

Software

Open Access

## AUG\_hairpin: prediction of a downstream secondary structure influencing the recognition of a translation start site

Alex V Kochetov\*<sup>1,2</sup>, Andrey Palyanov<sup>1</sup>, Igor I Titov<sup>1</sup>, Dmitry Grigorovich<sup>1</sup>, Akinori Sarai<sup>3</sup> and Nikolay A Kolchanov<sup>1,2</sup>

Address: <sup>1</sup>Institute of Cytology and Genetics, Lavrentieva 10, Novosibirsk 630090, Russia, <sup>2</sup>Novosibirsk State University, Novosibirsk 630090, Russia and <sup>3</sup>Kyushu Institute of Technology, Iizuka, 820-8502, Japan

Email: Alex V Kochetov\* - ak@bionet.nsc.ru; Andrey Palyanov - palyanov\_andrey@ngs.ru; Igor I Titov - titov@bionet.nsc.ru; Dmitry Grigorovich - odip@bionet.nsc.ru; Akinori Sarai - sarai@bse.kyutech.ac.jp; Nikolay A Kolchanov - kol@bionet.nsc.ru

\* Corresponding author

Published: 30 August 2007

Received: 25 January 2007

BMC Bioinformatics 2007, 8:318 doi:10.1186/1471-2105-8-318

Accepted: 30 August 2007

This article is available from: <http://www.biomedcentral.com/1471-2105/8/318>

© 2007 Kochetov et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

### Abstract

**Background:** The translation start site plays an important role in the control of translation efficiency of eukaryotic mRNAs. The recognition of the start AUG codon by eukaryotic ribosomes is considered to depend on its nucleotide context. However, the fraction of eukaryotic mRNAs with the start codon in a suboptimal context is relatively large. It may be expected that mRNA should possess some features providing efficient translation, including the proper recognition of a translation start site. It has been experimentally shown that a downstream hairpin located in certain positions with respect to start codon can compensate in part for the suboptimal AUG context and also increases translation from non-AUG initiation codons. Prediction of such a compensatory hairpin may be useful in the evaluation of eukaryotic mRNA translation properties.

**Results:** We evaluated interdependency between the start codon context and mRNA secondary structure at the CDS beginning: it was found that a suboptimal start codon context significantly correlated with higher base pairing probabilities at positions 13 – 17 of CDS of human and mouse mRNAs. It is likely that the downstream hairpins are used to enhance translation of some mammalian mRNAs *in vivo*. Thus, we have developed a tool, *AUG\_hairpin*, to predict local stem-loop structures located within the defined region at the beginning of mRNA coding part. The implemented algorithm is based on the available published experimental data on the CDS-located stem-loop structures influencing the recognition of upstream start codons.

**Conclusion:** An occurrence of a potential secondary structure downstream of start AUG codon in a suboptimal context (or downstream of a potential non-AUG start codon) may provide researchers with a testable assumption on the presence of additional regulatory signal influencing mRNA translation initiation rate and the start codon choice. *AUG\_hairpin*, which has a convenient Web-interface with adjustable parameters, will make such an evaluation easy and efficient.

### Background

Translation of most eukaryotic mRNAs is likely to be ini-

tiated by a linear scanning mechanism [1], although some alternative mechanisms are also possible [2,3]. According

to the scanning model, 40S ribosomal subunits can either initiate translation at the 5'-proximal AUG codon in a suboptimal context or miss it and initiate translation at downstream AUG(s). For mammalian mRNAs, the most important elements of AUG context are the adenine at position -3 and guanine at position +4 [1,4]. One might expect that mRNA should possess some features providing efficient translation, including the recognition of a genuine translation start site (TSS). However, the fraction of eukaryotic mRNAs with the start AUG codon in a suboptimal context is relatively large [5,6]. It is likely that at least some mRNAs with a suboptimal context of annotated start codon contain other signals providing additional information for efficient TSS recognition (*e.g.*, [7]).

It has been reported earlier that stable hairpins located in certain positions downstream of AUG codon in a suboptimal context can increase translation initiation efficiency [8]. The hairpin was placed at the distances of 5, 11, 17, and 35 nucleotides downstream of the CDS beginning, and start codon recognition efficiencies were evaluated in *in vitro* experiment (Fig. 1). It was reported that the hairpin located at the distance of either 5 or 35 nucleotides did not increase the AUG recognition. However, the hairpin located at the distance of 17 or (to lesser extent) 11 nucleotides enhanced translation initiation considerably in a sequence-independent manner. It was hypothesized [8] that stable hairpins could temporarily stop the movement of 40S ribosomal subunits until the disruption of the secondary structure. The hairpin located at certain critical distance(s) with respect to start codon (at 17<sup>th</sup> nucleotide in the experiment) could delay the 40S ribosomal subunit just in the position providing an efficient interaction between the met-tRNAi-located anticodon and the start AUG codon. It was also shown that such a downstream hairpin could increase translation from non-AUG start codons located in the optimal context [8,9].

