

## Auxiliary elements of mammalian pre-mRNAs polyadenylation signals

M. I. Zarudnaya, A. L. Potyahaylo, I. M. Kolomiets, D. M. Hovorun

Institute of Molecular Biology and Genetics, National Academy of Sciences of Ukraine  
150, Vul. Zabolotnoho, Kyiv, 03143, Ukraine

E-mail: dhovorun@imbg.org.ua

---

*Polyadenylation is one of the levels at which eukaryotic gene expression is regulated. The auxiliary elements of mammalian polyadenylation (poly(A)) signals are known to influence positively or negatively the efficiency of 3'-end processing. In this paper, we surveyed the literature on auxiliary elements of mammalian poly(A) signals. We also compiled the database of human poly(A) signals and searched it for auxiliary elements. This database contains 244 pre-mRNA sequences covering the cleavage region. Literature data and our database screen demonstrated that auxiliary elements, particularly, sequences binding U1 snRNP-specific U1 A protein, are widely presented in cellular pre-mRNAs. Besides, we analyzed our database for the availability of the polynucleotide sequences with a potential to form G-quadruplexes or i-motifs which we consider as possible auxiliary elements. Based on published experimental findings on formation of quadruplexes, the schematic structures of possible quadruplexes are proposed for several human pre-mRNAs. The structures of putative quadruplexes are also represented for G-rich auxiliary downstream element of SV40 L poly(A) signal.*

---

**Introduction.** Poly(A) tails of mammalian mRNAs are important for the regulation of mRNA stability, mRNA export and translation initiation [1–3]. Moreover, the polyadenylation process is essential for the regulation of gene expression [4]. For example, numerous pre-mRNAs contain several poly(A) signals which are located in the 3' untranslated regions (UTRs) of different exons or in the same region of a single exon. In the first case, the use of the alternative poly(A) sites results in synthesis of different proteins. In the second case, pre-mRNAs with different 3' UTRs are translated into the same protein which could be synthesized with different efficiency, when the elements regulating mRNA stability or translatability are present between tandemly arranged poly(A) signals. It is known that the polyadenylation process is tightly coupled with transcription termination and splicing, poly(A) signals being important for efficiency of both processes [1–3]. Thereby, the study of poly(A) signal organization which is not completely clear at present, has a great importance for better understanding of cell functioning.

© M. I. ZARUDNAYA, A. L. POTYAHAYLO, I. M. KOLOMIETS,  
D. M. HOVORUN, 2002

The polyadenylation reaction of mammalian pre-mRNAs proceeds in two stages: endonucleolytic cleavage of pre-mRNA and subsequent addition of poly(A) sequence to the newly formed 3'-end [1–3]. Poly(A) signal directing this reaction consists of two core elements which are located upstream and downstream of the cleavage site. Additionally, poly(A) signal can contain auxiliary elements. In our reviews [3, 5], we concluded that the core downstream element of poly(A) signal is not a degenerate U/GU-rich sequence, as it was assumed earlier [1, 2], but rather consists of different number of simple elements (5nt long as a minimum) located at different distances from each other. The shortest elements are four out of five base U-rich elements (UREs) and tracts consisting of two GU dimers and one U residue. To further examine the structure of mammalian poly(A) signals, we developed the database of human poly(A) signals. This database includes 244 DNA sequences in the cleavage site region which were randomly selected from GenBank at National Center for Biotechnology Information (NCBI). Our findings of primary structure of core elements of poly(A) signals

were reported in our review on the downstream elements [5].

In this paper, we discuss the upstream and downstream auxiliary elements of mammalian poly(A) signals with a special emphasis on possible role of RNA four-stranded structures in the polyadenylation process.

**Database of human poly(A) signals.** Our database of human poly(A) signals contains 244 DNA sequences corresponding to the pre-mRNA regions within 200 nt upstream of the cleavage site (or up to the 5'-end of the exon) and 200 nt downstream. To determine the downstream region, we compared the sequences of genomic DNAs and appropriate mRNAs also extracted from NCBI databases. We selected DNA sequences marked by Evidence Code C (confirmed gene model — model based on alignment of mRNA, or mRNAs plus ESTs, to the genomic sequence). Only mRNAs of status Reviewed with specified cleavage/polyadenylation site were examined. A half-part of our database is shown in Fig. 1, only 140 nt long sequences being listed. Core elements and some auxiliary elements of poly(A) signals known or supposed to participate in the polyadenylation process are underlined. The full database is available from the authors upon request.

The polyadenylation reaction is performed by a large set of proteins [1–3]. Among them there are the cleavage and polyadenylation specificity factor (CPSF) which binds to the upstream core element of poly(A) signal, and the cleavage stimulation factor (CstF) which interacts with the downstream core element. In addition, both proteins interact with each other. Previously, it was assumed that the upstream element of poly(A) signal is represented by the AAUAAA hexanucleotide in most animal pre-mRNAs [1, 2]. More recent studies have shown that the actual occurrence of AAUAAA is much lower, perhaps as low as 50–60 % [6–8]. We found that 69 % of pre-mRNAs in our database possess this element [5]. In this connection, we grouped the studied pre-mRNAs by the core upstream element type ([5] and Fig. 1 herein).

The pre-mRNAs having the canonical AAUAAA hexamer belong to the largest group I. The pre-mRNAs containing AAUAAA with single-base substitutions are divided into II–IV groups. Single-base substitutions in canonical hexamer were shown to significantly reduce the efficiency of pre-mRNA cleavage and polyadenylation with the only exception of the AUUAAA hexamer which directs the cleavage with the relatively high efficiency (66 % of the wild type) [9]. Group II is comprised of pre-mRNAs with this hexanucleotide.

Statistical analysis of mammalian DNA databases has revealed that the number of 3' EST containing AAUAAA with single-base substitutions correlates with their processing efficiencies *in vitro* [7]. The most effectively processed variants were found to be the natural functional elements [6]. These variants are described by the NNUANA consensus, where N is any nucleotide.

We allocated the pre-mRNAs with NNUANA elements in group III. The pre-mRNAs containing AAUAAA with other single-base substitutions make up a small group IV. These elements may be both functional and deleterious depending on the pre-mRNA source (refs. in [6]). The pre-mRNAs containing the canonical upstream element with substitutions of any two bases form group V. Little else is known about such elements in the literature.

