P ELEMENT TEMPERATURE-SPECIFIC TRANSPOSITION: A MODEL FOR POSSIBLE REGULATION OF MOBILE ELEMENTS ACTIVITY BY pre-mRNA SECONDARY STRUCTURE

A. GULTYAEV [1], T. REDCHUK [2], A. KOROLOVA [3], I. KOZERETSKA [3]

[1] Leiden Institute of Chemistry, Netherlands
[2] Palladin Institute of Biochemistry, Kyiv, Ukraine
[3] Taras Shevchenko National University of Kyiv, Ukraine
 E-mail: iryna.kozeretska@gmail.com

P element is a DNA transposon, known to spread in genome using transposase activity. Its activity is tissue-specific and normally observed at high temperatures within 24 °C to 29 °C. Here, we present a predicted RNA secondary structure domain of P element pre-mRNA which could potentially regulate the temperature sensitivity of the P element activity. In canonical P elements, the structure is a small hairpin with double-helical part interrupted by a symmetric loop and a mismatch. In M type P elements, the A.A mismatch is substituted by an A-U base pair, stabilizing the structure. The hairpin structure covers the region involving the IVS-3 5' splice site and both pseudo-splice sites F1 and F2. While the IVS-3 and F1 binding sites of U1 snRNA are located in the double-stranded part of the structure, the F2 site is exposed in the hairpin loop. The formation of this structure may interfere with landing of U1 snRNA on IVS-3 site, while F2 is positioned for the interaction. Alignment of P element sequences supports the proposed existence of the hairpin, showing high similarity for this region. The hairpin structure, stable at low temperatures, may prevent correct IVS-3 splicing. Conversely, temperature-induced destabilization of the hairpin structure may result in the splicing at the proper IVS-3 splice site. Taking into account the increasing amount of data demonstrating the important influence of RNA folding on phenotypes determined by alternative splicing a model for possible regulation of the activity of mobile elements by pre-mRNA secondary structure seems intriguing.

***Key words**: P element, RNA structure, alternative splicing, transposase.*

Introduction. Mobile genetic elements are widespread in eukaryotic genome and play a significant role in mutagenesis, genome regulation, and evolution [1−4]. *P* element is a DNA transposon that spread in wild populations of *Drosophila melanogaster* after most common laboratory strains had been isolated in the early 20th century [5]. Wild stocks carrying *P* element are referred to as P strains, and lab stocks that lack these elements are referred to as M strains [5, 6]. Crosses between P strain males and M strain females thus lead to a phenomenon called P-M hybrid dysgenesis, which is characterized by *P* element mobilization and reduced fertility in F1 progeny, restored in females with aging [7]. Reciprocal crosses between P strain females and M strain males produce genetically identical female progeny, but these hybrids are viable and fertile due to maternal deposition of P element piRNAs [8]. *P* element transposition is tissue- and temperature-specific; transposition occurs at high rates in the germ line but is not detected in somatic tissues. Furthermore, *P* element activity in the germ line is normally observed at temperatures within 24 to 29 °C and absent at temperatures lower than 24 °C [9].

P element is known to spread in genome using its transposase activity. The transposase pre-mRNA, encoded by *P* element, undergoes alternative splicing in a tissue-specific fashion. The canonical transposase gene consists of 4 exons. In germline cells, all the three introns are excised with the subsequent translation of an 87-kDa active protein. Translation of *P* element mRNA retaining the third intron in somatic tissues leads to the synthesis of a truncated 67-kDa product. The truncated protein not only fails to work as a transposase, but also inhibits the transposition [10]. The intron retention is determined by a 97 kDa protein named *P* element somatic inhibitor (PSI), which is specifically expressed in somatic cells and functions as a tissue-specific splicing repressor [11]. Splicing assays, mutagenesis and RNA binding studies indicate that PSI binds to unprocessed transposase pre-mRNA and recruits U1 snRNP to pseudo-splicing sites located 12 to 31 bases upstream of the 5' end of the 3d intron, precluding its excision [12−14].

As mentioned above, in germline cells splicing occurs only at relatively high temperatures above 24 °C. In spite of the fact that the origin of PSI-determined tissue specificity is now pretty well understood, little is known about the temperature dependence of the process.

