Turkish Journal of Biology

Volume 39 | Number 5 Article 3

1-1-2015

A small intron of Drosophila gamma carboxylase, not predictable in silico for mammalian splicing machinery, is spliced improperly in human cells

SOLMAZ MONIRI JAVADHESARI

ALIREZA ZOMORODIPOUR

SPIDEH VALIMEHR

TÜBİTAK

Turkish Journal of Biology

http://journals.tubitak.gov.tr/biology/

Research Article

Turk J Biol
(2015) 39: 674-681
© TÜBİTAK
doi:10.3906/biy-1411-5

A small intron of *Drosophila* gamma carboxylase, not predictable in silico for mammalian splicing machinery, is spliced improperly in human cells

Solmaz MONIRI JAVADHESARI[1,2], Alireza ZOMORODIPOUR[1,*], Spideh VALIMEHR[1]

[1]Department of Molecular Medicine, Faculty of Medical Biotechnology, National Institute of Genetic Engineering and Biotechnology (NIGEB), Tehran, Iran

[2]Department of Biology, Faculty of Basic Science, Azarbaijan Shahid Madani University, Tabriz, Iran

Received: 03.11.2014 • **Accepted/Published Online:** 06.03.2015 • **Printed:** 30.09.2015

Abstract: The splicing of two small introns of 58 and 67 nucleotides (nts) of the *Drosophila* gamma carboxylase (DγC) in cultured human cells was examined. The two introns are below the minimum size, which was proposed as a requirement for an intron to be spliced in human cells. None of the well-known software examined was able to provide a proper prediction for the DγC intron 1 in a mammalian host. Our in vivo experiments recognized only donor splice sites (not the acceptor site) of the DγC intron 1 by the mammalian cell, which, together with a new splicing acceptor site, generated an intron of 159 nts that was not predictable in silico. The experimental results showed that, for an intron to be spliced in mammalian splicing machinery, a minimum intron-length and acceptable branch point (BP) are more important than a pyrimidine-rich stretch in the 3′-region and wild-type splice sites. The in vivo approach, such as the one reported in this work, seems reliable for inspecting the interchangeability of introns among species.

Key words: Gamma carboxylase, intron splicing, small intron, *Drosophila melanogaster*

1. Introduction

Most eukaryotic protein-coding genes are interrupted by introns, which must be precisely removed from the transcript precursor in order to produce functional mRNAs (Gilbert, 1978). This phenomenon, which is commonly referred to as the premRNA splicing process, occurs through the formation of spliceosome (Nilsen, 2003). Recognition of a splice site (ss) by a spliceosome complex takes place via either intron- or exon-definition splicing mechanisms (Talerico and Berget, 1994; Berget, 1995; Sterner et al., 1996; Romfo et al., 2000). In the intron-definition mechanism, the 5′ss and 3′ss are initially recognized and paired across the intron (Lim and Burge, 2001), whereas in the exon-definition model the 5′ss and 3′ss are paired first across the exons (Berget, 1995). Based on the analysis of a number of complete genomes, in higher eukaryotes the introns are rather large, with an average size of 3–6 kb (Lander et al., 2001; Zhu et al., 2009), and recognition of splice sites for such introns takes place through the exon-definition mechanism (Berget, 1995). In lower eukaryotes such as yeast (Goguel and Rosbash, 1993) in which genome architectures are characterized by small introns (<100 nts) and large exons, the intron-definition mechanism is probably predominant for the recognition of splice sites (De Conti et al., 2013).

Drosophila melanogaster, with a simpler karyotype than that of humans due to reduced number of chromosomes, contains a combination of very small introns (51–80 nts), characteristic of lower eukaryotes, and larger introns, similar to those found in vertebrates (Hawkins, 1988; Mount et al., 1992; Kennedy and Berget, 1997; Ceprani et al., 2014). However, more than half of Drosophila introns are in the size-range of 59–67 nucleotides, which is shorter than the minimum size of the introns in vertebrates and mammals (Wieringa et al., 1984; Talerico and Berget, 1990; Mount et al., 1992). In spite of similarities in the intron splice sites and conservation of splicing system components in Drosophila and humans (Mount and Steitz, 1981; Guthrie and Patterson, 1988; Paterson et al., 1991), there are major differences between the sequences required for intron splicing in these two species (Guo et al., 1993). The species-specificity of intron splicing signals has also been attributed to the A + T content of introns and exons in different species (Goodall and Filipowicz, 1989, 1991). In contrast to mammals, in *C. elegans* and insects (e.g., Drosophila) and other species such as plants and Tetrahymena thermophila the A + T content of introns is higher than that of exons. No significant difference is observed between exons and introns in mammals and

* Correspondence: zomorodi@nigeb.ac.ir

yeast in this regard (Mount et al., 1992). In addition, the distance between 5′ss and branch-points (BPs) differs considerably in Drosophila and humans (Guo et al., 1993). The small Drosophila introns statistically lack the so-called pyrimidine tract, a major eukaryotic splicing feature, between the BP and 3′ss (Mount et al., 1992; Moore et al., 1993). Unlike most of the mammalian introns, the Drosophila introns do not have G in the position preceding the branched nucleotide (McCullough and Schuler, 1993).