The remaining pre-mRNAs we placed in group VI. Though these transcripts do not contain hexamers of I–V types in the vicinity of the cleavage site, most of them contain elements I–III (evident elements of poly(A) signal) elsewhere within the upstream region under examination. Some of these hexamers may be signals of cleavage at the distant site if they and the core downstream elements are brought together due to formation of the stem-loop structure between them [10]. In Fig. 1 we underlined the hexamers of I–III groups located not only in the vicinity of the cleavage site but everywhere, since some of such hexamers may be parts of alternative poly(A) signals. Actually, about 60 % of pre-mRNAs from our database contain more than one hexamer of I–III groups, and about 70 % of these pre-mRNAs have putative downstream elements located not far from the hexamers.

The schematic picture of a mammalian poly(A) signal is shown in Fig. 2. Using our database, we determined the distances between the cleavage site and the core upstream/downstream elements. The figure displays the most frequent distances, which are consistent with the data reported [9, 11]. The most frequent distance between two adjacent core elements of mammalian poly(A) signals in our database appeared to be in the range of 25–50 nt.

An analysis of the data published [3] and our database screening showed that poly(A) signals having only core elements may direct the polyadenylation reaction with varied efficiency. The efficiency may depend on the upstream element type, the number of downstream elements, primary structure of the region around core elements, secondary structure of core poly(A) signal, and other parameters. Also, many pre-mRNAs contain auxiliary elements which influence (positively or negatively) the core poly(A) signal efficiency [2, 3]. In the next section, we

Group I

ACTA2 TCATTCCTAGCCAAAGCTCTGACTCGTACCTATGTGTTTTTT**ATAAAA**CTGAAATAGGCCTACTGGTAA  
AIF1 ATATGGAAACAGAAAGCAA**ATTTGTAAG**CCAGAGTCAACAA**ATTAATAA**TTACCCOCTCCTCCAGATC  
ALDOB GTCTTAGCAGTGGTAGGTTGGGAAGGAGATAGCTGCAACCAAAAAAG**AAATAAA**TATTCTATA**AA**CCCTC  
ANXA8 ACTGGGTACCTGGAGATTCTGAAGTGCCTTTGCTGTGGTTTTCAAAAT**ATAAAA**GATTGTATTCAACTC  
ATP6V0B GTGTCCCCACCTCCACCCTCAAGCCATCTCTCTAGTGTTTGTG**AAATAAA**CTTGGTATTGCTGGGTC  
BACE2 GCCACTATCACA**TAGCTTACT**CTTGCTTAAAAT**ATAAAA**TCATGTTTTGATGAGAAAAACTATTCTA  
BAT3 CCTCATCAGGGGACCGTTTTCCCCCTCTTCCCTTCACAGTATTTAAG**AAATAAA**AGTCGGATTTTCTGGC  
BBP TAATCTTTTACAAAGGCAAGTTGCCACTTTGTCATTTTTGCTTTCTG**AAATAAA**AGTATAACTTATTCC  
CA7 GTCCACCAGCTCACAAAGCCACCTACATGACAGTCCATCCCTGTGAAT**ATAAAA**TAATGTATCC  
CDC10 TCATATATTGCATTTCTGTATTTTTGTTGTATTTGTA**AAAAA**TCCACAT**ATAAAA**CGGATTTGTGATG  
\*CDH16 CTCCAAATGTCAGGGTGT**TTGCCAA**AT**ATAAAA**GCCCCAGAGAACTGGGCTGGGCCCTATGGGATTTGTA  
CLIC4 TATCTTTGGAACATCAGCACCAGTATATTGGCTGGCAGCTATTGT**ATAAAA**AGTATATTTTCACT  
CST11 TGTGAGAGAA**TCC**TGCAATGGGCAATA**TCC**T**ATAATA**GGCA**ATAAA**TGTTTTGCAACATATTTCTTA  
CYP21A2 CTCGGCTGCAGAGGAT**T**GAGCTTAATTTCTGAGCTGGCCCTTTCCAGCC**ATAAA**TCAACTCCAGCTCCC  
DDX16 CTATACATTA**TTAA**TACC**TATTAAATAAA**AT**TATTTTT**GG**AAATAAA**GCTTGTGGGAACA**T**TTGGGATCT  
DEEFEST C**CC**CAAT**AC**CAAG**CC**AACTGGCATAGAGCCAACTGAG**ATAAA**TGCTATTT**ATAAA**AGTGATTTAATG  
DNAJA2 **TTTTTTTTTTTTTT**TACCATCTTTGTAAGG**TTT**CTGAAACTCGAT**ATAAAA**AGCGGGTTGGGTAAAT  
DPM1 TCTCAAT**TTT**CG**TTTT**CA**TTTT**CTGTATTGAGACCT**ATAATAAA**TGTATATTTTT**TTT**GCATA**AGT**  
EFEMP1 AAATGTCAC**TGCTTT**TACTTTGATGTATCATATTT**ATAATAAA**ATAATA**TTCCTTT**AGAAAGATCACTC  
ENO1 TCCGGGTGGCCACAGGCTAGATCCCGGTGGTTTTG**TGC**T**AAATAAA**AGCCCTCAGTGACCCATGAG  
F13A1 CTGGCTTAAC**T**TAGTATATTTATTAATTACAATG**TAATAGA**AGCTT**AAATAAA**GTAAACTGATTTAT  
FANCC TTTT**TAGCTC**TTT**AAAA**CTGTGAATTTGTTGCTATTTTT**TT**CTATCT**ATAATAAA**ACTTCAGTTTGCACT  
FBXL3A **TTAA**TT**TGTT**TATACTGTAAGATATATGTTAAATGCTTTTATCACTTTGAG**ATAAA**AGTTACTAA**TGC**  
FBXL7 CAAATCACTCGCTGCTGTATGAATCTAGAAAGCCTTAATTTACTACCAAG**ATAAA**AGCAATATGTTCTG  
FEZ2 TCATCAATATCATGAGCACTTTGGCTGGTCCAGTCAAATCAATGGCATTT**ATAAAA**TTTTTAA**AGT**  
FHOD1 TGCATATGCATGTCATGTGTGTGAGCTCCTTGAACGCACGGAGCA**ATAAAA**TTTTCTTAGCTAA**TCC**  
FLJ23277 GGT**TTT**GTGCTTATTTCAAAGTGGGAAATGTACCTGGCACGTT**GAAATA**AAAAAATCTGACTACCTATC  
FXR2 GGTGGATGAGAAGCGGTATTTATTTTCACTGTACAGTATTTAA**AAAG**AG**ATAAA**AAAA**TCCA**ATGGC  
GAPD TTGAAGAGGGGAGGGGCTAGGGAGCGCACCTTTGTCATGTACCATC**ATAAA**GTACCCTGTGCTCAACC  
GCNT3 AGGAGA**ACT**TTGATGGAAAGAAAGCTTCCCTTCTGTACTGTAACTT**AAATAAA**TAGCTCC**TGATC**  
GFER C**CC**CCACCGCCTTCCCTTTTCCCTGTCTTCCCTTAAAGTTTCACTCCCT**GAATAAA**ACTTCACTTTGCC**TT**  
GJA4 TGGTGCCTTGTTTTCACTACCTCTCTCTAGTTCTACTGTTCAAGCTTCTG**AAATAAA**AGGACTTGATCA  
GNE TTTAATGTAATATTTTTGTTACTGTTCTGAGCCAAATTTCAAAGAAA**ATAAA**TACATTTCC**TGTTG**  
GFS2 CACCCCCACCCCACTG**TGG**GTGAGGGTATCCCTG**TGTG**CCAGGCC**ATAAA**ACTTACCTGGCACT  
GEMK T**CTTTTT**TCACTTGTATCACTGATGATTTCTTACCA**TGCTG**TTTT**TAATCA**ATA**AA**TAATTTCTGAGAGACT  
HBA2 TCCCAACGGGCCCTCCTCCCTCTCTG**AC**CGGCCCTTCC**TGG**CTTT**G****ATAAA**AGTCTGAGTGGGCAGC  
HBD GAAC**TTGG**GAACACA**TGCC**TACTTCAAGGGTATGGCTTCTGCCT**ATAAA**GAATGTTCAAGCTCAACTTC  
HBG1 CAGAGCTT**TC**AAAGATAGGC**TT**TATTTCTGCAAGC**ATAACA**AT**ATAAA**CTATTTCTGCTGAGAGATCAC  
HDAC7A ATG**TAT**GCATGTT**TT**TGTA**AA**AGGAAGAAATGGAAAAA**ATCT**GAA**CAATAAA**TG**TTTT**ATT**TGCTTT**  
HLA-A GGT**TGAT**GAGTT**AA**T**AA**AGGAGAAAT**TCC**TAA**ATTT**GAGAGCA**ATAAA**TGGAAAGACATGAGAA**C**  
HMOX1 AG**CCCC**AGCTTATCTCTCTCTCCGGCTGTGTAATGCTCCAGCACTC**ATAAA**AGTGGGCTTTGCAAGCT  
HNRPE **TTTT**TTTGGATG**TG**ATTG**TAT**TTCCAA**TTT**CTTGTCTCA**TGA**ATTT**CAATAAA**ACT**AAAA**ATCTAT**TC**  
ICAM4 **ATAAA**AT**AAAA**AT**AAAA**AT**AAAA**AT**TTGG**GGGGGAACCTCTGGAATC**ATAAA**AGGTTCTTCA**CC**AGC  
KRFTHA1 C**AT**CCAGAA**ACT**CTCTGATGTG**AG**TGGGCTCCCTTTAA**CTT**CA**CTT**TA**AT**CT**CT**CT**TTGG**  
\*KRFTH5 C**AG**TAGAGGGCTTTTGT**TTGG**AGAA**CAATAAA**AGTGCATTTGCT**AG**GGCCCTGATGCC**TA**CTTGCACCA  
LGALS4 C**T**ATG**GG**AAAAACAGAA**TTA**T**CC**CTTAGGACTCCTTTCTAA**GGCC**CT**ATAAA**ATG**CT**TSAGGGGTG**CTC**