While the temperature effect was first reported over 30 years ago, the nature of this mechanism remains unclear, perhaps due to the absence of proper methods to address this question. On the other hand, alternative splicing is known to be frequently regulated by RNA structures folded in pre-mRNA [15]. In particular, splicing efficiency may be sensitive to the thermodynamic stability of structures folded in the regions of splice sites [16].

In this study, we present a predicted RNA secondary structure domain of *P* element pre-mRNA which could potentially regulate the temperature sensitivity of the *P* element transposase activity in germ cells.

Materials and methods. RNA secondary structures were predicted on the basis of both comparative structure analysis and folding thermodynamics. In order to identify the structures conserved in related *P* elements, thus more likely to be functional, we used the RNAalifold algorithm, designed for prediction of common structures in RNA sequence datasets [17]. For development of initial models, representative *P* element sequences of different groups, such as canonical, were aligned and the multiple alignments were used as an input in the RNAalifold program. The regions of the *P* element intron 3 together with flanking exon sequences were used. The predictions yielded by RNAalifold algorithm were further analysed using free energy calculations by mfold program [18] and multiple alignments of *P* element sequences available in GenBank. This allowed us to estimate thermodynamic stability of predicted structures and the extent of their conservation in *P* element subfamilies.

Results. The most conserved structure in the area of intron 3 (IVS-3) was predicted by RNAalifold algorithm (see Methods) in the 3'-proximal part of exon 3, in the region of the 5' splice site (Fig. 1). In canonical *P* elements, the structure is a relatively small hairpin with double-helical part interrupted by a symmetric loop and a mismatch. In M type *P* elements, the A.A mismatch is substituted by an A-U base pair, apparently stabilizing the structure (Fig. 2). In the elements of O type and K type, some of the base pairs are disrupted, destabilizing the hairpin (not shown).

Interestingly, the hairpin structure covers the region involving the sequences of the IVS-3 5' splice site and both pseudo-splice sites F1 and F2 [12, 13] (Fig. 1). While the IVS-3 and F1 binding sites of U1 snRNA are mostly located in the double-
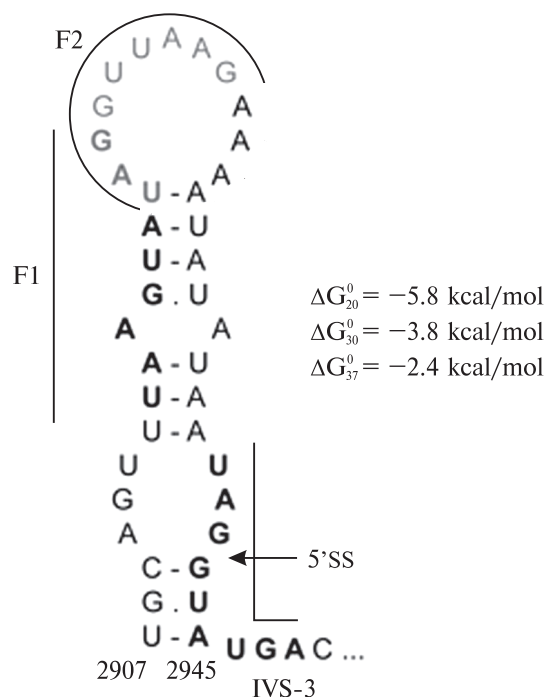

Fig. 1. *P* element RNA secondary structure prediction for the region around 5' splicing site of IVS-3. The canonical sequence of *P* element from *D. melanogaster* X06779, coordinates of F1, F2 pseudo SS and accurate 5'SS marked according [19]

$\Delta G^0_{20} = -5.8$ kcal/mol
$\Delta G^0_{30} = -3.8$ kcal/mol
$\Delta G^0_{37} = -2.4$ kcal/mol

stranded part of the structure, the F2 binding site is exposed in the hairpin loop. Obviously, the formation of this structure may interfere with landing of U1 snRNA on one of the three binding sites: most likely the IVS-3 and F1 sites are hidden in the stem, while F2 site is positioned for the interaction with U1 snRNA. Remarkably, the hairpin is significantly destabilized within the range of temperatures known to switch on the splicing process: while at 20 °C the free energy is −5.8 kcal/mol, its value is −3.8 kcal/mol at 30 °C and only −2.4 kcal/mol at 37 °C, as calculated using mfold program with the version 2.3 thermodynamic parameters [18].