This hypothetical mechanism has not been fully verified by *in vivo* experiments yet. However, it was often suggested that such secondary structures can be used to modulate translation efficiency of certain viral and cellular mRNAs with a suboptimal context of start AUG codons [10-14] or with a non-AUG start codons [15-17]. For example, the efficiency of TSS in a suboptimal context in Dengue virus mRNA is increased by a hairpin of moderate stability located 17 nucleotides downstream. The hairpin effect on the TSS recognition depended on its stability and position with respect to AUG [14]. It was also recently found that the stable hairpin located downstream of start AUG codon in Sindbis virus subgenomic 26S RNA provides efficient translation even though eIF2 $\alpha$  is phosphorylated and translation of most cellular mRNAs is blocked [18]. The authors hypothesized that the hairpin can stall the ribosomes on the correct site to initiate translation, thus

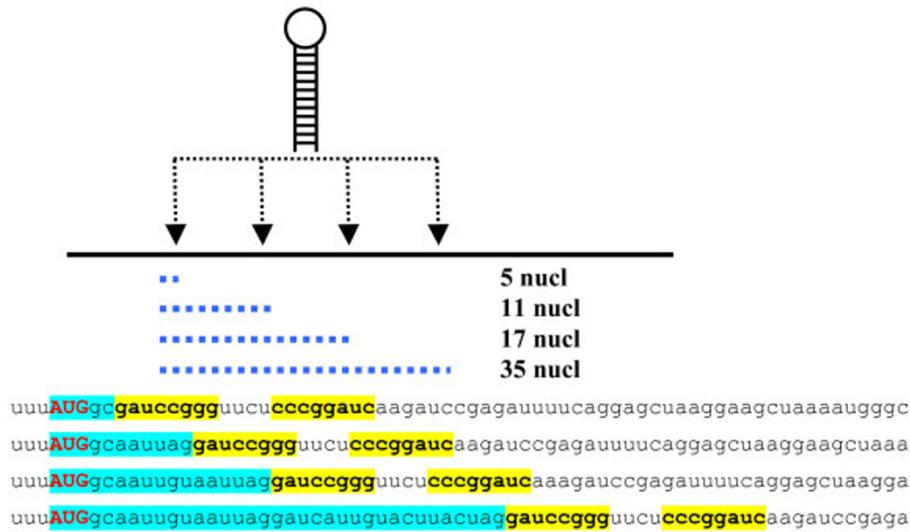
bypassing the requirement for a functional eIF2 and, thereby, specifically supporting translation of certain viral and (probably) cellular mRNAs. Notably, this function of a downstream hairpin is related to a general cellular translation control rather than a compensation of the "weakness" of the upstream start codon context. Thus, the information on the presence of potential compensatory hairpins may be useful for further experimental investigation of both general and specific mRNA translational properties.

Here, we describe the computational tool (*AUG\_hairpin*) targeted at the prediction of secondary structure elements possibly compensating for suboptimal context of translational start codon. We also analyzed the structural features of human and mouse mRNAs and found significant correlation between the base pairing probabilities in positions 13–17 of CDS and the TSS context. This relationship supports the hypothesis on the functional significance of precisely located downstream hairpins for the TSS recognition in some cellular mRNAs.

### Implementation

According to the experimental data, the hairpins started either upstream or downstream of certain "critical" region of CDS did not compensate for the "weak" AUG context. In particular, continuous secondary structure started at 5<sup>th</sup> nucleotide of coding sequence did not increase translation initiation efficiency despite it included the critical 12<sup>th</sup> and 18<sup>th</sup> positions (Fig. 1; [8]). Based on this observation, *AUG\_hairpin* predicts the stem-loop structures, in which 5'-borders are located within the critical region (from 12<sup>th</sup> to 18<sup>th</sup> nucleotides by default). An appropriate stem-loop structure can also be a part of a more complex secondary structure started upstream of the critical region. We hypothesized that the 40S ribosomal subunit moving from the 5'-end of mRNA can pause consequently on each stable stem of a complex stem-loop structure waiting for its melting. In this case we assumed that an eligible hairpin has to be separated from upstream secondary structure elements by some impaired segment (*e.g.*, loop) (for detailed description, see tutorial at the program [www-ites.com](http://www-ites.com)).