CYP4E2	GTCCAGGTTGTCATATAATAATATGCTGTGAGCATCTTCCATGAC <u>ATTAAA</u> TCATCTTAGGAAACATT	<u>ATTTTGT</u> GGTCTCCTCAAAA <u>TGTGTGTGTT</u> AAAGTTCAAATCAGTCTTAATTTAAAAAAGACATCATATT
F10	GGCTTGCCCAAGGCCAAGAGCC <u>TGCCCGG</u> AGGTCATAACGTCCTCC <u>ATTAAA</u> GTGAGATCCCCTC	AAGGCTGGT <u>TTTGT</u> CTCTCGATTGCCGCTTGCCTGGCTTCTCCGCGCCTGTTGAGGTGGGAAGGTGAA
FL11	CCGCTAGGG <u>GATAAA</u> AGCAGATGTGGATGCCCTTTAAGAGAT <u>ATTAAA</u> TGCTTTTATTTCATAATATAAAAA	TCAG <u>PATTTTT</u> TAATATAAAAAATGGCGCTAA <u>TTTTAA</u> CAAAAACGACAGAAAAAACCCCAAGGTACAATCT
MAK	AAAGCTGTGCAGTCTTTTGTACTGGGACACTTTTAAACTCTGAATAGGC <u>ATTAAAA</u> AAAAATATGGGTAA	CTAG <u>TCCTTTT</u> TCATTA <u>GGTTTTAA</u> AAACCTTATATCTGAATTC <u>TATTTGTGTGTGTGT</u> TAATTAGT
PMM2	TTCCTTGTGCAATGACTGTAAATGACCTTCCAAAGTGAAGAGTAGCAC <u>ATTAAA</u> AGTGATTTTATTGTTT	<u>CCTGTTT</u> CTCATTTCCGAT <u>TTCTTTTT</u> CTCTCTATGTTCCAGATTCTGGAATGGAACCCAGCCCTTCCA
CAGACACTGCTCTAGTACCATTCCCTCTCTAGCCCCAGGAGCA <u>ATTAAA</u> AGGTACAGTTAAAACTCT	<u>AACTTTCT</u> CTCT <u>TGTGT</u> GAGTCAATGGGGGTGGGGCAGGTCTCAGTAATACCCCTTCCGCCACACTTCA	
ROB01	TGCATCTGAAAAATCCAATAATCAGAAAAAT <u>ATTTTTGTG</u> CACATTATTT <u>ATTAAA</u> AAATGTTCTCAAAATAC	<u>ATCTTCTTT</u> CT <u>TGTGT</u> GTGATTTCTTTTTCAT <u>TGTGT</u> GCTGATCAAA <u>TTTTAT</u> TCCATTTAATCACAAAG
SRD5A2	TACTGAAAAATGCAAAATGCAATCATA <u>ATAAA</u> TAAGTGTTTTGTGTTCAT <u>ATAAA</u> TACCTTTAAATCATG	GATGTTAAGCAG <u>TTTTGT</u> TTGAAT <u>ATAAA</u> TATCCTGGGACTTCTGGAG <u>TTTTTC</u> CACAAGGTATTGGGGTG
TPSB1	<u>CTGCCC</u> CAGATAGCTGGTGGTGGGGCTAATCTCTGAGTCTGGACCTC <u>ATTAAA</u> GTGCATGGAAATC	ACTGCTGTGCATCGCT <u>TGTGTCTCGTGT</u> GGAATGTCACCTGGGAGAGAAGGGTCCAGGTGTCTGAGGA
WNT5B	CGGGCCCGCGTGTCTCATCTCTGCCCCAGGTGTACGGTTCTCTCTGCAC <u>ATTAAA</u> TGCCCTTCATGG	AGG <u>TTTTG</u> CTCCCTTCTCTTATTTGGACCCACGCTGAT <u>CTTT</u> TCATGAGTCTCCTTTATTTTTTATTTGG
ZNF278	GAGCAAGCCCAAGTTGACCTTGATGTGAATTGATCTGATCAGACTGT <u>ATTAAAAA</u> TGTTAGTACATTA	CTCTATCTGCTCCCTTGAGTCCAG <u>TTTTTC</u> AGCTGTGGGGG <u>CTTT</u> GGGACCCAGAGGTGGGCAGTCCAT
Group III		
ACRBP	GTCCCAGAGAGGGCCACGGTGGGAGCTGCGCCCTCCTTAAAAGATGACTTTA <u>CATAAA</u> ATGTTGATCTTC	AGCC <u>TGTGT</u> GGCCCTGTTCTGGTTATGATCCCAATCCTCAGCACTGCTACCCTGCCCGTGGCAGTCAGTC
ALDH3A2	CAGATATACCCCTATTGGAGACAATCCTTTGAT <u>CATAAA</u> TTCGCCCAACT <u>ATAAA</u> TCATTTTATGTCCTT	ATCATAGATGG <u>TATTTG</u> TCAAC <u>TTTTTTTTTCTC</u> ATT <u>TTAAA</u> ATTC <u>TT</u> AGCTGAGAAG <u>TTTTTC</u> CAAAAC
B3GALT5	GACTTTACTTTGTACTCAGAGAAGAGCCCTCACATGGCTGTGTCCAT <u>TATAAA</u> TGTTGG <u>ACTTAA</u> CTCTT	ACTTTGAACCTCAGGAAAAGTTACTTGCTGCAAAAATGG <u>TTTTG</u> ACCATGTTCAATTGCAAAAAGTACGTCGA
BAG4	TAGACTTTAGCCCTCTGGACTTCTGTTTGTGTTTGTATTTGCAGTTTACA <u>ATAA</u> TAGATTATTTCTCT	ACTTCTTGGTAC <u>TTTTGT</u> AGTAATA <u>TGTTTAA</u> TATAT <u>TTTAA</u> TTCAGAGACTATATTGACAGCCAGATAG
BRPF1	AGATTCCCCACCTTTCAAACCTGGTTTGTATTATTTCAAAGGAAGAA <u>AAATATA</u> TTGATCTTGTAGAAAT	<u>AAACT</u> GTCAATTTAGAA <u>TTCTTTG</u> TTCTTCCGCCCTGCCAGGGCAAAGGGATACAAAAGGGAGTTCTGT