Evidence demonstrating the splicing defects of introns in heterologous systems has been reported (Wieringa et al., 1984; Fu et al., 1988). In vitro studies showed that small Drosophila introns of less than 70 nts such as mle gene are not spliced in mammalian extract, probably due to the lack of a pyrimidine tract (Guo et al., 1993; Talerico and Berget, 1994). Increasing the size of such introns greatly reduced the efficiency of splicing introns longer than 79 nucleotides in Drosophila extracts but had an opposite effect in human extracts, in which introns longer than 78 nucleotides were spliced with much greater efficiency (Guo et al., 1993; Talerico and Berget, 1994). Alteration of sequences adjacent to the 3′ss of a typical small intron of Drosophila, such as White gene, to create a pyrimidine stretch was necessary for splicing in human, but not in Drosophila, extracts (Guo et al., 1993).

Genes encoding γ-carboxylase (γC) have been identified in vertebrates and two invertebrate species, mollusks and *D. melanogaster* (Goodstadt and Ponting, 2004; Robertson, 2004; Bandyopadhyay et al., 2006). The potential of the Drosophila γC (DγC) for carboxylation of some human vitamin-K–dependent (VKD) proteins was demonstrated previously in both in vitro (Bandyopadhyay et al., 2006) and in vivo (Vatandoost et al., 2012) studies. In order to investigate the potential of the DγC to γ-carboxylate a human substrate in a mammalian host after cloning of the DγC gene, 2144 base pairs (bp), and its corresponding cDNA (2019 bp) in separate expression vectors, we studied the expression of DγC in a mammalian cell line (in preparation). We used the Drosophila complete native gene for expression of DγC to examine whether the two DγC small introns, with 58 and 67 bps, are recognizable by a mammalian host. If they were, we could take advantage of the presence of a small native intron to improve the DγC expression.

It has been argued that human and Drosophila splicing machineries are not necessarily interchangeable (Guo et al., 1993). A limited number of in vitro experiments have indicated the deficiencies of mammalian cells for recognition of small Drosophila introns (Guo et al., 1993; Talerico and Berget, 1994). However, no in vivo data have been reported in this regard. Now, we had the opportunity to address this question in vivo. In this paper we report the data obtained from transcriptional analysis of the DγC gene in a mammalian host in parallel with our in silico inspections.

2. Materials and methods

2.1. Bacterial strain, mammalian cell line, and nucleic acid resources

The DH5α strain of *Escherichia coli* (Stratagene, USA) was used for cloning steps. Human embryonic kidney 293T (HEK293T) cell line (ATCC, USA) was used as the expression host. Plasmid pcDNA3 (Invitrogen, USA) was used for the construction of the DγG-expressing plasmid.

2.2. Media, enzymes, chemicals, and kits

Luria–Bertani was used as the bacterial culture medium, and ampicillin (100 μg/mL) was added when required to maintain selection pressure. Enzymes including HindIII, XhoI, T4 DNA ligase, and reverse transcriptase (M-MuLV) and kits for PCR product purification, plasmid isolation, and TriPure isolation reagent in addition to X-tremeGENE HP DNA transfection reagent were purchased from Roche, Germany. Fetal bovine serum (FBS) was obtained from Biowest, South America. The PCR product cloning kit (InsT/Aclone) and Taq DNA polymerase were obtained from Fermentas, Canada.

2.3. Nucleic acid manipulations

All DNA manipulations were carried out based on standard procedures (Sambrook and Russell, 2001). The *Drosophila melanogaster* genomic DNA was extracted using the TriPure isolation reagent (Roche, Germany) according to the manufacturer's instructions and subjected to polymerase chain reaction (PCR) to amplify the DγC gene. Total Drosophila RNA was extracted using RNA-high pure kit (Roche, Germany) and subjected to reverse-transcriptase polymerase chain reaction (RT-PCR) to amplify the corresponding DγC cDNA. Oligonucleotides were synthesized by TAG, Copenhagen, Denmark: DγCF (5′- AAGCTTGCCAC CATGGCCAACTCCAAGCGAAAG-3′) with HindIII site (underlined) in forward direction and DγCR (5′-CTCGAGGTTAGCCTTCTCCTTCAATCTCCC-3′) with XhoI site (underlined) in reverse direction, were designed based on the DγC mRNA sequences (accession no.: NM_079161.4). PCR and RT-PCR were carried out after 5 min at 94 °C and followed by 30 cycles, each starting with 30 s at 94 °C, 45 s at 56.5 °C, and 140 s at 72 °C, followed by a 10 min final extension at 72 °C (Altschul et al., 1990).