5'-UTR and CDS nucleotide sequences have to be entered separately to the program through [www-page](http://www-ites.com) (this provides the program with information on the start codon position). *AUG\_hairpin* analyzes the mRNA segment compiled from 10 nucleotides long 5'-UTR portion located immediately upstream of the start AUG codon and 100 5'-proximal nucleotides of CDS. Algorithm consists of the following main steps: (1) Prediction of RNA optimal secondary structure for 5'UTR-CDS fragment. For this purpose the program *foldRNA* from Vienna RNA package v.1.4 [19] was implemented as subroutine. (2) Checking

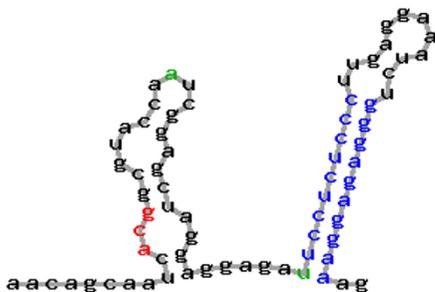


**Figure 1**

Description of experiment reported by Kozak ([8]): The hairpin (inverted repeats are marked with bold type and yellow boxes) was placed at the distances of 5, 11, 17, and 35 nucleotides from the CDS beginning (marked with blue boxes).

the occurrence of a perfect (or imperfect) stem located a certain distance downstream of start AUG codon (from 12<sup>th</sup> to 18<sup>th</sup> nucleotides by default; user can change this range from the CDS beginning till 30<sup>th</sup> nucleotide). Conventionally, a stem is perfect when it does not contain any

Energy of secondary structure = -24.6 kcal/mol  
 Position of Hairpin start: 30  
 Energy of double strands in Hairpin: -19.9 kcal/mol



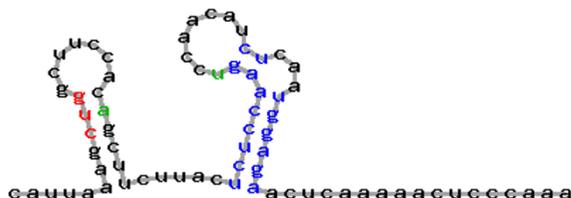
**Figure 2**

Predicted local secondary structure at the CDS beginning of *Agamous* mRNA (printscreens of *AUG\_hairpin* output [www-wwwnet.com](http://www.wwwnet.com); nucleotide sequence of 5'-UTR-CDS fragment was taken from [15]). Borders of the critical region (12<sup>th</sup> and 30<sup>th</sup> pos.) are marked with green colour; stem region of the eligible hairpin is marked with blue colour; start codon is marked with red colour.

interrupting loops; an imperfect stem includes short mismatches (one-nucleotide bulges or 1+1 inner loops) which presumably do not interrupt stacking interactions. Program's output presents visualization of the optimal secondary structure and provides calculation of the predicted thermodynamic stability of the secondary structure and the helices (if any) started within the defined critical region. The program was written on C++ and runs in a Unix environment.

It is likely that a downstream hairpin can increase translation even located outside the critical region defined by Kozak ([8]; from 12<sup>th</sup> to 18<sup>th</sup> nucleotides of CDS) although at a lower efficiency. It was experimentally found that the hairpin could influence translation even located at pos. 30 [10] or at pos. 27 of CDS [18]. Taking into account that the hairpin located at CDS position 35 did not increase an upstream AUG recognition in *in vitro* experiment [8], we limited the *AUG\_hairpin* prediction interval by 30<sup>th</sup> nucleotide of CDS. We analyzed two mRNA nucleotide sequences with annotated non-AUG translation start site taking them as examples. CDS of *Agamous* mRNA (*Arabidopsis thaliana*) starts from ACG codon [15] and CDS of *NsRpoT-C* mRNA (*Nicotiana sylvestris*) starts from CUG codon [20]. According to the *AUG\_hairpin* prediction, there are stable hairpins located downstream of start codons in these mRNAs (*Agamous*: at pos. 30, E = -19.9 kcal/mol, Fig. 2; *NsRpoT-C*: at pos. 22, E = -12.7 kcal/mol, Fig. 3).