CST	ATGTTCTATTCTGCCCTCCCTTAAAGGGGAGACCTCAGAA <u>ACTAAA</u> GGAAATTTGATGTTGTGTTTTGTT	AATCAGCCCTCAGTGGTGGTCACTTGGGGTGGGGT <u>TTTTGGGG</u> GAACCCAGAGACCACAGTGGTGC
FECH	CCATTTAATGCCATTGTCACCTTGGATTTATGAGTGAAGTGTTC <u>TAAAA</u> ATATGAAATATGTCAG	ATCAGAGTCTGATCTTCTAT <u>TGTTG</u> TATTAATGGAT <u>TTAAA</u> AGATCCCGGTGGTTCCATGAAGA <u>TTTT</u>
HBQ1	ACTGTGGGTGGGTGGCCGGGATCCCCAGGC <u>CACT</u> TCCCGTGTGTTGAGTAAAGCCCTTCCAGGAGC	AGCC <u>TTCT</u> TGCCGTGCTCTCTCAGGCTCAGGACGCGAGAGGAAGGCCGCCCTCCCAAGGAAAGGGC
MAD4	CACCTGTCCACCGTGTGGGCCGTGTGTCTTATGTCATTTGATATAAATACAGATTTTATATCTC	ACCCTGCTGGCTCCGGCT <u>CTTTCT</u> GTGTG <u>TTGT</u> CACTGTCAGCGTCTTCCGCCCTGCCCTGGCCCTCCA
MSMB	CTGGCCCTAATAGTCAATGATTTGTGTAGCCATGCCTATC <u>AGTAAA</u> AGATTTTGTAGCAAACTTGA	TAT <u>TGTGT</u> CT <u>TTTTA</u> TTTTATAATTTATGTAATGCAATGATTTAATACACATTTATAAGAAATACAAAACA
*RORE	<u>AGGGGG</u> AGGCATATT <u>TATAAA</u> ATGCTAT <u>AAATATAAA</u> ATGTAGCAAAACTTGACAGACTAGAAAAAACA	GATC <u>TGTGT</u> TATCTAGGAACTAATGTACCCCAAGCCAAAACATAATTCCTGTGAAGTTTACAGTTACA
Group IV		
DUSP4	CATCTTGAGAAGCA <u>GCCGG</u> TTGGGTGGGAGGAGG <u>AGAAA</u> GGGAAGAATTAGGTTGAAATGCTTTTTT	AAAAAAAAGAAAAGAAAAAAGACAGCATCTCACTATGTTGCCAAGGCTCATCTCAAGCTCTTGGGC
GRAP2	CAGGGGTGPTGGAAAGCAGTGGGGAGTTGGGAGGGGGCAGGGAATGAAATGGAGTTTGTCTCTGGC	CTTCAGCTGTCACTGCTTCTCTCTGTCTTGGGA <u>TTTTTC</u> ACGGGAGAACAGCTAAGCAGAGACCACACCT
TRIAD3	AGAGGTTGCAGTGAGCCAAGATTGTGCCACTGCACCTCCAGCCTGAGC <u>AAACAAA</u> AGCAAAACTCTATCTT	AAAAAAAAGAAAAAAGAAAAAAGTAAAGATACCCAT <u>TTCTTTCT</u> GCCTTTGCCAGGATGGAGGCTTTGAAGACA
Group V		
CDKN3	AATGAAACCACCAGTGTATCAACTTGAATGTAATGTACATGTGCAGATATT <u>CTTAAA</u> GTTTTATTGAC	AAAACCTC <u>TTTGTCTCTTT</u> CA <u>TTTTCT</u> TATAGGAATAGGAAAAT <u>TTGG</u> ATAGGAAAAT <u>TG</u> TATGCA <u>ATT</u>
FLRT1	CATTTCCCTTGCAGATTTGCAGAAACATGGCATCTTTCACGTGATCTTTG <u>AAACATGAT</u> GTAGTCGAT	AAAAAAAAAAAAAACA <u>ACTTTTTTTTCT</u> CATAGGCTGAAGCCCTTCTCAGTTCATGCACCACGCTCCGTAGA
CNNT1	AATACTGTACTGTGTAGTGTCTCCGTATGTCATCTCAGGGAGC <u>TTAAA</u> ATGGGCTTGATTTAACAATTG	<u>TTTTTGTGT</u> AT <u>TTTTG</u> CTTGAACAACGCACAC <u>TTTTTGA</u> CAACCAAAAAATGACAAT <u>TTCTA</u> GTTTA
NDN	<u>AAATATA</u> GAAAAATATACAAAAACAACAGTAAGATTGCTGT <u>TAATCACAT</u> CAATATGGGA <u>ATAAAAA</u> ACA	AAAAATAATTTCTCCCTTAAG <u>TTTTCT</u> CACATTTTATCAAAAATTAATAGATGCTCTTGACATCTATTAAT
Group VI		
ACY1	ACCTCCTCTTCCCTTCCAAAT <u>ATAAAA</u> GTCTATGGACAGGGCTGTCTCTGAAGTACTAACACAAGGAC	ACTCGTGGAGCAAGA <u>TTTTTC</u> TTTTCTCTGGGGACATGTTACCATCTCCATTTCAAGATGAGGAAACTG
GPC6	CATTTTCAACAAAAAGCAACAGAGAAAA <u>ATAAAA</u> TGAACCTTAAACACTGTAAGTTCAGCATGACAGCC	AACCTCCAGTGAAGC <u>TTTTTGT</u> TGACAAAATCCGGTTTAAAAATGCTTATGGGGAGAAGCTTAAAG <u>TT</u>
PTERO	ATATACAACAGAAACACTTTGAGTTGAA <u>AGTAAA</u> CACAAGCTGGCTGCTTCCCTGTGGCAACGTGGCT	ATTGTGAGCCTCCG <u>TGTGTGT</u> GCTTCCAGGAGGTTTGTCTCCTTCAATGCACCCGACACTGCCCTTCC
WIRE	AGCTTAAAAATTTAATCACAGGATTTGCCACTGCACCTCCAGCCTGGGGCAGAGAGCCAGACTCCGCTC	AAAAAAAAGAAAAAAGAAAAAATTTAATCACACACATCACAGGCTTAAACTGTCTTAGAGACTTTCCAG