The amplified DγC-related fragments, including the DγC complete gene and its corresponding cDNA, were inserted in HindIII/XhoI sites in two expression plasmids following a first round of cloning in T/A vectors (pTZ57R/T) which ended with construction of plasmids pDγCint and pDγC. A double HindIII/XhoI digestion of pDγC created two fragments of 2046 and ~5360 bps, corresponding to the

intron-less gamma carboxylase coding sequence and the pcDNA3-related fragment, respectively. A double, HindIII/XhoI digestion of pDγCint created two fragments (2171 and ~5360 bps) corresponding to the intron-containing DγC and the pcDNA3-related fragment, respectively (data not shown). After verification by restriction analysis and nucleotide sequencing, the resulting DγC-expressing plasmids were used to transfect HEK293T cells separately for further expression analysis. A nested PCR approach was taken to amplify the band corresponding to the spliced transcript using a pair of nested primers, DγCNF: 5′ AGCTGCCTGGCCTTTCTG 3′ and DγCNR: 5′ TTTTGGACCAGTTGCCTAGC 3′, for further sequence analysis.

A complete sequence of DγC gene, obtained in this work, is now available (GenBank accession no.: KM056976.1).

2.4. The mammalian cell growth and transfection conditions

The mammalian cell line was grown in DMEM/F12 medium containing 10% FBS (as complete medium) at 37 °C in a 5% (v/v) CO_2 atmosphere. The cultured medium was refreshed every 2–3 days. Upon 60%–75% confluency the medium was removed, and the cells were washed once with Ca^{2+} and Mg^{2+}-free D-PBS and treated with a solution containing 0.25% trypsin and 1 mM EDTA for 5 min at 37 °C. During incubation, it was aspirated every 1–2 min to dissociate the cells. Trypsin was neutralized with 10% FBS in the final trypsin solution, and the cells were pelleted at 100 × g for 5 min at 4 °C. The pellet was gently resuspended in approximately 2–5 mL of complete medium (DMEM + F12 + 10% FBS).

The cells obtained from passages 3–9 were seeded into 6-well culture plates (2.5×10^5 cells/well/2 mL of complete medium). After 21 ± 3 h, they were subjected to transfection procedure. Briefly, 1 µg of (~1–5 µL) the DNA construct (either pDγC or pDγCint) was added to 97 µL of basal DMEM/F12 and incubated at 25 °C for 5 min. Subsequently, X-tremeGENE HP (3 µL) was added to the mixture and mixed gently, and incubation continued for 30 min. The final transfection mix was added directly to each well. The transfection step continued for up to 72 h, and the cells were collected for further expression analysis.

2.5. Networking and software

The complete nucleotide sequences of the DγC gene (gene ID: 38194 GC) and cDNA, were retrieved from GenBank (accession nos.: NT-037436.3 and NM-079161.4, respectively) [http://www.ncbi.nlm.nih.gov] (Benson et al., 2006). For homology searches and sequence alignments, Blast programs (Altschul et al., 1997) were used. In order to predict the splicing pattern of the introns in both human and *D. melanogaster* cells, the sequence of the intron-containing fragment was evaluated by a number of available online programs including Genscan (http://genes.mit.edu/GENSCAN.html) (Burge and Karlin, 1997), Genmark (http://opal.biology.gatech.edu/eukhmm.cgi) (Lomsadze et al., 2005), Fgenesh (http://linux1.softberry.com/berry.phtml?topic=fgenesh&group=programs&subgroup=gfind) (Solovyev et al., 2006), and Netgene (http://www.cbs.dtu.dk/services/NetGene2/) (Hebsgaard et al., 1996). Subsequently, program outputs were interpreted based on the data available in NCBI (GenBank) and experimental data.

3. Results

3.1. Isolation of the DγC gene and cDNA

Following the extraction of the Drosophila total RNA, our efforts towards the amplification of the DγC coding sequence resulted in two DNA fragments of about 2 kb and 2.2 kb. Considering the lengths of two DγC introns with 58 and 67 bps, the larger and shorter fragments were assumed to be the DγC gene and cDNA, respectively (Figure 1).

The presence of introns in the larger fragment was confirmed by alignment of the nucleotide sequences of the two amplified fragments. As shown in Figure 2, comparison of the nucleotide sequence of the amplified DγC gene against GenBank (NCBI) detected a "T" to "C" substitution at position 14 downstream of the 5′ss of the first intron, a 5 bp deletion between positions 13 and 17 upstream of the 3′ss, and a "T" to "C" substitution at position 22 upstream of the 3′ss in the second intron. The sequence of the DγC cDNA shows that the two introns of the cognate gene are spliced properly in the native host, ruling out any possible effect of the detected mutations on the splicing of their corresponding introns in Drosophila. In silico translation of the DγC cDNA led to a 672 amino acid (aa) protein similar to the DγC, differing in 6 amino acids, which seems to be single nucleotide polymorphisms (SNPs).