Energy of secondary structure = -13.5 kcal/mol  
 Position of Hairpin start: 22  
 Energy of double strands in Hairpin: -12.7 kcal/mol



**Figure 3**  
 Predicted local secondary structure at the CDS beginning of *NsRpoT-C* mRNA (printscreen of *AUG\_hairpin* output [www-](http://www-) page; nucleotide sequence of 5'-UTR-CDS fragment was taken from [AB084950](http://www.ncbi.nlm.nih.gov/GenBank/AB084950) GenBank entry and [20]). Borders of the critical region (12<sup>th</sup> and 30<sup>th</sup> pos.) are marked with green colour; stem region of the eligible hairpin is marked with blue colour; start codon is marked with red colour.

### Results and discussion

It was assumed that stable hairpins precisely located downstream of translation start sites (AUG or non-AUG) enhance translation initiation rate of some viral and cellular mRNAs [10-17]. Computational analysis of yeast mRNAs demonstrated the interrelationship between the base pairing probabilities at pos. 14–17 of CDS and the start codon context: mRNAs with a suboptimal TSS context were characterized by a more frequent presence of secondary structure at these positions [21]. In this work we tested the interdependency between the base pairing probabilities in the 5'-terminal CDS segment and start codon context in mammalian mRNAs. For this purpose human and mouse cDNA samples characterized by complete CDS and 5'-UTR size larger than 29 nucleotides were extracted from GenBank. The subsamples of cDNAs characterized by either optimal (purine at pos. -3 upstream the annotated TSS; 18750 human and 7700 mouse cDNAs) or suboptimal (pyrimidine at pos. -3 upstream the annotated TSS; 3469 human and 1233 mouse cDNAs) start codon context were used for further comparative analysis. The mRNA base pairing probabilities (BPP; [22]) were calculated using the Vienna RNA secondary structure package [19]. mRNA segments containing 5'-UTR (limited by 100 nucleotides in size) and the beginning of CDS (100 nucleotide in size; thus, total size of segments varied from 130 to 200 nucleotides depending on the 5'-UTR length) were used. Positional BPP values were calculated and compared between the samples of mRNAs with optimal and suboptimal start codon contexts. The difference in average BPP values at the CDS positions between these samples is shown in Table 1. One may see that mRNAs

**Table 1: The difference in average BPP values and G+C content in 6<sup>th</sup> – 30<sup>th</sup> CDS positions of human and mouse cDNAs characterized by optimal and suboptimal start codon contexts (purine (Pu) and pyrimidine (Py) in pos. -3 upstream AUG, respectively)†**

CDS pos.	human		mouse	
	BPP <sub>Py</sub> -BPP <sub>Pu</sub>	GC <sub>Py</sub> -GC <sub>Pu</sub> ,%	BPP <sub>Py</sub> -BPP <sub>Pu</sub>	GC <sub>Py</sub> -GC <sub>Pu</sub> ,%
6	-0.014	-3.93	0.007	-1.16
7	0.005	-2.23	-0.016	-3.14
8	0.011	1.34	-0.002	0
9	0.006	-2.36	-0.006	-1.07
10	0.006	1.15	-0.014	-1.21
11	-0.008	2.14	-0.004	-1.47
12	-0.007	-3.79	0.005	2.46
13	<b>0.015*</b>	-0.41	-0.004	-0.38
14	<b>0.012**</b>	1.65	<b>0.020*</b>	4.47
15	<b>0.015***</b>	-1.99	<b>0.018**</b>	-1.08
16	<b>0.014*</b>	-0.93	<b>0.036***</b>	-1.48
17	<b>0.012*</b>	4.21	<b>0.019*</b>	7.53
18	-0.003	-3.23	-0.005	-5.46
19	0.004	0.59	0.010	1.31
20	0.005	0.68	-0.008	2.8
21	-0.025	-4.94	-0.012	-1.29
22	-0.005	-0.50	-0.013	-1.6
23	-0.018	3.79	-0.020	5.15
24	-0.020	-3.51	0.001	-3.21
25	0.007	-0.70	-0.013	-0.32
26	<b>0.017*</b>	2.14	-0.003	4.01
27	0.006	-3.66	-0.022	-4.23
28	-0.001	-0.03	-0.022	1.02
29	0.007	5.30	-0.022	5.21
30	-0.016	-3.68	-0.035	-5.06