Fig. 1. Database of human poly(A) signals. The pre-mRNAs (DNAs) distribution into groups is explained in the text. Only every second sequence from every group is shown here. The sequences marked by asterisk are included into the groups on condition that the distance between the appropriate hexamer and the core downstream element does not exceed 54 nt. The sequence before blank space is corresponded to the mRNA, the sequence after blank space is corresponded to the 3' fragment of cleaved pre-mRNA. If the cleavage site was not indicated in mRNA sequences from NCBI database, we took the last non-A nucleotide before the poly(A) tail as 3' terminal. If the several cleavage sites were indicated, the blank space was positioned after the last site. Hexamers I-III are underlined throughout the Figure (they are printed in bold in -10/-70 region). Hexamers IV and V are underlined only within groups IV and V in -10/-35 region. UREs and 2GU/U elements are underlined and printed in bold throughout the downstream region. The following auxiliary elements are underlined: binding sites of U1 A (N- and C-terminal RBDs, the latter is also printed in italic), polypyrimidine and oligo(U) tracts. Also, the G- or C-repeats in the sequences with a potential to form G-quadruplexes or i-motifs are underlined. The binding sites of SRp20 are printed in bold and not underlined.

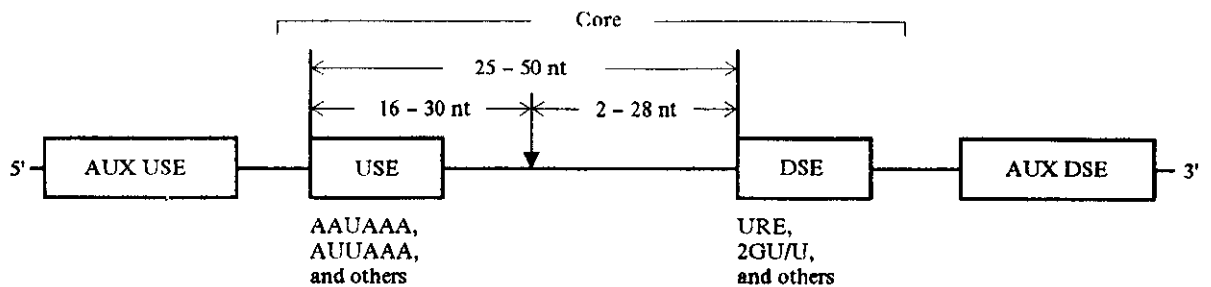


Fig. 2. The alignment of mammalian poly(A) signal elements. The cleavage site is indicated by the black arrow

discuss the published data on these elements and analyze their occurrence in our database.

**Auxiliary elements of mammalian poly(A) signals.** We summarized the available information on the auxiliary elements in Tables 1 and 2, showing the sequences of the elements and the proteins binding specifically to them. Spliceosomal proteins U1 A and U1 70K (Table 1, lines 1–3) belong to the group of proteins with poly(A) polymerase (PAP) regulatory domains (PRDs) [12–14]. PAP fulfils the second stage of the polyadenylation reaction, addition of poly(A) tail, and in the most cases also participates in the first stage, cleavage of a transcript [1–3]. Both U1 A and U1 70K (as a part of U1 snRNP) specifically bind to the substrate RNA and inhibit poly(A) addition via their PRDs. This inhibitory mechanism is used in U1 A autoregulation [12, 15], in control of expression of immunoglobulin M secretory mRNA [16], and in repression of bovine papilloma virus (BPV) late gene expression [13]. In the case of U1 A, two molecules of protein are required to inhibit PAP [15]. U1 A directly binds to pre-mRNAs via the N terminal RNA binding domain (RBD), its binding to U1 A pre-mRNA being stronger than to IgM transcript. Phillips et al. [16] found U1 A-binding motifs in pre-mRNAs of other immunoglobulin isotypes and suggested that the expression of these mRNAs might be regulated similarly to the expression of IgM mRNA. We discovered that about 9 % of pre-mRNAs available in our database contain two or more putative binding sites for the N terminal RBD of U1 A. In Fig. 1 these sites presented by the AUUGC/UAC [15] or AU/GGCN<sub>2,3</sub>C [16] sequences are underlined.