Of course, the (temperature-dependent) formation of the hairpin in the 5' splice site region is not only dependent on its own folding free energy, but on the possible alternative conformations that may interfere with the hairpin folding. For instance, an alternative stem-loop structure is also possible in the same region, involving a larger part of IVS-3 (Fig. 2). Apparently, its folding would prevent the formation of the hairpin shown in Fig. 1.

```
Dmel       tggcagcatgcgatcgagaggtggacaattcgaccatcccactccactgcagtttaagta
Dneb       tggcagcatgcgatcgagaggtggacaattcgaccatcccactccactgcagtttaagta
Dwilcons1  tggcagcatgcgatcgagaggtggacaattcgaccatcccactccactgcagtttaagta
Dstucons1  aataacaagccgatcgagaggtggacaaatcgaccatccgactccactgcagtttaagta
Dstucons2  tggcagcatgcgatcgagaggtggacaattcgaccatccgactccactgcaattttaagta
Dstu7      tggcagcatgcgaccgaggtgtggacaattcgaccatccgactccactgcagattaagta
Dmed       tggcagcatgcgatcgagaggtggacaattcgaccatccgactccactgcagtttaagta
Lmik       tggcgctacgcgatcgaggggtggacactttgaccatccaactccattacagtttaagta
Dhel       tggggctatgcgatcgaggggtggacagtttgaccatccaactccactacagtttaagta
Dbif       tggcgctatgcgatcgaggggtggtcagtttgaccatccaactccactacagtttaagta
Spal       tggcgctatgcgatcgaggggtggtcagtttgaccatccaaccccactacagtttaagta
                                           hairpin:          (((...(((.(((
                                alternative hairpin:          (((((.(((

            |--F2---|             |--5'ss-|
Dmel       taggttaagaaaatatataataggtatgacaaatttaaaagaatgcgtaaacaaaaatgt
Dneb       taggttaagaaaatatataataggtatgacaaatttaaaagaatacgtaaacaaaaatgt
Dwilcons1  taggttaagaaaatatataataggtatgacaaatttaaaagaatgcgtaaacaaaaatgt
Dstucons1  taggttaagaaaatatataataggtatgacaaatttaaaaaaataagttaacaaaaatgt
Dstucons2  taggttaagaaaatatataataggtatgacaaatttacaagaataagttaccaaaaatgt
Dstu7      taggttaagaaaatatataataggtatgacaaatgtaaaaaaataaattaacaaaaatat
Dmed       taggttaagaaaatatatagtaggtatgacaaatttaaaagaataagttaataaaaatgt
Lmik       taggttaagaaaataattaataggtatgtcaaatttacaagaatgaattacaaaatgaat
Dhel       taggttaaaaaaatatttaataggtatgtcaaatttagaaaaatgaattacaa-attaat
Dbif       taggttaagaaaatatttaataggtatgtcaaatttagaaaaatgaattacaa-attaat
Spal       taggttaagaaaatatttaataggtatgtcaaatttagaagaatgaattacaaaattaat
            (..........))))).)))...)))
            (...(((((....(((((.....)))))....)))))....)))).)))))
```

Fig. 2. Alignment of the sequences in the region of IVS-3 5' splice site in canonical and M type *P* elements. Representative sequences were selected according to the phylogenetic tree of Loreto et al. [20]. Paired regions of the predicted RNA structures in canonical *P* element (according X06779) are shown by brackets. Canonical: *D. melanogaster* X06779 2860−2980 (Dmel), *D. nebulosa* M17424 1860−1980 (Dneb), *D. willistoni* AY578739 1834−1954 (Dwilcons1), *D. sturtevanti* AY578782 1691−1811 (Dstucons1), *D. sturtevanti* AY578776 1834−1954 (Dstucons2), *D. sturtevanti* AY578781 1792−1912 (Dstu7), *D. mediopunctata* AF313770 1827−1947 (Dmed). M type: *Lordiphosa miki* AF313772 478−598 (Lmik), *D. helvetica* AF313771 1846−1966 (Dhel), *D. bifasciata* X60990 1885−2005 (Dbif), *Scaptomyza pallida* M63342 1878−1998 (Spal). Substitutions as compared to the reference Dme1 are marked by grey boxes (for the putative hairpin structure regions only). F2 − the second pseudo-SS and the true 5'SS are marked accordingly [19]