The two 58 and 67 bps introns of the DγC follow the common features of the typical Drosophila small introns, which are less frequent in mammals. As expected, the A + T content of these introns [>60% (intron 1) and 71% (intron 2)] is higher than that of the DγC exons which produced 44.63%, 51.87%, and 45.07% A + T content for exons 1–3, respectively. Each of the two introns contains only a minimal region of interrupted pyrimidines. Sequence analysis of the two DγC introns detected some candidate sequences with weak small similarities to the Drosophila splicing BP (CTAAT) (Mount et al., 1992). We examined the splicing of the DγC small introns in a mammalian host, both in silico and in vivo, as outlined in the subsequent steps.

3.2. In silico prediction of DγC introns splicing

The nucleotide sequence of DγC gene was subjected to analysis using four well-known splicing-prediction

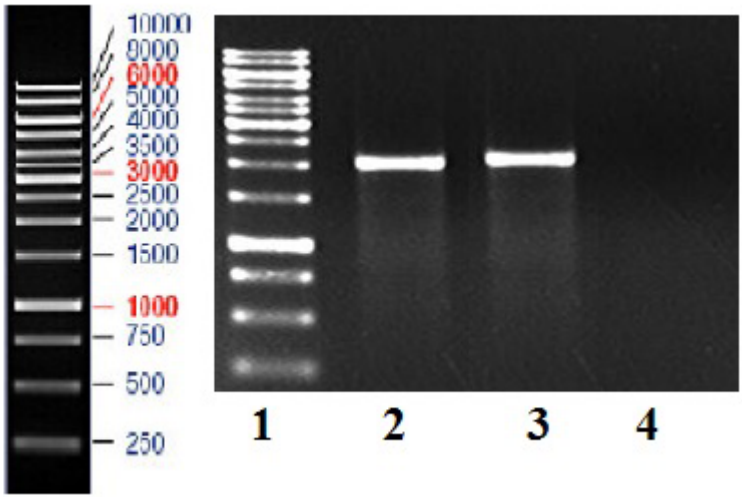

Figure 1. Amplifications of DγC-corresponding gene and cDNA using the gene-specific primers. Lane 1: DNA size marker, 2: DγC cDNA, 3: DγC gene, and 4: negative PCR control.

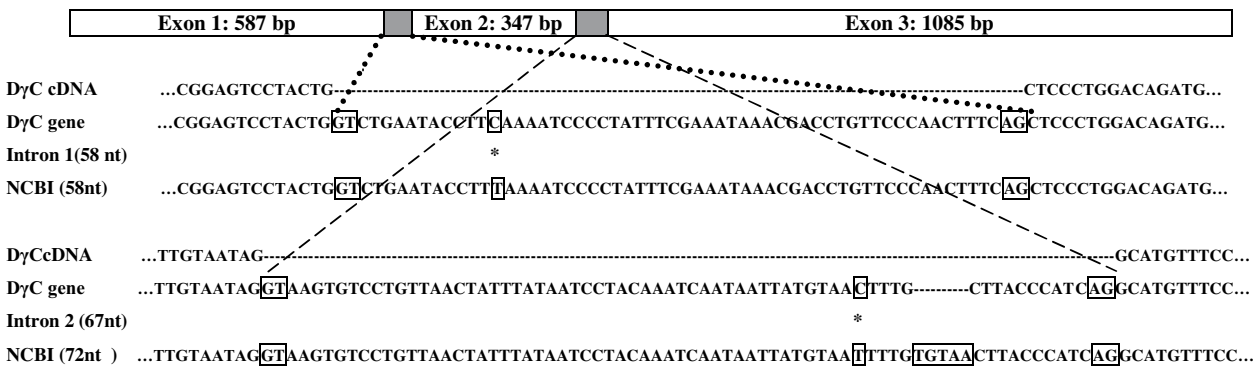

Figure 2. Schematic view of DγC exon/introns. Alignment of the nucleotide sequences of the DγC cDNA against the DγC gene and GenBank. The intronic donor and acceptor splice sites are boxed. SNPs are indicated by *.

programs: Genscan, Genmark, Fgenesh, and Netgene, applicable for genes of various eukaryotic origins. The data obtained from the applied programs are summarized in the Table.

Based on the data obtained from the analysis of the DγC gene by the above-mentioned software, its second intron (67 nt) was predicted to be recognizable by both vertebrate and invertebrate splicing machineries (Table), whereas different donor and acceptor splice sites were suggested for the DγC intron 1 by different programs. These sites are summarized in Figure 3. The lengths of the products expected from splicing of the DγC appeared to be different in each program.