U1 70K inhibits PAP when bound to U1 snRNA. In this case, the auxiliary upstream element (AUX USE) which participates in regulation of the polyadenylation reaction represents the pseudo 5' splicing site (5' ss) to which U1 snRNP binds (Table 1, line 3). Screening our database, we have not found pre-mRNAs with consensus 5' ss (C/AAGGUA/GA-

GU) located in the upstream region. However, we have not searched for 5' ss homologs, which can also act as inhibitory elements ([17] and refs. therein).

U1 A can influence the polyadenylation reaction not only negatively. Its interaction with PAP results in inhibition of poly(A) addition. In contrast, U1 A interaction with the other basal factor of polyadenylation machinery (CPSF) leads to 3'-end processing stimulation by stabilization of CPSF binding to substrate RNA [18]. While U1 A interacts both with its own pre-mRNA and IgM pre-mRNA via the N-terminal RBD, it binds to the SV40 L pre-mRNA (Table 1, line 4) via the C-terminal RBD [19]. Preliminary results obtained by Lutz et al. [18] show that U1 A enhances CPSF-dependent polyadenylation of the precleaved SV40 L pre-mRNA *in vitro* not binding to RNA substrate. Nevertheless, the authors believe that this binding is important for *in vivo* reaction. In our opinion, the interaction of U1 A with RNA can at least increase the local concentration both of U1 A and CPSF (bound to U1 A) in the vicinity of the core poly(A) signal, thereby facilitating 3'-end processing reaction.

The motif, similar to U1 A-binding elements in SV40 L pre-mRNA, was also found in the pre-mRNA of ground squirrel hepatitis B virus (GSHV) in the region required for increasing the efficiency of the polyadenylation reaction directed by noncanonical UAUAAA hexamer [20]. It should be noted that multiple upstream auxiliary elements are present in GSHV pre-mRNA, but the mechanisms of their functioning remain unclear [20, 21]. When the present article has been ready for publication, Natalizio et al. reported that elements resembling AUX USEs in SV40 L pre-mRNA can also be found in many cellular pre-mRNAs [22]. The authors showed that these elements stimulate 3'-end processing of three human collagen pre-mRNAs. One of them is shown in Table 1 (line 5). In our database search, we found that 8 pre-mRNAs have putative binding sites for C-terminal RBD of U1 A described by UAU<sub>2-5</sub>GUNA

**Table 1**  
Upstream elements which influence the efficiency of core poly(A) signals

N	Pre-mRNA	Element(s)	Distance from the hexamer, nt	Protein(s) or RNA which specifically interact with the element	Comments
1	U1 A	AUUGUAC AUUGCAC	52 26	U1 A (N-terminal RBD)	Two molecules of U1 A inhibit PAP
2	IgM heavy chain secretory	AGGCGGC AUGCUAGC AUGCAAAC	119 98 50	U1 A (N-terminal RBD)	Mechanism of inhibition is supposed to be the same as in case of U1 A pre-mRNA
3	Bovine papilloma virus (BPV-1)	AAGGUAAGU (pseudo 5' splicing site)	20	U1 snRNA	U1 70 K as a part of U1 snRNP inhibits PAP
4	Simian virus 40 late (SV40 L)	AUUUGUGAA AUUUGUGAU AUUUGUAAC	51 42 22	U1 A (C-terminal RBD)	U1 A stimulates polyadenylation by interaction with CPSF-160
5	Human collagen type I (COL1A2)	AUUGUACC AUUUUGUAU	32 23	Possibly U1 A (C-terminal RBD)	Mechanism of inhibition is supposed to be the same as in case of SV40 L pre-mRNA
6	Human immunodeficiency virus (HIV-1) (3' LTR)	CAGCUGCUUUUUGCC- UGU	94	CPSF	The element stabilizes CPSF binding to pre-mRNA providing efficient 3'-end processing
7	Equine infectious anemia virus (EIAV)	58 nt long U3 region which contains 42 % U residues	114	CPSF	The element stabilizes CPSF binding to pre-mRNA providing efficient 3'-end processing
8	Human lamin B2	89 nt long sequence containing several oligo(U) tracts	100	CPSF 55 kDa protein	The element stabilizes CPSF binding to pre-mRNA mediating the efficient 3'-end formation; 55 kDa protein interferes with CPSF binding and inhibits 3'-end processing
9	Adenovirus L3	UUCUUUUU	52	Possibly CPSF	The element stabilizes the formation of the cleavage complex
10	C2 complement	53 nt long sequence containing the AUGCUUGUUUCACUU- UCAC fragment	53	PTB CstF	PTB stimulates cleavage; CstF stimulates cleavage and polyadenylation
11	Histone H2a	101 nt long sequence from the coding region containing the ACAACAAGAAGACGCGC- AUCAU fragment	—	SRp20 9G8	Mechanism of 3'-end processing stimulation is unknown
12	2'-5' oligo A synthetase enzyme	UAAUGUCUAAUGUA- UUAUCAUAUAC	31	Unknown	Mechanism of 3'-end processing stimulation is unknown; U residues are crucial for the element activity
13	Human papilloma virus (HPV-16)	79 nt long element containing the fragment GAAUUAGUGU- UGUUUGUUGUGUAUAUG- UUUGUAUGU	132	CstF U2AF 65 HUR	Mechanism of 3'-end processing stimulation is unknown

Continuation of *tabl. 1*

N	Pre-mRNA	Element(s)	Distance from the hexamer, nt	Protein(s) or RNA which specifically interact with the element	Comments
14	Xenopus albumin	CAAACUCACUGAGGAAC- ACCU AAAAGUCCUUCAGCUG- AAAAGAGCAU	99 55	~62 kDa	Perhaps, the later phase of poly(A) addition is blocked
15	Transferrin	AGUUUAUUCUCAGAUGU- GGGAGG	—	~62 kDa	Mechanism of poly(A) length limiting in the same as in case of albumin pre-mRNA. The element is located relatively close to AAUAAA
16	HIV-EP2/Schnurri 2	AGAUCUUCAGAGAAA- AGAG	1649	~62 kDa	Mechanism of poly(A) length limiting is the same as in case of albumin pre-mRNA.
17	Pre-mRNA reporters with SV40 L poly(A) signal	Possibly 3' most exon-exon junction	—	SRm160	Interaction of SRm160 with CPSF promotes cleavage