Multiple alignment of *P* element sequences (Fig. 2) supports the proposed existence of the first hairpin 2907−2945 in the canonical and M type elements, showing rather high sequence similarity for this region among the analyzed genes. Only two substitutions led to mispairings: 553 T > A transition in *L. miki* and 1906 A > G in *D. mediopunctata*. The rest of the nucleotide differences are observed in regions predicted to be loops in the reference sequence. Moreover, transversions 554 A > T (bp numbering according to AF313772) even closed small internal loop (A.A mismatch) in the hairpin stem of M type structure. Similarly, 526 G > A (also observed in M

type species) converts the non-Watson-Crick GU pair into the canonical AU. The presence of transversions increasing the hairpin stability in M type elements is in accordance with the splicing patterns of *P* element-derived mRNAs of *D. bifasciata* and *S. pallida* and could possibly explain the immobility of the M type *P* elements subfamily [21].

As can be seen from the alignment, the alternative hairpin 2911−2971 region is less conservative among the analyzed species.

Discussion. Since up to 85 % of some eukaryotic genomes studied so far consist of mobile element [22], and sequences similar to drosophila *P* element have

been detected in the genomes of yeast, fish, birds, mammals, and even humans [23, 24], interest in understanding of the mechanisms of *P* element regulation reaches far beyond the biology of *D. melanogaster.*

It has long been known that the structure of pre-mRNA may have a strong effect of on spliceosome assembly, shown both *in vitro* and *in vivo* [25]. A variety of such examples have already been described in literature for different unrelated species, revealing their ubiqitiousness. A well-known illustration of the pre-mRNA secondary structure influence on splicing is the human gene *tau* [26, 16]. In this case, a stem-loop structure controls recognition of the 5'splicing site of an intron by spliceosome. Destabilization of the stem loop opens an access of splicing machinery to the splice site.

Similarly, the structure predicted in the *P* element pre-mRNA (Fig. 1) may suppress recognition of the IVS-3 5'splice site. Furthermore, this folding exposes one of the pseudo − splice sites in the hairpin loop, presumably facilitating U1 snRNA binding to the F2 pseudo splice site sequence. Thus, the hairpin structure, stable at low temperatures, may prevent correct IVS-3 splicing. At the elevated temperatures, the hairpin becomes less stable and/or is disrupted in favour of an alternative folding. This is expected to result in both increased efficiency of IVS-3 splice site recognition and a decrease in the F2 site binding by U1 snRNA.

An increased binding of U1 snRNA to a 5' splice site located in the loop of pre-mRNA secondary structure has been recently shown for the serotonin receptor 2C gene [27]. In this case, a dependence of U1 snRNA binding on RNA structure makes splicing sensitive to RNA conformational changes: the switching between alternative structures, triggered by the presence of a small molecule − pyrvinium palmoate, regulates splicing efficiency. Similarly, temperature-induced destabilization of the structure exposing the F2 pseudo splice site of *P*elements in the hairpin loop (Fig. 1), may result in the splicing at the proper IVS-3 splice site.

Such a transition may involve formation of alternative structures competing with the hairpin formation. We could readily identify a potential for formation of such structures (Fig. 2), especially because the hairpin structure does not have a low folding free energy. These structures are less conserved than the hairpin in the F1-F2 pseudo-splice site region, and may be specific for various

P element subfamilies. On the other hand, conformational transitions in the splice site regions may also involve complex tertiary structures such as e.g. pseudoknots, suggested to regulate the splicing of influenza virus mRNAs [28, 29]. We cannot exclude a possibility of similar effects of 3D RNA folding on the U1 snRNA binding to *P* element 5' splice site.

Apparently, the existence of the putative structures described in our study and their possible effect on the splicing of the 3rd intron still has to be confirmed by mutational studies. Taking into account the increasing amount of data demonstrating the important influence of RNA folding on phenotypes determined by alternative splicing [15, 30], a model for possible regulation of the activity of mobile elements by pre-mRNA secondary structure seems intriguing.

