

Possible utilization of -1 Ribosomal frame shifting in the expression of a human SEMA6C isoform

Xiaolan Huang¹, Qiang Cheng¹ & Zhihua Du^{2*}

¹Department of Computer Science, Southern Illinois University at Carbondale, IL 62901, USA; ²Department of Chemistry and Biochemistry, Southern Illinois University at Carbondale, IL 62901, USA; Zhihua Du – Email: zdu@chem.siu.edu; *Corresponding author

Received March 14, 2013; Accepted March 24, 2013; Published August 07, 2013

Abstract:

We have used bioinformatics approaches to identify a potential case of -1 ribosomal frame shifting in the mRNAs of the three variants of human SEMA6C protein. The mRNAs contain a heptanucleotide slippery sequence followed by a compact H-type pseudoknot. Unlike -1 frameshifting signals in viral or viral-like mRNAs, the slippery sequence and downstream pseudoknot in SEMA6C mRNAs locate 423 nucleotides (encoding 141 amino acids) upstream of the stop codon. The potential -1 frameshifting event would produce a polypeptide of 238 residues encoded by the -1 reading frames. Sequence similarity searches using BLAST indicate that ~90% of the 238 residues match actual protein sequences annotated as SEMA6C proteins in the database. We propose that the mRNAs of human SEMA6C utilize a pseudoknot dependent -1 ribosomal frameshifting mechanism to express novel SEMA6C isoforms.

Key words: RNA, pseudoknot, -1 ribosomal frame shifting, SEMA6C.

Background:

-1 ribosomal frameshifting is a programmed recoding event during which a certain percentage of the translating ribosomes shift back by one nucleotide along the mRNA to the -1 reading frame [1, 2]. Two cis-acting elements in the mRNA are typically required for efficient -1 ribosomal frameshifting: a heptanucleotide "slippery sequence" with a consensus of X XXY YYZ (the zero reading frame is indicated by the triplets) and a downstream RNA structure.

When frameshifting happens, the two tRNAs recognizing XXY and YYZ condons in the zero reading frame shift back one nucleotide to pair with the XXX and YYY condons in the -1 reading frame. While the actual frameshifting occurs at the slippery sequence, an RNA structure that is located several nucleotides downstream of the slippery sequence plays a stimulatory role for efficient frameshifting. Most often, the frameshift stimulating RNA structure is a compact H-type pseudoknot. An H-type pseudoknot is a structural motif of RNA formed when a stretch of nucleotides within a stem-loop

(hairpin) basepairs with a complementary sequence outside that loop [3-5]. -1 ribosomal frameshifting is a well-established mechanism that is utilized by many RNA viruses to express their structural and enzymatic proteins at a defined ratio [6-9], which is important for the viral life cycle. For the expression of cellular genes, only a few cases have been reported for the involvement of -1 ribosomal frameshifting. So far, only three mammalian genes are known to use the -1 frameshifting mechanism. These are the human paternally expressed gene 10 (PEG10) [10] and the paraneoplastic antigen Ma3 & Ma5 genes [11]. All of these genes are derived from retroelements and they encode viral-like proteins. They all use a retrovirus-like -1 frameshifting mechanism to express the overlapping -1 reading frame sequences with 15-30% efficiencies.

Here we report a possible case of -1 ribosomal frameshifting in the translation of the human SEMA6C (semaphorin) mRNAs. Sequence analysis indicates that the frameshit product most likely represents a novel isoform of the protein.

Methodology: mRNA sequences

Three mRNA (cDNA) sequences for the three variants of *Homo sapiens* sema domain, transmembrane domain (TM), and cytoplasmic domain, (semaphorin) 6C (SEMA6C) protein were obtained from the NCBI Reference Sequence database, with accession IDs NM_001178061 (4024 bp), NM_030913 (3928 bp) & NM_001178062 (3904 bp) for variants 1, 2 & 3 respectively.

Identification of potential frameshifting events

The mRNA sequences were subjected to analysis by a program written in C++. The search for potential -1 frameshifting events was carried out in three steps. First, the program searches for X XXY YYZ heptanucleotide slippery sequences that register in the correct reading frame (XXY & YYZ denote two codons in the zero reading frame). Second, the program tests whether an H-type pseudoknot can form within 10 nucleotides downstream of the identified slippery sequences. Details for identifying pseudoknots have been described elsewhere. Briefly, an H-type pseudoknot contains four essential elements: two helical stems (S1 & S2) and two connecting loops (L1 & L2), with another loop (L3) as being optional. The program scans through the mRNA sequence and tests whether the sequence elements for pseudoknot formation are present within a given sequence window (defined by the default ranges of stem and loop lengths). If a sequence contains two pairs of complementary stretches (forming S1 & S2) that are separated by two or three connecting loops (L1, L2, and L3), then the sequence has the potential to fold into a pseudoknot. In the final step, providing that a slippery sequence and a downstream pseudoknot are identified, the program tests whether the -1 reading frame encodes a polypeptide with 100 or more amino acids residues after the slippery sequence, i.g. 300 or more nucleotides in the -1 reading frame without a stop codon. Of course, -1 ribosomal frameshifting may also lead to the synthesis of less than 100 amino acids residues in the -1 reading frames. However, this scenario is not considered in the current study.