**Table 2**  
Downstream elements which influence the efficiency of core poly(A) signals

N	Pre-mRNA	Element	Distance from the cleavage site, nt	Protein(s) or RNA which specifically interact with the element	Comments
1	SV40 L	GGGGGAGGUGUGGG	32	hnRNP H/H'	Mechanism of 3'-end processing stimulation is unknown
2	HIV-1 (5' LTR)	CUGGUGAGUAC (major 5' splicing site)	195	U1 snRNP	Use of weak poly(A) signal is inhibited; mechanism of inhibition is unknown
3	Calcitonin (CT)	127 nt enhancer containing CU- CCGCUCEUCUCCAGGUA- AGAC core sequence	168	U1 snRNA; SRp20; PTB; ASF/SF2 (all proteins bind to the core sequence)	Mechanism of 3'-end processing stimulation is unknown

consensus [22], these sites are located within 70 nt upstream of the cleavage site. In fact, ten U1 A binding sites of collagen and SV40 L pre-mRNAs examined [19, 22] may be described by U/AU<sub>2-4</sub>G-U/ANA/U consensus. As many as 18 pre-mRNAs available in our database were found to contain these elements located close to the cleavage site, while other 24 transcripts contain such elements in the remote

upstream region. Taking into account the frequency of binding sites both for N and C terminal RBDs of U1 A in the poly(A) signal region of pre-mRNAs (the literature data analysis and our database screening), it is possible to suggest that regulation of polyadenylation reaction via U1 A protein is rather frequent event.

Table 1 (lines 6—9) shows those pre-mRNAs



which use the same mechanism to enhance polyadenylation [23–28]. The stimulation observed results from stabilizing the CPSF binding to substrate pre-mRNA without participation of auxiliary proteins like U1 A. It occurs via direct binding of CPSF to U-rich auxiliary upstream elements [23, 24]. Evidently, CPSF has two RNA binding sites, but it is still unknown if CPSF binds to AAUAAA and U-rich sequence simultaneously, or the U-rich element acts only as an entry site for CPSF which subsequently slides along RNA towards core poly(A) signal [25]. Interestingly, CPSF is UV cross-linked to the U-rich element only in the context of AAUAAA [23, 24]. Consensus for U-rich element to which CPSF binds is presently undetermined. Possibly, oligo(U) tracts are the special feature of these elements. Their occurrence is rather high in pre-mRNAs of our database: approximately 13 % of the transcripts studied contain several oligo(U) tracts in the vicinity of the upstream core element of poly(A) signal.

The auxiliary upstream element found in C2 complement pre-mRNA includes the fragment which serves as a binding site for heterogeneous nuclear protein hnRNP I (Table 1, line 10), also called the polypyrimidine-tract (PPT) binding protein (PTB) [29]. PTB binding to C2 pre-mRNA AUX USE enhances the cleavage reaction, the mechanism remaining unknown. This auxiliary element is also required for activating poly(A) addition via the general polyadenylation factor CstF [29]. It is the first and yet a single demonstration that CstF enhances the second stage of polyadenylation reaction and exerts its effect by binding to AUX USE. In our database, 9 pre-mRNAs were found to contain PPTs with consensus YYYUUCUUY (refs. in [30]) located within 70 nt upstream of the cleavage site.

Huang et al. showed that the transport elements from three naturally intronless pre-mRNAs (histone H2a, herpes simplex virus thymidine kinase (HSV Tk), and hepatitis B virus (HBV)) in addition to promoting mRNA nuclear export act as polyadenylation enhancers and as splicing inhibitors [31]. The activatory elements of these three pre-mRNAs are different, the mechanisms by which they stimulate the polyadenylation reaction are still unknown. Though some proteins binding to the transport elements were found, their role in polyadenylation has not yet been clarified. In Table 1, we therefore demonstrate only one of these elements (from histone H2a pre-mRNA, line 11). The 22 nt long fragment of this element was shown to be required for all three activities [32]. Two members of the serine/arginine-rich (SR) protein family (SRp20 and 9G8) specifically bind to this fragment. It is also

worth noting that PTB specifically interacts with a fragment of the transport element from HBV [30]. PTB [29] and SRp20 (see text below) are known to participate in the polyadenylation process. Screening our database, we found that 5 pre-mRNAs have the binding site for SRp20 (C/UA/UCUUCAU [32]) in the upstream region.

Several pre-mRNAs are known to have the auxiliary upstream elements, though little information exists about them. Among them are adenovirus L1, Epstein-Barr virus DNA polymerase (EBV pol) and 2'-5' oligo A synthetase enzyme (OASE) transcripts. In L1 pre-mRNA, the unidentified auxiliary elements, located between 50–113 nt upstream and 52–170 nt downstream of the cleavage site, enhance the stability of CPSF binding to substrate pre-mRNA [33]. CPSF is likely to directly interact with the Ad L1 AUX USE. In EBV pol pre-mRNA auxiliary element (UUUGUA) is located 8 nt upstream of the noncanonical UA-UAAA hexamer, besides the elements located farther on may also function as AUX USEs [34]. The authors suppose that EBV SM early protein may participate in enhancement of 3'-end processing reaction [35]. The auxiliary element identified in OASE pre-mRNA [36] is presented in Table 1 (line 12), the mechanism of its functioning being unknown.

Several papillomaviruses, besides BPV (Table 1, line 3), contain inhibitory elements controlling production of the late RNAs ([37] and refs. therein). The negative regulatory element (NRE) of human papillomavirus type 16 (HPV-16) appears to regulate polyadenylation (unpublished data of McGuire and Graham reported in [37]), nuclear export and mRNA stability (refs. in [37]). NRE was shown to specifically interact with three proteins: U2AF 65, CstF and Elav-like HUR [38]. The fragment of NRE required for U2AF 65 binding [17] is given in Table 1 (line 13). It contains GU/U-rich tracts which are also potential CstF-binding sites. However, the whole NRE (79 nt long sequence) is required for CstF binding. All these proteins (U2AF 65, CstF and HUR) bind to complex negative regulatory element of HPV-31 [37]. Cumming et al. [37] think that CstF-binding sites of HPV-31 element could compete with the downstream core element of poly(A) signal for CstF binding, reducing thereby the rate of polyadenylation reaction. They also suppose that U2 AF 65, which has PAP regulatory domain homologous to those in U1 A and U1 70K [14, 39], inhibits poly(A) addition. HUR may act by stabilizing late mRNAs in the cytoplasm. AUX USEs listed in lines 14–16 of Table 1 exert an unusual effect on polyadenylation. They are responsible for occurrence of pre-mRNAs with short discrete < 20 nt long poly(A) tails [40,

41]. Spacing between these USEs called poly(A)-limiting elements (PLE) and poly(A) signal plays no detectable functional role and may be more than 1000 nt long (Table 1, line 16). But the presence of these elements in the terminal exon is essential for their functioning [42]. PLEs are likely to inhibit the later phase of poly(A) synthesis — addition of poly(A) tail to short oligo(A) tract which is probably formed by the cleavage complex [43]. In database search Das Gupta et al. [42] identified several hundred genes with PLE-like elements and showed that the mRNAs of those genes that were examined by them have < 20 nt poly(A) tails. They also identified the ~62 kDa protein which specifically binds to PLEs. Screening our database, we failed to find pre-mRNAs with PLE elements (AGUCCUUYRGCURNRRRR [42]).