† CDS positions where mean BPP values were significantly higher in mRNA samples with a suboptimal start codon context are marked (BPP<sub>Py</sub>-BPP<sub>Pu</sub> values are bolded; the significance level (according to Mann-Whitney *U*-test) is shown by numbers: \* *p* < 0.05; \*\* *p* < 0.01; \*\*\* *p* < 0.001)

with a suboptimal TSS context are characterized by significantly higher average BPP values in pos. 13–17. We also calculated average BPP values at 5'-end CDS positions (from 6<sup>th</sup> to 56<sup>th</sup> nucleotides; average BPP = 0.615112 ± 0.0006 (human mRNAs with a suboptimal TSS context); 0.615775 ± 0.0006 (human mRNAs with optimal TSS context); 0.613197 ± 0.002 (mouse mRNAs with a suboptimal TSS context); 0.614936 ± 0.0009 (mouse mRNAs with optimal TSS context)). Thus, mRNAs with a suboptimal start codon context were not characterized by a higher average BPP value at the CDS beginning.

It may be speculated that this difference results from a local deviation in G+C content (GC pairs make the major impact on the secondary structure stability) or from some codon-dependent periodic pattern [23,24]. However, average positional differences in G+C frequencies do not correlate with the difference in BPP values: it is unlikely

**Table 2: The difference in characteristics of 5'-end proximal hairpins started in 6<sup>th</sup> – 20<sup>th</sup> CDS positions of human and mouse cDNAs characterized by optimal and suboptimal start codon contexts (purine (Pu) and pyrimidine (Py) in pos. -3 upstream AUG, respectively; E<sub>tot</sub>, energy of an eligible hairpin, kcal/mol; E<sub>st</sub>, energy of the hairpin stem region, kcal/mol; L<sub>st</sub>, size of the hairpin stem region, base pairs)†**

CDS pos.	human			mouse		
	E <sub>tot</sub> <sup>Py</sup> -E <sub>tot</sub> <sup>Pu</sup> kcal/mol	E <sub>st</sub> <sup>Py</sup> -E <sub>st</sub> <sup>Pu</sup> kcal/mol	L <sub>st</sub> <sup>Py</sup> -L <sub>st</sub> <sup>Pu</sup> b.p.	E <sub>tot</sub> <sup>Py</sup> -E <sub>tot</sub> <sup>Pu</sup> kcal/mol	E <sub>st</sub> <sup>Py</sup> -E <sub>st</sub> <sup>Pu</sup> kcal/mol	L <sub>st</sub> <sup>Py</sup> -L <sub>st</sub> <sup>Pu</sup> b.p.
6	0.45	0.21	-0.14	0.09	-0.16	0.08
7	0.40	0.38	-0.19	0.02	0.35	-0.26
8	-0.25	0.15	-0.22	1.04	-0.027	-0.16
9	1.79	0.11	-0.203	0.24	-0.53	0.02
10	1.89	0.01	0.13	-0.47	-0.29	0.11
11	0.79	0.21	-0.09	-0.87	-0.80	0.16
12	1.12	0.11	-0.02	0.86	1.29	-0.64
13	<b>-1.28*</b>	0.39	-0.13	<b>-3.65*</b>	<b>-1.02*</b>	0.19
14	<b>-2.41*</b>	<b>-1.09**</b>	0.10	1.12	1.65	-0.74
15	-0.87	-0.29	0.09	-1.84	0.54	-0.47
16	<b>-1.03**</b>	<b>-0.53*</b>	<b>0.26*</b>	3.42	-0.10	0.11
17	2.87	<b>-1.15**</b>	<b>0.71*</b>	<b>-3.73*</b>	<b>-0.60*</b>	0.46
18	-0.94	0.13	-0.24	3.08	-0.43	0.04
19	<b>-2.99*</b>	0.82	-0.48	1.88	0.75	0.10
20	-1.82	-0.58	0.06	-2.74	0.47	-0.32

† CDS positions where mean E<sub>tot</sub> and E<sub>st</sub> values were significantly lower (and L<sub>st</sub> values were significantly higher) in mRNA samples with a suboptimal start codon context are marked (E<sub>tot</sub><sup>Py</sup> - E<sub>tot</sub><sup>Pu</sup>, E<sub>st</sub><sup>Py</sup> - E<sub>st</sub><sup>Pu</sup>, and L<sub>st</sub><sup>Py</sup> - L<sub>st</sub><sup>Pu</sup> values are bolded; the significance level (according to Mann-Whitney U-test or Kolmogorov-Smirnov two-sample test) is shown by numbers: \* p < 0.05; \*\* p < 0.01; \*\*\* p < 0.001)

that the observed dependency between the TSS context and base pairing probabilities reflects an unusual G+C distribution at the CDS beginning rather than the more frequent representation of the hairpin-containing mRNAs in a sample with a suboptimal TSS context (Table 1). These data demonstrate that precisely positioned hairpins may increase translation efficiency of some mammalian mRNAs *in vivo* and that the positions defined by Kozak ([8]; from 13<sup>th</sup> to 17<sup>th</sup> nucleotide of CDS) are most frequently (or efficiently) used for this purpose.