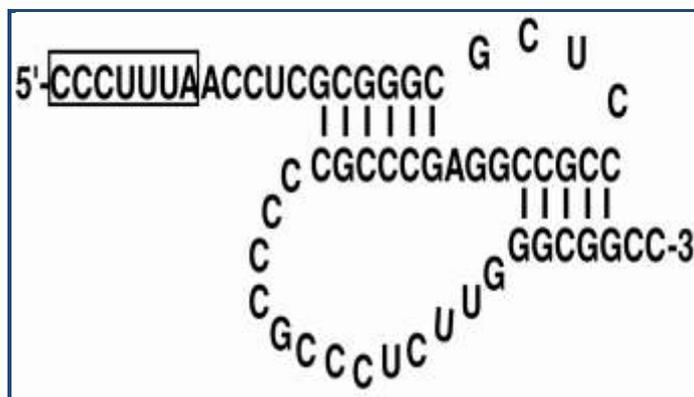


Figure 1: Potential signals for -1 ribosomal frameshifting in the human SEMA6C mRNAs. The heptanucleotide slippery sequence is boxed. A potential H-type pseudoknot is located five nucleotides downstream of the slippery sequence.

Sequence analysis of the frameshift product

Similarity searches for the amino acids sequence of the polypeptide generated by -1 ribosomal frameshifting were performed by using the BLAST program available at NCBI [http://blast.ncbi.nlm.nih.gov/], against the non-redundant protein sequences database. Multiple sequences alignment of

amino acids sequences was carried out by the ClustalW program [12].

Results & Discussion:

Human SEMA6C has three variants in the NCBI Reference Sequence database (962, 930, and 922 aa for variant 1, 2 and 3 respectively). Compared to variant 1, the mRNAs of variants 2 and 3 apparently lack an exon encoding 32 aa and 40 aa respectively. The three variants otherwise have identical protein and DNA sequences.

As shown in (Figure 1), a potential slippery sequence of C CCU UUA was identified, with the CCU codon coding for residue Pro-821 (residue numbering refers to the variant 1 sequence). C CCU UUA is an established slippery sequence utilized by giardavirus (GLV) at the viral *gag-pol* junction for -1 frameshifting [13]. Five nucleotides downstream from the slippery sequence, a potential H-type pseudoknot were identified. A -1 ribosomal frameshifting event induced by these signals would produce a polypeptide of 238 amino acids residues encoding by the -1 reading frame. In comparison, the normal protein sequence contains 142 amino acids residues after the frameshifting site (Figure 2). Sequence similarity between the frameshift product and the normal sequence is low, as can be seen from the pairwise alignment.

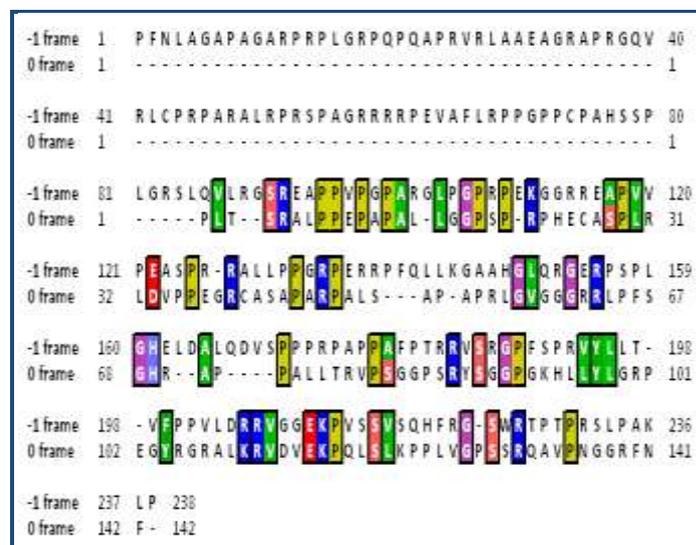


Figure 2: Comparison of the polypeptide sequences encoded by the 0 and -1 reading frames after the slippery sequence. The alignment is carried out by ClustalW. Identical or highly similar residues are boxed and highlighted with background colors.

We performed a similarity search for the frameshift product using BLAST against the non-redundant protein sequences database. When using the 238 amino acids sequence as the query, only one hit was found: AAI14624, which is annotated as a human SEMA6C protein with 537 residues. Residues 74-238 of the query are identical to the last 165 residues (aa 373-537) of AAI14624 except at one residue (Ala-137 vs Val-446). More interestingly, the protein AAI14624 is encoded by a complete cDNA clone (IMAGE: 40036285), indicating that the protein is actually expressed. AAI14624 contains 537 residues; residues 1-371 are identical to corresponding sequence of SEMA6C variant 1 and residues 373-537 match the C-terminal 165 residues of the frameshift product (Figure 3). Therefore, AAI14624 likely

represents a novel alternatively spliced isoform of human SEMA6C with the N-terminal ~70% of the sequence being encoded by the zero reading frame and the C-terminal ~30% of the sequence being encoded by the -1 reading frame (using the mRNA of variant 1 as a reference). We also used the residues 1-73 of the frameshift product as the query in a BLAST search. Only one hit was found: AES06293, which is annotated as a partial sequence of a SEMA6C protein from *Mustela putorius furo*. Residues 15-60 of the query show a high degree of similarity to the last 46 residues of AES06293 (identities 54%, positives 63%, no gap).