ТЕМПЕРАТУРОЗАВИСИМАЯ ТРАНСПОЗИЦИЯ *P* ЭЛЕМЕНТА: МОДЕЛЬ РЕГУЛЯЦИИ АКТИВНОСТИ МОБИЛЬНЫХ ЭЛЕМЕНТОВ ВТОРИЧНОЙ СТРУКТУРОЙ пре-мРНК

А. Гультяев, Т. Редчук, А. Королева, И. Козерецкая

Предсказана вторичная структура пре-мРНК *P* элемента, которая, возможно, регулирует его активность и температурную чувствительность. Структура представляет собой шпильку, более стабильную в *P* элементах М типа в сравнении с каноническими. Регион, образующий шпильку, находится в области 5' сайта сплайсинга третьего интрона (IVS-3), включая оба описанных псевдосайта сплайсинга − F1 и F2, в то время как истинный сайт и F1 расположены главным образом в двухцепочечной области шпильки, F2 экспонирован в петле. Выравнивание последовательностей *P* элементов продемонстрировало высокую степень сходства для указанного региона, что свидетельствует в пользу существования предсказанной структуры. Формирование шпильки, по нашему мнению, может препятствовать связыванию U1 snRNA с истинным сайтом сплайсинга, и, напротив, индуцированная температурой дестабилизация шпильки может приводить к корректному сплайсингу IVS-3. Таким образом, предложена гипотеза о влиянии вторичной структуры пре-мРНК мобильного элемента на его активность.

REFERENCE

1. *Kazazian H.H.Jr.* Mobile elements: drivers of genome evolution // Science. − 2004. − **303**, № 5664. − P. 1626−1632.
2. *Feschotte* C., *Pritham E.* DNA transposons and the evolution of eukaryotic genomes // Annu. Rev. Genet. − 2007. − **41.** − P. 331−368.

3. *Biemont C., Vieira C.* Junk DNA as an evolutionary force // Nature. − 2006. − **443**. − P. 521−524.

4. *González J., Petrov D.A.* The adaptive role of transposable elements in the Drosophila genome // Gene. − 2009. − **448**, № 2. − P. 124−133.

5. *Anxolabèhére D., Kidwell M.G., Periquet G.* Molecular characteristics of diverse populations are consistent with the hypothesis of a recent invasion of *Drosophila melanogaster* by mobile P elements // Mol. Biol. Evol. − 1988. − **5**, № 3. − P. 252−269.

6. *Rubin G.M., Kidwell M.G., Bingham P.M.* The molecular basis of *Drosophila melanogaster* hybrid dysgenesis : The nature of induced mutations // Cell. − 1982. − **29**, № 3. − P. 987−994.

7. *Khurana J.S., Wang J., Xu J. et al.* Adaptation to *P* element transposon invasion in *Drosophila melanogaster* // Cell. − 2011. − **147**, № 7. − P. 1551−1563.

8. *Brennecke J., Malone C.D., Aravin A.A. et al.* An epigenetic role for maternally inherited piRNAs in transposon silencing // Science. − 2008. − **322**, № 5906. − P. 1387−1392.

9. *Angels W.R., Prestons C.R.* Hybrid dysgenesis in *Drosophila melanogaster*: the biology of female and male sterility // Genetics. − 1979. − **92**, № 1. − P. 161−174.

10. *Misra S., Rio D.C.* Cytotype control of *Drosophila* P-element transposition: the 66 kd protein is a repressor of transposase activity // Cell. − 1990. − **62**, № 2. − P. 269−284.

11. *Siebel C.W., Admon A., Rio D.C.* Soma-specific expression and cloning of PSI, a negative regulator of *P* element pre-mRNA splicing // Genes & Dev. − 1995. − **9**, № 3. − P. 269−283.

12. *Chain A.C., Zollman S., Tseng J.C., Laski F.A.* Identification of a cis_Acting Sequence Required for Germ Line-specific Splicing of the *P* element ORF2-ORF3 intron // Mol. Cell Biol. − 1991. − **11**, № 3. − P. 1538−1546.

13. *Amarasinghe A.K., MacDiarmid R., Adams M.D., Rio D.C.* An in vitro-selected RNA-binding site for the KH domain protein PSI acts as a splicing inhibitor element // RNA. − 2001. − **7**, № 9. − P. 1239−1253.