According to Fgenesh, splicing patterns of the DγC in both Drosophila and human cells are similar but different from wild-type splicing. Based on this prediction (Figure 4), the positions of the donor and acceptor splice sites of the DγC intron 1 were different from those of their native counterparts, whereas the second DγC intron (67 nt) was predicted to be spliced correctly by both human and Drosophila splicing machineries. Thus, a final product of 633 aa was expected from the spliced DγC transcript predicted by Fgenesh, which is 39 aa shorter than the native enzyme. Netgene was able to predict only the DγC intron 1 native acceptor splice site (but not its donor site) properly, when human splicing machinery was selected as host (Table).

3.3. In vivo analysis of the DγC-intron splicing in a mammalian cell line

Transfections of the HEK293T cells with either pDγC or pDγCint were evaluated by RT-PCR to demonstrate the transcriptions of their corresponding transgenes. As shown in Figure 5, the electrophoresis pattern of the RT-PCR products of the pDγCint-corresponding transgene displayed two bands of around 2.1 and 2 kb. The longer band corresponded to the DγC unspliced transcript, and the shorter one seemed related to the mature (properly spliced) DγC transcript. Taking a nested PCR approach,

Table. Splicing pattern of the DγC introns. (A) Using splicing prediction software Genscan, Genmark, Fgenesh, and Netgene; and (B) based on experimental (in vivo) evidence. Base pair (bp), nucleotide (nt), amino acid (aa).

Splicing analysis			Predictions				Expected protein length (aa)
			Intron 1		Intron 2		
	Program	Species	Length (bp)	Position (nt)	Length (bp)	Position (nt)	
In silico	Genscan	Vertebrates	28	547–574	67	993–1059	682
	Genmark	*Drosophila*	58	588–645	67	993–1059	672
	Genmark	*H. sapiens*	28	547–574	67	993–1059	682
	Fgenesh	*Drosophila*	175	547–721	67	993–1059	633
	Fgenesh	*H. sapiens*	175	547–721	67	993–1059	633
	Netgene	*H. sapiens*	??	??–645	67	993–1059	
	Netgene	*C. elegance*	??	588–??	67	993–1059	
In vivo	Species						
	Drosophila	*Drosophila*	58	588–645	67	993–1059	672
	Mammalian cell	*H. sapiens*	159	588–746	------	----------	226

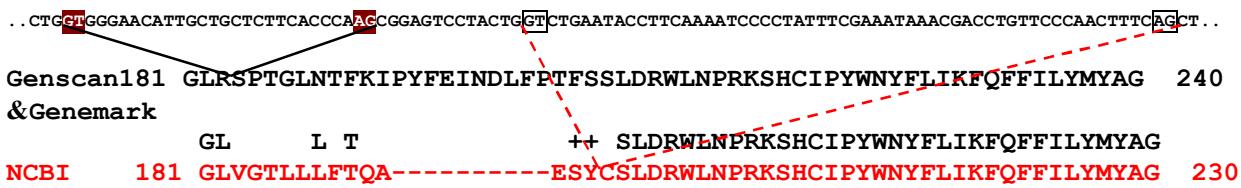

Figure 3. Splicing pattern of the DγC intron I and the corresponding aa sequence based on Genscan and Genmark prediction programs in comparison with the data obtained from GenBank (NCBI). The donor/acceptor sites predicted by Genescan and *Drosophila* cell (in vivo) are shown in colored and colorless boxes, respectively.

Figure 4. Splicing pattern of the DγC intron 1 and the corresponding aa sequence based on Fgeneh prediction program in comparison with the data obtained from GenBank (NCBI). The donor/acceptor sites predicted by Fgenesh and *Drosophila* cell (in vivo) are shown in colored and colorless boxes, respectively.

the shorter band was reamplified and subjected to sequence analysis, which demonstrated a splicing pattern of the DγC transcript precursor by HEK293T host cell, which is different from its native splicing in Drosophila. Figure 6 shows the sequence analysis of this transcript. According to these data, only the donor splice site of the DγC intron 1 was recognizable by the mammalian host, and the corresponding acceptor ss appeared to be different from that of the native DγC intron 1 that caused the generation of an intron of 159 nts. Interestingly, a sequence similarity of 8 bp was detected on the intron donor and acceptor splice site boundaries. In contrast to our in silico data in the Table, the second DγC intron was not recognizable by the mammalian host in vivo.

