used for quantification.

The non-coding region (lower panel) presents a different pattern. For symmetry, interactions are divided into three frames, as in the coding region, but that division is useful for comparison only. Note that differences among the frames are little; in one case there was none at all; the pre-translation messenger region, therefore, does not contain a clear signal for framing. Furthermore, there are less interactions here than in the coding

region, meaning that rRNA binds preferably to the coding zone in the correct reading frame. Again, the trend is maintained using different thresholds.

Fig. 2 shows an analogous study, this time using the 530 loop sequence. Clearly, it has a similar profile to that obtained using the best sequence (Fig. 1), thus making a strong support for the claim that this biologically relevant region is very close to the optimal potentially interacting sequence (Mendoza, et al. 1995). Besides, interactions with messengers at the non-coding region show no difference among any of the three reading frames.

Finally, as a comparison, we used another sequence that was the worst potentially interacting in a previous study (Fig. 3). The most striking characteristic of this figure is the very low interaction profile of the correct reading frame, contrasting with former figures, having even lower values than those corresponding to reading frame I in non-coding regions in all three figures. Interestingly, both wrong reading frames have slightly more interactions with this sequence than those with the best or 530 regions. Besides, in the non-coding zone, the general aspect is the same, although differences among reading frames are shorter than in the coding region.

3.2. Exon detection

In the last section, we obtained a clear pattern in which the correct reading frame is clearly interacting largely with both 530 and the best sequences, while doing it poorly using the worst sequence. While this happened in the coding region, the non-coding region did not show that trend; thus, it seems that these interactions can be used as a criterion (or part of it) to distinguish between coding and non-coding regions in a given nucleotide sequence.

As a first step, for each messenger in our database, interactions with the 530 sequence were calculated; then, values were filtered and sorted into those coming from coding and non-coding regions. Later, all values were divided into three frames, though without specifying which was the right one, and which was not. Then, for both regions, the frame with more potentially stable interactions (PSI) was pointed out. Finally, we ap-

plied a criterion in which if the frame with more PSIs obtained using the 530 region was the same with less PSIs using the worst sequence, then that region contained a coding sequence.

Applying that criterion to all of the messengers of our database, we obtained a 61.5% yield of true positives, that is, actual coding regions detected as such with the above stated principle. This may not seem quite distinct from a random assignment of being a coding region or not, thus we used a slight variation. We used the three rRNA sequences; we postulated that if the frame with less PSIs using the worst sequence coincided with the frame with most PSIs using either the best or the 530 sequences, then that was a coding region. Now, the true-positives yield rised to 70% of our database.

A detection rate of 70% is an encouraging result, because we could exploit with some success the mRNA-rRNA interaction patterns. Clearly, these results indicate that there is a signal inside the coding region that differentiates it from the non-coding; this signal is strong enough to be noticed with the above experiments, but it is weak enough to be used as an infallible exon detector, at least when using interactions only with small rRNA sequences.

4. Discussion

We think that any device capable of detecting exons in DNA sequences would be quite helpful for molecular biologists, which presently are accumulating a huge amount of data coming out from genome projects. Because of our biological interests, we turned our attention to a small part of the actual protein translation mechanism, and evaluated its capacity for exon detection.

Results give a good insight for a way to tackle the problem down. As seen in Figs. 1 and 2, there is a notable difference between interaction patterns of coding and non-coding regions with rRNA sequences, the latter having stronger interactions with coding zones. But even more, there is a large preference of the correct reading frame over the wrong ones. This clear pattern depends on the rRNA fragment used, however; but we have shown that the 530 loop is one of the most appropriate for such a task.

It should be emphasised that the presented pro-

mising results, were obtained using only ten base-long rRNA sequences, too short compared with the actual biological machinery used for exon detection. Even though we had a 70% of true positives, that figure is not good enough to be used for sequence analysis. Our work points to a way of constructing an automated exon detector using known and putative molecular interactions. Thus, in order to obtain a high reliability in exon detection, perhaps some interactions not contemplated here should be taken into account; for example, those among proteins, tRNAs and mRNAs.

Finally, we want to state that the analysis here presented considers two aspects, and that one reinforces the other. On the one hand, we are proposing a first advance on how a biological mechanism can be used to our advantage in the automated study of nucleotide sequences. And on the other hand, we are trying to understand the role of rRNA-mRNA interactions in the protein translation mechanism, via computer simulations.

Acknowledgement

This work was partially supported by project DGAPA # HIN30293 and grant CONACYT 2117-m9303

References

- Bilofsky, H.S., Burks, C., Fickett, J.W., Goad, W.B., Lewitter, F.I., Rindone, W.P., Swindell, C.D. and Tung, C.S., 1986, The GenBank genetic sequence data bank. *Nucleic Acid Res.* 14(1), 1–4.
- Dontsova, O., Kopylov, A. and Brimacombe, R., 1991, The location of mRNA in the ribosomal 30S initiation complex, site-directed cross-linking of mRNA analogues carrying several photo-reactive labels simultaneously on either side of the AUG start codon. *EMBO J.* 10, 2613–2620.
- Dontsova, O., Dokudovkaya, S., Kopylov, A., Bogdanov, A., Rinke-Appel, J., Junke, N. and Brimacombe, R., 1992, Three widely separated positions in the 16S RNA lie in or close to the ribosomal decoding region, a site-directed cross-linking study with mRNA analogues. *EMBO J.* 11, 3105–3116.
- Freier, S.M., Kierzek, R., Jaeger, J.A., Sugimoto, N., Caruthers, M.H., Neilson, T. and Turner, D.H., 1986, Improved free-energy parameters for predictions of RNA duplex stability. *Proc. Natl. Acad. Sci. USA* 83, 9373–9377.
- Lagunéz-Otero, J. and Trifonov, E.N., 1992, mRNA periodical infrastructure complementary to the proof-reading site in the ribosome. *J. Biomol. Struct. Dynam.* 10, 455–463.
- Lagunéz-Otero, J., 1993, rRNA-mRNA complementarity: Implications for translation initiation. *Trends Biochem. Sci.* 18, 406–408.
- McKuskie Olson, H., Lasater, L.S., Cann, P.A. and Glitz, D.G., 1988, Messenger RNA orientation on the ribosome. *J. Biol. Chem.* 263, 15196–15204.
- Mendoza, L., Mondragón, M. and Lagunéz-Otero, J., 1995, The role played by the 530 ribosomal site in translocation — interaction with the termination region of mRNA. Submitted for publication.
- Powers, T. and Noller, H., 1994, The 530 loop of 16S rRNA: a signal to EF-Tu? *Trends Genet.* 10(1), 27–31.
- Raué, H., Musters, W., Rutgers, C., Van't Riet, J. and Planta, R., 1990, rRNA: from structure to function; in: *The ribosome, structure, function and evolution*, R.A. Garrett, P.B. Moore, D. Schlessinger and J.R. Warner (eds.) (American Society for Microbiology, Washington, DC).
- Shine, J. and 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.
- Trifonov, E.N., 1987, Translation framing code and frame-monitoring mechanism as suggested by the analysis of mRNA and 16S rRNA nucleotide sequences. *J. Mol. Biol.* 194, 643–652.