AUG\_hairpin was used to analyze the stem-loop structures whose 5'-borders were located between the 6<sup>th</sup> and 30<sup>th</sup> nucleotides of CDS in human and mouse mRNA samples. Three parameters were taken into account: total energy of the stem-loop structure (E<sub>tot</sub>), energy of the stem (E<sub>st</sub>), and the stem size (L<sub>st</sub>, 5'-proximal perfect stems started within defined region were taken). It was found that mRNAs with a suboptimal start codon context were characterized by significantly more stable secondary structure elements started in positions 13, 14, 16, 17, 19 (Table 2). According to these results, the 5'-border of the critical region is also located at position 13 of CDS. We further selected sub-samples of human and mouse mRNAs with stable stem-loop structures whose 5'-borders were located between the 13<sup>th</sup> and 19<sup>th</sup> nucleotides of CDS. It was found that 546 human mRNAs (16% of the sample with a suboptimal start codon context) were characterized by the pres-

ence of a stable stem-loop structure (E<sub>tot</sub> < -20 kcal/mol) in the defined region. Average energies of eligible stem-loop structures (E<sub>tot</sub>) were -34.9 kcal/mol and -33.2 kcal/mol for human mRNAs with a suboptimal and the optimal start codon contexts, respectively (the distributions of E<sub>tot</sub> values differ significantly according to Kolmogorov-Smirnov two-sample test, p < 0.01; the difference between mean E<sub>tot</sub> values was also significant according to the Mann-Whitney U-test, p < 0.00005). Similarly, 187 mouse mRNAs with a suboptimal start codon context were characterized with a highly stable hairpin (E<sub>st</sub> < -20 kcal/mol) in this region (E<sub>tot</sub> = -32.9 kcal/mol).

It was reported that a correctly positioned hairpin even of a moderate stability (-8.2 kcal/mol) enhanced the recognition of upstream AUG codon in a suboptimal context [14]: thus, it may be assumed that at least some of these mammalian mRNAs possess higher translation initiation efficiency due to the presence of "compensatory" hairpins than it may be expected from the context of annotated start codon (lists of human and mouse mRNAs with some additional information on secondary structure characteristics are available as Additional file 1 and Additional file 2, respectively). It should be, however, noted that suboptimal context of start codon is not necessarily compensated by "compensatory hairpin": in many cases, contexts of translational start sites are likely to be evolutionary

attenuated to decrease translation level of mRNAs encoding regulatory proteins (e.g., [5]).

### Conclusion

The presence of a potential secondary structure downstream of start AUG codon in a suboptimal context (or downstream of a potential non-AUG start codon) can provide researchers with a testable assumption on the additional regulatory element influencing translation initiation level. *AUG\_hairpin* is based on an elegant hypothesis supported by the *in vitro* [8] and *in vivo* experimental data [10,12-14] as well as the results of computational analysis ([21]; Tables 1 and 2 in this manuscript).

It should be noted that the applied algorithm depends on the interpretation of available (rather limited) experimental data and the prediction accuracy may also be limited. Only few hairpin positions were tested in experiments. Secondary structure elements influencing translation start site recognition *in vivo* may have distinct characteristics (e.g., species-specific or tissue-specific). Currently it is also not possible to predict the interdependence of hairpin stability and its influence on start codon recognition as well as the influence of mRNA-protein and mRNA-ribosome interactions during translation initiation process on the mRNA secondary structure. Finally, the recognition of start codons in a suboptimal context can be modulated through other (currently poorly known) signals [7,25-28], and the absence of a "compensatory" hairpin does not necessarily mean that the TSS recognition is inefficient. However, despite these limitations, *AUG\_hairpin* may be used to reveal potential "compensatory" hairpins in the case of discrepancy between the gene expression pattern and mRNA features (e.g., highly expressed gene is characterized by a suboptimal context of annotated translation start site, proteomic or phylogenetic data suggest the usage of non-AUG potential start codons [29], etc.).