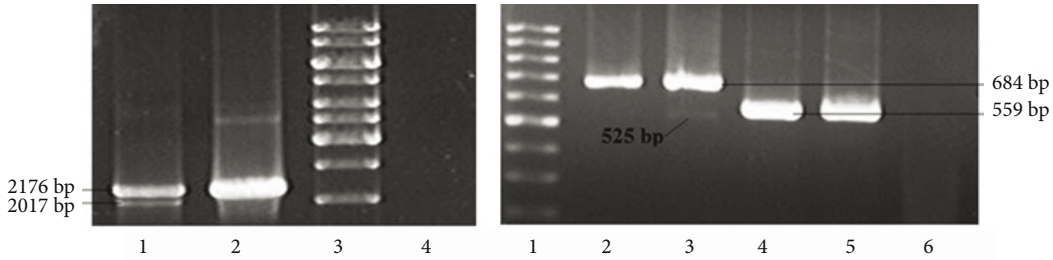

Figure 5. Electrophoretic pattern of the products obtained from PCR and RT-PCR, performed on the plasmid DNAs and transcripts of the transfected cells, respectively. Left panel: using full-length primers (DγCF and DγCR); lanes 1 and 2: DγC fragments amplified from cDNA and DNA of cells transfected by p*DγC*int, lane 3: 1kb DNA size marker, lane 4: negative control. Right panel: using nested primers (*DγC*NF and *DγC*NR); lane 1: 100 bp DNA size marker, lanes 2 and 4: DγC fragments amplified from DNAs of cells transfected by pDγCint and pDγC, respectively. Lanes 3 and 5: DγC fragments amplified from cDNAs of cells transfected by pDγCint and pDγC, respectively. Lane 6: negative control.

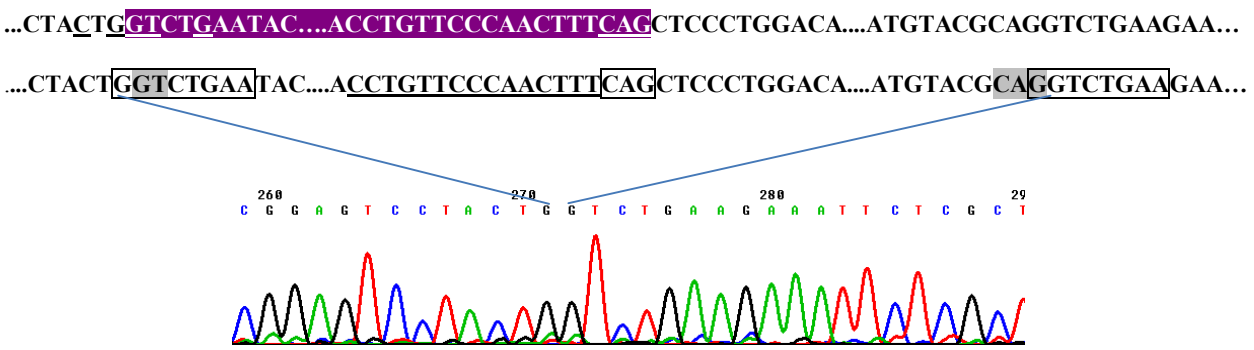

Figure 6. Sequence analysis of the DγC cDNA amplified from the total RNA of the cells harboring p*DγC* int. Colored box in the first row represents the wild-type intron 1. In the second row, the 5′ss and 3′ss recognized by mammalian cells in vivo are presented as gray boxes, and the location of the resulting splice junction site is indicated by arrow heads. The boxed CAG in the middle represents the wild-type 3′ss after a pyrimidine-rich region (underlined). The two 8 nts repeats are boxed in the alternately recognized intronic borders.

4. Discussion

Early studies of heterologous intron splicing indicated the species specificity of this biological phenomenon (Wieringa et al., 1984; Fu et al., 1988). In this regard, small introns of Drosophila, in which many of the typical features of vertebrate introns are missing, have provided the opportunity to evaluate the above-mentioned state experimentally. In spite of some similarities between the Drosophila and mammalian splicing systems, there are enough species-specific differences that the splicing of typical small introns of Drosophila by mammalian cell machinery may be prevented. A minimal length of an intron in HeLa cells was previously suggested (Wieringa et al., 1984). In this context, two requirements (length greater than 80 nucleotides and a good pyrimidine stretch) were deemed necessary for introns to be spliceable in human cell extracts (Guo et al., 1993).

The present work was the first in vivo study to examine the interchangeability of introns among distantly related species. Experimental evidence provided in this work showed that among the DγC native splicing sites only the intron 1 donor site was recognizable by the mammalian host, and neither the intron 1 acceptor site nor the splicing sites of the second intron were detectable by the same splicing machinery. Instead, a new splicing acceptor site was recognized by the mammalian host, which caused a splicing pattern for the DγC first intron that is different from its native splicing. Our in vivo data were in contrast with the results obtained from the in silico predictions. Indeed, none of the examined software programs were able to provide a proper prediction for the DγC intron 1 in a mammalian host. Neither were they able to predict the newly formed 159 bp intron for a mammalian host. Among the applied programs, Netgene, which is applicable for prediction of donor/acceptor splice sites in various species, was only able to properly predict the acceptor splice site (but not the donor site) of the DγC intron 1 when the human cell was considered as host.