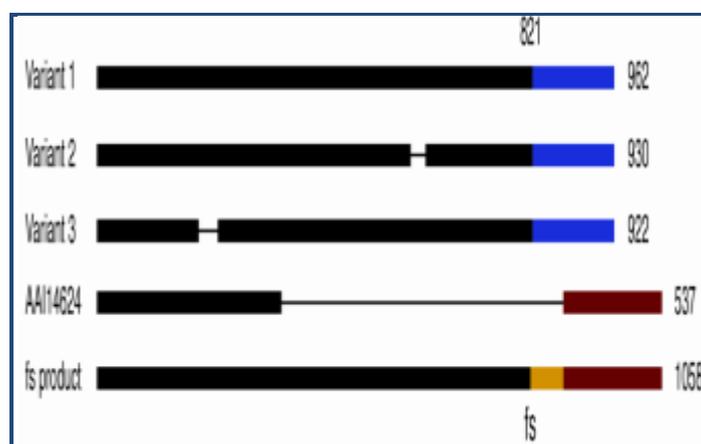


Figure 3: A schematic diagram of the human SEMA6C variants and isoforms, including the protein product generated by the potential -1 ribosomal frameshifting. Thick solid bars represent amino acids sequences. Thin lines indicate gaps in the sequences. Numbers indicate the number of residues in each protein and the position of the frameshift site in variant 1. All three variants have the same potential frameshifting signals. Only the frameshift (fs) product generated by the variant 1 mRNA is shown. Colors of the sequences are used to indicate the identity of the sequences. Sequences in black indicate they are identical in all proteins (except gaps). Sequences of the same color are identical (or almost identical), while a difference in color indicates sequence dissimilarity.

From the above sequence analysis, it is clear that ~90% of the 238 amino acids residues of the frameshift product match to SEMA6C protein sequences in the database. These results provide strong support for the utilization of -1 ribosomal frameshifting mechanism in the translation of human SEMA6C

mRNAs. A SEMA6C protein isoform produced by -1 ribosomal frameshifting would contain 1058 amino acids residues, which is 96 residues longer than the longest known SEMA6C variant (**Figure 3**).

Conclusion:

A probable case of pseudoknot-stimulating -1 ribosomal frameshifting is reported here for the mRNAs of three variants of human SEMA6C proteins. This case differs from the previously established cases of human PEG10 [10] and Ma3 & Ma5 gene expressions [11] in two important aspects. First, SEMA6C is a true cellular protein without a viral origin, while PEG10, Ma3 and Ma5 are all retrovirus-like proteins. The frameshift site in SEMA6C mRNA locates more than 400 nucleotides upstream of the 0 reading frame stop codon, while the frameshift sites in the PEG10, Ma3 and Ma5 mRNAs all locate at the end of the 0 reading frame (also retrovirus-like).

Results from sequence analysis of the frameshift product shows, beyond any reasonable doubt, that the polypeptide sequence generated by -1 frameshifting in SEMA6C mRNA is genuine SEMA6C protein sequence. Very likely, the proposed -1 ribosomal frameshifting event represents a real case of recoding in human gene expression. Of course, more evidence provided by further biochemical studies is required to establish the case.

Acknowledgement:

This work was supported by the start-up fund and a seed grant from Southern Illinois University Carbondale to Z.D.

References:

- [1] Gesteland RF *et al. Science*. 1992 **257**: 1640
- [2] Baranov PV *et al. Gene*. 2002 **286**: 187
- [3] Pleij CW *et al. Nucleic Acids Res*. 1985 **13**: 1717
- [4] Pleij CW *et al. Trends in Biochemical Science*. 1990 **15**: 143
- [5] Dam ET *et al. Biochemistry*. 1992 **31**: 11665
- [6] Brierley I, *J Gen Virol*. 1995 **76**: 1885
- [7] Farabaugh PJ, *Microbiol Rev*. 1996 **60**: 103
- [8] Giedroc DP *et al. J Mol Biol*. 2000 **298**: 167
- [9] Brierley I *et al. Nat Rev Microbiol*. 2007 **5**: 598
- [10] Clark MB *et al. J Biol Chem*. 2007 **282**: 37359
- [11] Wills NM *et al. Journal of Biological Chemistry*. 2006 **281**: 7082
- [12] Thompson JD *et al. Curr Protoc Bioinformatics*. 2002 **Chapter 2**: 23.
- [13] Li L *et al. J Virol*. 2001 **75**: 10612

Edited by P Kanguane

Citation: Huang *et al. Bioinformation* 9(14): 736-738 (2013)

License statement: This is an open-access article, which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes, provided the original author and source are credited