### Availability and requirements

Project name: *AUG\_hairpin*

Project home pages:

[http://www.mgs.bionet.nsc.ru/mgs/programs/aug\\_hairpin/](http://www.mgs.bionet.nsc.ru/mgs/programs/aug_hairpin/)

[http://gibk26.bse.kyutech.ac.jp/aug\\_hairpin/](http://gibk26.bse.kyutech.ac.jp/aug_hairpin/)

Any restrictions to use by non-academics: licence needed

### Abbreviations

UTR, untranslated region of mRNA; CDS, protein coding sequence; BPP, base pairing probability.

### Authors' contributions

AK and IT conceived of the study, carried out the computational analysis and drafted the manuscript. AP participated in the implementation of *AUG\_hairpin* algorithm. NK and AS participated in the design and coordination of the study. DG prepared the www-site. All authors read and approved the final manuscript version.

### Additional material

#### Additional file 1

List of human mRNAs with a suboptimal context of annotated start codon potentially containing "compensatory hairpins". This table provides a list of human mRNAs with a suboptimal context of annotated start codon containing stem-loop structures whose 5'-borders are located between the 13<sup>th</sup> and 19<sup>th</sup> nucleotides of CDS. The Table contains information on secondary structure (position of a hairpin's 5'-border, energy of the hairpin and stem(s) given separately) as well as an information from DE (DESCRIPTION) field of GenBank entries (which provides the names of corresponding genes).

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2105-8-318-S1.xls>]

#### Additional file 2

List of mouse mRNAs with a suboptimal context of annotated start codon potentially containing "compensatory hairpins". This table provides a list of mouse mRNAs with a suboptimal context of annotated start codon containing stem-loop structures whose 5'-borders are located between the 13<sup>th</sup> and 19<sup>th</sup> nucleotides of CDS. The Table contains information on secondary structure (position of a hairpin's 5'-border, energy of the hairpin and stem(s) given separately) as well as an information from DE (DESCRIPTION) field of GenBank entries (which provides the names of corresponding genes).

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2105-8-318-S2.xls>]

### Acknowledgements

This work was supported by the Russian Foundation for Basic Research (grant No. 05-04-48207) and RAS programs (Dynamics of Gene Pools). A.V.K., N.A.K. and D.G. thank Siberian Division of Russian Academy of Sciences (Complex Integration Program 5.3) and RAS Program "Molecular and Cellular Biology" for partial support. A.V.K. is also grateful to JSPS short-term fellowship program.

### References

1. Kozak M: **Regulation of translation via mRNA structure in prokaryotes and eukaryotes.** *Gene* 2005, **361**:13-37.
2. Jackson RJ: **Alternative mechanisms of initiating translation of mammalian mRNAs.** *Biochem Soc Trans* 2005, **33**:1231-1241.
3. Baird SD, Turcotte M, Korneluk RG, Holcik M: **Searching for IRES.** *RNA* 2006, **12**:1755-1785.
4. Pisarev AV, Kolupaeva VG, Pisareva VP, Merrick WC, Hellen CU, Pestova TV: **Specific functional interactions of nucleotides at key -3 and +4 positions flanking the initiation codon with components of the mammalian 48S translation initiation complex.** *Genes Dev* 2006, **20**:624-636.
5. Rogozin IB, Kochetov AV, Kondrashov FA, Koonin EV, Milanezi L: **Presence of ATG triplets in 5' untranslated regions of**