On the other hand, although the examined software predicted a successful splicing of the second intron in a mammalian with cell, it was not feasible in vivo. In fact, the examined programs appeared as ineffective for prediction of typical Drosophila small introns, such as those of the DγC. Moreover, they are not suitable for prediction of interchangeability or splicing of small introns among distantly related species. The failure of the of the examined programs to predict Drosophila small introns could be due to the need for more input criteria concerning the intronic structures and splicing requirements for the applied programs. Therefore, additional features rather than intron size, such as the intronic A + T content in addition to the cis-acting elements including; pyrimidine tract and the splice BP, might be considered when in silico prediction of small intron splicing in a heterologous host is the aim. An in vivo approach such as the one presented in this work seems to be reliable enough to preview the fate of an exogenous intron in a heterologous host and to demonstrate the limitations of the application of introns as molecular tools among eukaryotic species.

Splicing of small introns in Drosophila extracts was observed using Drosophila Kc cells extract (Rio, 1988). Kennedy and Berget (1997) reported the splicing failure of a small Drosophila intron of mle gene in mammalian cell extract in spite of the presence of a minimal region of interrupted pyrimidines downstream to the BP and two longer uninterrupted C-rich tracts between the 5' splice site and branch. They suggested that a very short intron, lacking a classical pyrimidine tract between the BP and the 3′ splice site for assembly of active spliceosomes, requires accessory pyrimidine sequences in the short region between the 5′ss and BP. In contrast to their assumption, the DγC native intron 1 was not spliceable in a mammalian host despite a good pyrimidine stretch in its 3′ region. However, occurrence of splicing for the 159 nts intron, which has the pyrimidine stretch on its 5′ region (Figure 6), is in accordance with their assumptions. The optimum distance between the BPs and the 3′ss of an intron that is recognized by the mammalian cell is 21–34 nts (Gao et al., 2008). In the case of the DγC 159 bp intron, a consensus BP sequence (TCCTAAT) (Gao et al., 2008) was localized 34 nts upstream to its corresponding 3′ss. This feature is similar to the BP consensus found in both human and yeast introns (Gao et al., 2008). An interesting feature of the DγC 159 bp intron is the sequence similarity of 8 bp that is observed on both splice site boundaries. So far no such repeats have been reported as required accessories of intron splicing. Thus, any possible correlation between these short direct repeats and the intron fate remains to be elucidated.

The experimental results presented in this work, in accordance with previous findings, confirmed the species-specific criteria of the introns and emphasize the minimal length requirement for an intron to be spliced properly by a mammalian host. A mammalian splicing system, at least in this case, preferably recognizes introns longer than the minimal size requirement of mammalian small introns, rather than those with a pyrimidine-rich stretch in their 3′-region and wild-type splice sites. Our data also confirmed the presence of a proper BP for the 159 bp intron that might also play a key role in the splicing process in a mammalian cell. Therefore, it is thought that for mammalian splicing machinery a minimum intron length, in addition to an acceptable BP, is necessary for an intron to be spliced.

Acknowledgments

This work was performed as part of the requirement for the fulfillment of the degree of PhD in Molecular Genetics of Solmaz Moniri at NIGEB and was supported in part by an internal grant (no.: 419) at NIGEB.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990). Basic local alignment search tool. J Mol Biol 215: 403–410.

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25: 3389–3402.

Bandyopadhyay PK, Clark K, Stevenson BJ, Rivier JE, Olivera BM, Golic KG, Rong YS (2006). Biochemical characterization of *Drosophila* gamma-glutamyl carboxylase and its role in fly development. Insect Mol Biol 15: 147–156.

Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL (2006). GenBank. Nucleic Acids Res 34: D16–20.

Berget SM (1995). Exon recognition in vertebrate splicing. J Biol Chem 270: 2411–2414.

Burge C, Karlin S (1997). Prediction of complete gene structures in human genomic DNA. J Mol Biol 268: 78–94.

Ceprani F, Spirito F, Piergentili R (2014). Identification of new genes required for the maintenance of chromosome integrity in *Drosophila melanogaster*. Turk J Biol 38: 880–897.

De Conti L, Baralle M, Buratti E (2013). Exon and intron definition in pre-mRNA splicing. Wiley Interdiscip Rev RNA 4: 49–60.

Fu XY, Colgan JD, Manley JL (1988). Multiple cis-acting sequence elements are required for efficient splicing of simian virus 40 small-t antigen pre-mRNA. Mol Cell Biol 8: 3582–3590.

Gao K, Masuda A, Matsuura T, Ohno K (2008). Human branch point consensus sequence is yUnAy. Nucleic Acids Res 36: 2257–2267.

Gilbert W (1978). Why genes in pieces? Nature 271: 501.

Goguel V, Rosbash M (1993). Splice site choice and splicing efficiency are positively influenced by pre-mRNA intramolecular base pairing in yeast. Cell 72: 893–901.

Goodall GJ, Filipowicz W (1989). The AU-rich sequences present in the introns of plant nuclear pre-mRNAs are required for splicing. Cell 58: 473–483.

