COMPUTER ANALYSIS OF mRNA UNTRANSLATED REGIONS OF HYPOXIA-INDUCED CORN GENES

1 Titov I.I., 1 Kochetov A.V., 1 Kolchanov N.A., 2 Sarai A.

1 Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia, e-mail: titov@bionet.nsc.ru
2 RIKEN Institute, Tsukuba, Ibaraki 305-0074, Japan
*Corresponding author

Key words: computer analysis, mRNA, translation, 5'UTRs, 3'UTRs, secondary structure

Resume

Motivation: Untranslated regions (UTRs) of mRNA frequently house determinants of translational activity. UTR variability and ambiguous localizations of these signals in the sequence in combination with their intricate compositions, including elements of primary and secondary structures, underlie the difficulties arising while detecting translation determinants. Searching for such signals requires a combination of contextual analysis techniques with search for invariant elements of the secondary structure.

Results: Two groups were distinguished among the analyzed mRNA UTRs of hypoxia-induced corn (Zea mays) genes. Typical of the first group are longer 5'UTRs with a pronounced secondary structure that brings into spatial proximity the cap site and site of translation initiation. 3'UTRs of this group contain more AUG triplets compared with the second group. 5'UTRs of the second group are shorter and lack a stable secondary structure. Presumably, an intergroup distinction between mechanisms of translation regulation underlies this segregation of mRNAs with respect to the contextual and structural characteristics.

Introduction

Signals affecting mRNA translational activity could play an important role in gene expression control. The majority of cellular 5'UTRs are scanned by ribosomes from the cap site to the site of translation initiation. Length of the leader sequence, false translation initiation sites, secondary structures, or proteins binding to this region may modulate this process. 3'UTRs also contain quite a few determinants of translational efficiency; however, mechanisms of their actions are yet vague. Thus, translation determinants may be present over the entire length of UTRs, whose length amounts from several dozens to several thousands nucleotides, may be degenerate, and depend strongly on one another. These signals have complex structure, are highly diverged, and dispersed within sequences, challenging conventional bioinformatics methods to reveal them. A successful approach to the problem may consist in developing computer methods that combine analysis of sequence homology and alignments with secondary structure calculations and its comparison. We present an example of such an integrated analysis for revealing mRNA-located translational signals.

Methods

UTRs of hypoxia-induced genes were extracted from database on translational enhancers (Kochetov et al., 2001; http://wwwmgs.bionet.nsc.ru/mgs/dbases/trsig/). To localize the candidate mRNA regions capable of forming non-randomly stable structures and to refine the secondary structure translational signals, we used the algorithm GArna (Vorobiev et al., 2002; http://wwwmgs.bionet.nsc.ru/Programs/2dStructRNA/). The final evolutionary invariant model was constructed using the program GenBee (http://www.genebee.msu.su/genebee.html).

Results and Discussion

5'UTRs. The sample of 5'UTRs comprised six sequences. These sequences distinctly fell into two groups with reference to the stability of their secondary structure and their lengths: the first group appeared to form stable structure, while second did not (Table). Moreover, stability of each sequence belonging to the second group was inferior compared to the typical stability for the sequence of the same nucleotide composition (see Z-score value). Interestingly, all the mRNA 5'UTRs studied so far by different authors (see Titov et al., 2002 for review) displayed the secondary structure that was at least equal in its stability to that of the shuffled random sequences. It is likely that the regions from the second group had undergone a selection directed against stability of the secondary structure.

Further search for invariant secondary structure involved only the three sequences of the first group. The sequences in question were aligned with additional weights ascribed to invariant complementary pairs. Consequently, a model of the secondary structure common for the entire group was constructed (Fig.).
Table. Sequence length and secondary structure energy of hypoxia-induced mRNA leaders.

<table>
<thead>
<tr>
<th></th>
<th>mzesus1b</th>
<th>mzeadh1cm</th>
<th>zmadh2n</th>
<th>zmaldoar</th>
<th>zmenola</th>
<th>zmsucs2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Length, nt</td>
<td>123</td>
<td>107</td>
<td>126</td>
<td>74</td>
<td>54</td>
<td>71</td>
</tr>
<tr>
<td>Energy, kcal</td>
<td>−19.3</td>
<td>−19.5</td>
<td>−16.8</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Z-score</td>
<td>−0.98</td>
<td>−0.59</td>
<td>0.57</td>
<td>0.33</td>
<td>0.62</td>
<td>0.63</td>
</tr>
</tbody>
</table>

An interesting specific feature of this structure is a spatial proximity of the cap site and the site of translation initiation. This proximity may be involved in an interdependent regulation of binding of translation factors with these sites. This regulation may involve both the inhibition (for example, via a simple shielding of one site when binding to the other) and stimulation provided by an additional stabilizing interaction between two factors bound to mRNA. Elucidation of these interactions requires experimental study.

3’UTRs. The sample of 3’UTRs comprised 11 sequences. The total number of AUG triplets in the sample, 30, appeared approximately equal to the expected value amounting to 31.1, calculated on the assumption on independence of the neighboring positions. However, the six genes whose 5’UTRs were considered above fell into the same distinct groups of three sequences in each (Table): the first group contained three AUG triplets (versus 27 expected randomly), whereas the second only 20 (versus expected 25).

While both the excess or shortage of AUG triplets in 5’UTRs are comparatively well studied (Bernardi, 2000; Kochetov et al., 1998), their increased content in 3’UTRs were detected only in RNAs of certain viruses (Hann et al., 1997). Presumably, they are involved in a short translation re-initiation in 3’UTRs, thereby retaining ribosomes on mRNA and enhancing their efficient transfer to the translation start.

Thus, the mRNAs of corn genes studied allowed us to detect a group of mRNAs distinguishable by the following contextual and structural characteristics. Their 5’UTRs exhibit the secondary structure that brings the cap site with the site of translation initiation and their 3’UTRs contain increased number of AUG triplets. Presumably, simultaneous occurrence of these two specific features represents a signal indicating the presence of the translation determinants facilitating the rotation of ribosomes on the mRNA.

Acknowledgements

The work of IIT and AVK was performed while their visit to RIKEN with a support of STA Foundation.

References