- eukaryotic cDNAs correlates with a "weak" context of the start codon. *Bioinformatics* 2001, **17**:890-900.
6. Kochetov AV, Sarai A, Rogozin IB, Shumny VK, Kolchanov NA: **The role of alternative translation start sites in generation of human protein diversity.** *Mol Genet Genomics* 2005, **273**:491-496.
  7. Kochetov AV: **AUG codons at the beginning of protein coding sequences are frequent in eukaryotic mRNAs with a suboptimal start codon context.** *Bioinformatics* 2005, **21**:837-840.
  8. Kozak M: **Downstream secondary structure facilitates recognition of initiator codons by eukaryotic ribosomes.** *Proc Natl Acad Sci USA* 1990, **87**:8301-8305.
  9. Kozak M: **Context effects and inefficient initiation at non-AUG codons in eucaryotic cell-free translation systems.** *Mol Cell Biol* 1989, **9**:5073-5080.
  10. Hwang W-L, Su T-S: **The encapsidation signal of hepatitis B virus facilitates preC AUG recognition resulting in inefficient translation of the downstream genes.** *J Gen Virol* 1999, **80**:1769-1776.
  11. Ciullo M, Del Pozzo G, Autiero M, Guardiola J: **Downstream sequence adjacent to AUG affects translation of chloramphenicol acetyl transferase in eukaryotic cells.** *DNA Cell Biol* 2000, **19**:39-46.
  12. Kwon HS, Lee DK, Lee JJ, Edenberg HJ, Ahn YH, Hur MW: **Posttranscriptional regulation of human ADH5/FDH and Myf6 gene expression by upstream AUG codons.** *Arch Biochem Biophys* 2001, **386**:163-171.
  13. Yang L, Chen J, Chang CC, Yang XY, Wang ZZ, Chang TY, Li BL: **A stable upstream stem-loop structure enhances selection of the first 5'-ORF-AUG as a main start codon for translation initiation of human ACAT1 mRNA.** *Acta Biochim Biophys Sin* 2004, **36**:259-268.
  14. Clyde K, Harris E: **RNA secondary structure in the coding region of dengue virus type 2 directs translation start codon selection and is required for viral replication.** *J Virol* 2006, **80**:2170-2182.
  15. Riechmann JL, Ito T, Meyerowitz EM: **Non-AUG initiation of AGAMOUS mRNA translation in Arabidopsis thaliana.** *Mol Cell Biol* 1999, **19**:8505-8512.
  16. Nguyen M, He B, Karaplis A: **Nuclear forms of parathyroid hormone-related peptide are translated from non-AUG start sites downstream from the initiator methionine.** *Endocrinology* 2001, **142**:694-703.
  17. Takahashi K, Maruyama M, Tokuzawa Y, Murakami M, Oda Y, Yoshikane N, Makabe KW, Ichisaka T, Yamanaka S: **Evolutionarily conserved non-AUG translation initiation in NAT1/p97/DAP5 (EIF4G2).** *Genomics* 2005, **85**:360-371.
  18. Ventoso I, Sanz MA, Molina S, Berlanga JJ, Carrasco L, Esteban M: **Translational resistance of late alphavirus mRNA to eIF2alpha phosphorylation: a strategy to overcome the antiviral effect of protein kinase PKR.** *Genes Dev* 2006, **20**:87-100.
  19. Hofacker IL: **Vienna RNA secondary structure server.** *Nucleic Acids Res* 2003, **31**:3429-3431.
  20. Kobayashi Y, Dokiya Y, Kumazawa Y, Sugita M: **Non-AUG translation initiation of mRNA encoding plastid-targeted phage-type RNA polymerase in Nicotiana sylvestris.** *Biochem Biophys Res Comm* 2002, **299**:57-61.
  21. Kochetov AV, Kolchanov NA, Sarai A: **Interrelations between the efficiency of translation start sites and other sequence features of yeast mRNAs.** *Mol Genet Genomics* 2003, **270**:442-447.
  22. McCaskill JS: **The equilibrium partition function and base pair binding probabilities for RNA secondary structure.** *Biopolymers* 1990, **29**:1105-1119.
  23. Meyer IM, Miklos I: **Statistical evidence for conserved, local secondary structure in the coding regions of eukaryotic mRNAs and pre-mRNAs.** *Nucleic Acids Res* 2005, **33**:6338-6348.
  24. Shabalina SA, Ogurtsov AY, Spiridonov NA: **A periodic pattern of mRNA secondary structure created by the genetic code.** *Nucleic Acids Res* 2006, **34**:2428-2437.
  25. Lukaszewicz M, Feuermann M, Jerouville B, Stas A, Boutry M: **In vivo evaluation of the context sequence of the translation initiation codon in plants.** *Plant Sci* 2000, **154**:89-98.
  26. Sawant SV, Kiran K, Singh PK, Tuli R: **Sequence architecture downstream of the initiator codon enhances gene expression and protein stability in plants.** *Plant Physiol* 2001, **126**:1630-1636.
  27. Zhao K-N, Tomlison L, Liu WJ, Gu W, Frazer IH: **Effects of additional sequences directly downstream from the AUG on the expression of GFP gene.** *Biochim Biophys Acta* 2003, **1630**:84-95.
  28. Shabalina SA, Ogurtsov AY, Rogozin IB, Koonin EV, Lipman DJ: **Comparative analysis of orthologous eukaryotic mRNAs: potential hidden functional signals.** *Nucleic Acids Res* 2004, **32**:1774-1782.
  29. Touriol C, Bornes S, Bonnal S, Audigier S, Prats H, Prats AC, Vagner S: **Generation of protein isoform diversity by alternative initiation of translation at non-AUG codons.** *Biol Cell* 2003, **95**:169-178.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:  
[http://www.biomedcentral.com/info/publishing\\_adv.asp](http://www.biomedcentral.com/info/publishing_adv.asp)