Goodall GJ, Filipowicz W (1991). Different effects of intron nucleotide composition and secondary structure on pre-mRNA splicing in monocot and dicot plants. Embo J 10: 2635–2644.

Goodstadt L, Ponting CP (2004). Vitamin K epoxide reductase: homology, active site and catalytic mechanism. Trends Biochem Sci 29: 289–292.

Guo M, Lo PC, Mount SM (1993). Species-specific signals for the splicing of a short *Drosophila* intron in vitro. Mol Cell Biol 13: 1104–1118.

Guthrie C, Patterson B (1988). Spliceosomal snRNAs. Annu Rev Genet 22: 387–419.

Hawkins JD (1988). A survey on intron and exon lengths. Nucleic Acids Res 16: 9893–9908.

Hebsgaard SM, Korning PG, Tolstrup N, Engelbrecht J, Rouze P, Brunak S (1996). Splice site prediction in *Arabidopsis thaliana* pre-mRNA by combining local and global sequence information. Nucleic Acids Res 24: 3439–3452.

Kennedy CF, Berget SM (1997). Pyrimidine tracts between the 5' splice site and branch point facilitate splicing and recognition of a small *Drosophila* intron. Mol Cell Biol 17: 2774–2780.

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W et al. (2001). Initial sequencing and analysis of the human genome. Nature 409: 860–921.

Lim LP, Burge CB (2001). A computational analysis of sequence features involved in recognition of short introns. Proc Natl Acad Sci USA 98: 11193–11198.

Lomsadze A, Ter-Hovhannisyan V, Chernoff YO, Borodovsky M (2005). Gene identification in novel eukaryotic genomes by self-training algorithm. Nucleic Acids Res 33: 6494–6506.

McCullough AJ, Schuler MA (1993). AU-rich intronic elements affect pre-mRNA 5' splice site selection in *Drosophila melanogaster*. Mol Cell Biol 13: 7689–7697.

Moore MJ, Query CC, Sharp PA (1993). 13 splicing of precursors to mRNA by the Spliceosome. Cold Spring Harbor Monograph Archive 24: 303–357.

Mount SM, Burks C, Hertz G, Stormo GD, White O, Fields C (1992). Splicing signals in *Drosophila*: intron size, information content, and consensus sequences. Nucleic Acids Res 20: 4255–4262.

Mount SM, Steitz JA (1981). Sequence of U1 RNA from *Drosophila melanogaster*: implications for U1 secondary structure and possible involvement in splicing. Nucleic Acids Res 9: 6351–6368.

Nilsen TW (2003). The spliceosome: the most complex macromolecular machine in the cell? Bioessays 25: 1147–1149.

Paterson T, Beggs JD, Finnegan DJ, Luhrmann R (1991). Polypeptide components of Drosophila small nuclear ribonucleoprotein particles. Nucleic Acids Res 19: 5877–5882.

Rio DC (1988). Accurate and efficient pre-mRNA splicing in *Drosophila* cell-free extracts. Proc Natl Acad Sci USA 85: 2904–2908.

Robertson HM (2004). Genes encoding vitamin-K epoxide reductase are present in *Drosophila* and trypanosomatid protists. Genetics 168: 1077–1080.

Romfo CM, Alvarez CJ, van Heeckeren WJ, Webb CJ, Wise JA (2000). Evidence for splice site pairing via intron definition in *Schizosaccharomyces pombe*. Mol Cell Biol 20: 7955–7970.

Sambrook J, Russell DW (2001). Molecular Cloning: A Laboratory Manual. Cold Spring Harbor, NY, USA: Cold Spring Harbor Laboratory Press.

Solovyev V, Kosarev P, Seledsov I, Vorobyev D (2006). Automatic annotation of eukaryotic genes, pseudogenes and promoters. Genome Biol 7: 1–12.

Sterner DA, Carlo T, Berget SM (1996). Architectural limits on split genes. Proc Natl Acad Sci USA 93: 15081–15085.

Talerico M, Berget SM (1990). Effect of 5' splice site mutations on splicing of the preceding intron. Mol Cell Biol 10: 6299–6305.

Talerico M, Berget SM (1994). Intron definition in splicing of small *Drosophila* introns. Mol Cell Biol 14: 3434–3445.

Vatandoost J, Zomorodipour A, Sadeghizadeh M, Aliyari R, Bos MH, Ataei F (2012). Expression of biologically active human clotting factor IX in Drosophila S2 cells: gamma-carboxylation of a human vitamin K-dependent protein by the insect enzyme. Biotechnol Prog 28: 45–51.

Wieringa B, Hofer E, Weissmann C (1984). A minimal intron length but no specific internal sequence is required for splicing the large rabbit beta-globin intron. Cell 37: 915–925.

Zhu L, Zhang Y, Zhang W, Yang S, Chen JQ, Tian D (2009). Patterns of exon-intron architecture variation of genes in eukaryotic genomes. BMC Genomics 10: 1471–2164.