14. *Ignjatovic T., Yang J.C., Butler J. et al.* Structural basis of the interaction between P-element somatic inhibitor and U1-70k essential for the alternative splicing of P-element transposase // J. Mol. Biol. − 2005. − **351**, № 1. − P. 52−65.

15. *Buratti E., Baralle F.E.* Influence of RNA secondary structure on the pre-mRNA splicing process // Mol. Cell Biol. − 2004. − **24**, № 24. − P. 10505−10514.

16. *Varani L., Hasegawa M., Spillantini M.G. et al.* Structure of tau exon 10 splicing regulatory element RNA and destabilization by mutations of frontotemporal dementia and parkinsonism linked to chromosome 17 // Proc. Nat. Acad. Sci. USA. − 1999. − **96**, № 14. − P. 8229−8234.

17. *Bernhart S.H., Hofacker I.L., Will S. et al.* RNAalifold: improved consensus structure prediction for RNA alignments // BMC Bioinform. − 2008. − **9**, № 11. − P. 474.

18. *Zuker M.* Mfold web server for nucleic acid folding and hybridization prediction // Nucl. Acids Res. − 2003. − **31**, № 13. − P. 3406−3415.

19. *Siebel C.W., Fresco L.D., Rio D.C.* The mechanism of somatic inhibition of *Drosophila* P-element pre-mRNA splicing: multiprotein complexes at an exon pseudo-5' splice site control U1 snRNP binding // Genes Dev. − 1992. − **6**, № 8. − P. 1386−1401.

20. *Loreto E.L.S., Zambra F.M.B., Ortiz M.F., Robe L.J.* New *Drosophila* P-like elements and reclassification of *Drosophila* P-elements subfamilies // Mol. Genet. Genom. − 2012. − **287**, № 7. − P. 531−540.

21. *Haring E., Hagemann S., Pinsker W.* Transcription and splicing patterns of M- and O-type P elements in *Drosophila bifasciata*, *D. helvetica*, and *Scaptomyza pallida* // J. Mol. Evol. − 1998. − **46**, № 5. − P. 542−551.

22. *Schnable P.S.* The B73 maize genome: complexity, diversity, and dynamics // Science. − 2009. − **326**, № 5956. − P. 1112−1115.

23. *Hagemann S., Pinsker W.* Drosophila P transposon in the human genome? // Mol. Biol. Evol. − 2001. − **18**, № 10. − P. 1979−1982.

24. *Hammer S.E., Strehl S., Hagemann S.* Homologs of *Drosophila* P transposons were mobile in zebrafish but have been domesticated in a common ancestor of chicken and human // Mol. Biol. Evol. − 2005. − **22**, № 4. − P. 833−844.

25. *Goguel V., Wang Y., Rosbash M.* Short artificial hairpins sequester splicing signals and inhibit yeast pre-mRNA splicing // Mol. Cell Biol. − 1993. − **13**, № 11. − P. 6841−6848.

26. *Grover A., Houlden H., Baker M. et al.* 5' Splice site mutations in tau associated with the inherited dementia FTDP-17 affect a stem-loop structure that regulates alternative splicing of exon 10 // J. Biol. Chem. − 1999. − **274**, № 21. − P. 15134−15143.

27. *Shen M., Bellaousov S., Hiller M. et al.* Pyrvinium pamoate changes alternative splicing of the serotonin receptor 2C by influencing its RNA structure // Nucl. Acids Res. − 2013. − **41**, № 6. − P. 3819−3832.

28. *Gultyaev A.P., Heus H.A., Olsthoorn R.C.L.* An RNA conformational shift in recent H5N1 influenza A viruses // Bioinformatics. − 2007. − **23**, № 3. − P. 272−276.

29. *Moss W.N., Dela-Moss L.I., Kierzek E. et al.* The 3' splice site of influenza A segment 7 mRNA can exist in two conformations: a pseudoknot and a hairpin // PLoS One. − 2012. − **7**, № 6. − e38323.

30. *Yang Y., Zhan L., Zhang W. et al.* RNA secondary structure in mutually exclusive splicing // Nature Struct. Mol. Biol. − 2011. − **18**, № 2. − P. 159−168.