The splicing-associated factor SRm160 (SR-related nuclear matrix protein) is given in Table 1 (bottom line) in view of the coupling of polyadenylation and splicing. Vagner et al. consider poly(A) signals to act as 3'-terminal exon splicing enhancers in addition to their role in the 3'-end processing [39]. They suppose that PAP, as a part of the cleavage complex, can interact with U2AF 65 and help to tether this factor to the pyrimidine tract of the adjacent 3'-splice site, thereby stimulating splicing. Correspondingly, the splicing signals can be considered to act as polyadenylation enhancers. The mechanism by which splicing can stimulate polyadenylation is clarified in recent work of McCracken et al. [44], who found that SRm160 specifically binds to general polyadenylation factor CPSF and promotes the cleavage of pre-mRNAs. SRm160 forms a part of a splicing-dependent complex deposited by spliceosome 20–24 nt upstream of exon-exon junctions [45]. The enhancement of polyadenylation via SRm160 is not pre-mRNA specific and probably takes place in case of different splicing substrate.

The last upstream auxiliary element under discussion here is a part of poly(A) signal of the adenovirus-2 (Ad 2) L4 pre-mRNA. Sittler et al. [46] believe that more effective part of this positively acting element (USEa) plays a structural role. It forms the ascending arm of the hairpin structure with AAUAAA in the 14-nt loop, where the hexamer is well exposed to cleavage and polyadenylation factors. However the authors do not formally exclude the fact that USEa (UCUCUGUGCUGA) is also a recognition site for a regulatory protein, possibly as a part of the helix.

Downstream auxiliary elements known so far are not so numerous as upstream ones. Some of them are shown in Table 2. AUX DSE of the SV40 L pre-mRNA is represented by G-rich region (Table 2, line

1) which specifically interacts with hnRNP H/H' [47, 48]. The mechanism of 3'-end processing stimulation is undetermined to date. It was recently shown that many pre-mRNAs (about 34 % of the surveyed transcripts) contain short G-tracts in the downstream region of poly(A) signals. All tested elements were found to bind H/H' protein and stimulate the polyadenylation process [49]. In this work, the consensus for H/H' binding site was not identified, but the other studies demonstrated that this protein recognizes the GGGA/GGGGC [50], and probably GGGU containing tracts [51]. Interestingly, the GGGA sequence is recognized not only by H/H', but by all proteins of hnRNP H family [50] which also includes F and 2H9 proteins. The relevance of 2H9 protein to the polyadenylation process is currently unclear, while hnRNP F was shown to diminish 3'-end processing [52]. Most pre-mRNAs studied using our database contain GGGA/U or GGGGC tracts. About 40 % of transcripts contain such elements located immediately downstream and about 20 % — immediately upstream of the cleavage site.

The 5' splicing site can influence the polyadenylation process negatively when being positioned not only upstream of the cleavage site (BPV-1, Table 1, line 3), but downstream as well. This is the case of human immunodeficiency virus type 1 (HIV-1) (Table 2, line 2). However, in the case of HIV-1, in contrast to BPV-1, the first stage of the polyadenylation process but not the second one is inhibited [53]. Ashe et al. [54] showed that the binding site for U1 70K in stem-loop 1 of the U1 snRNP is associated with the cleavage inhibition. Regulation of poly(A) signal use in 5' LTR of HIV-1 pre-mRNA has two peculiarities. First, replacement of relatively weak wild-type poly(A) signal with the strong one abolishes 5' ss inhibitory effect [53]. Second, the sequence between the cap structure and AAUAAA is required for the regulation [55]. The mechanism of HIV-1 AUX DSE functioning is currently unknown. Searching our database, we found two transcripts with 5' ss in the remote downstream region.

In the bottom line of Table 2, the element which positively influences the calcitonin (CT) pre-mRNA polyadenylation [56, 57] is presented. Interestingly, the core sequence of this enhancer composed of pyrimidine tract and 5' splicing site could be modeled on a pseudomicroexon with 0 nt within the exon [57]. Splicing factors U1 snRNA, SRp20, PTB, and ASF/SF2 specifically bind to the enhancer core. The mechanism of CT pre-mRNA 3'-end processing stimulation by AUX DSE is presently unclear. According to the first model, PTB prevents U2AF binding to the enhancer pyrimidine tract that is

inhibitory for the enhancer activity [57]. By the second model, PTB or SRp20 can directly interact with polyadenylation factors [56, 57]. The third model is based on the fact that PTB can bind to a pyrimidine tract located between AAUAAA and the cleavage site in CT pre-mRNA. Dimerization of PTB bound to both the poly(A) signal and the enhancer core brings the other enhancer core-binding factors into proximity with polyadenylation factors. Screening our database, we found 9 pre-mRNAs having SRp20-binding sites in the downstream region, and 17 transcripts contain PTB-binding sites in the downstream region beyond the core poly(A) signal.

The downstream auxiliary elements are also found in such pre-mRNAs as IgM [58], adenovirus-5 L3 [59], adeno-associated virus [59], and HPV-31 [37]. However, the data on these AUX DSEs are poorly informative. AUX DSE is possibly present in human glycinamide ribonucleotide formyltransferase pre-mRNA (GART) [60]. Kan and Moran showed that it contains 24 nt long poly(U) tract in intronic region located immediately downstream of GU tract supposed to be the downstream element of poly(A) signal [60]. Poly(U) tract of the same length is also present close downstream of the core mouse poly(A) signal. The authors think that the conservation of poly(U) region is not fortuitous and this tract may be functional. Possibly it plays structural role. Some pre-mRNAs in our database also contain poly(U) tracts (~20 nt long) in the downstream region.

Thus, the upstream and downstream auxiliary elements are widely dispersed in viral and cellular pre-mRNAs and exert their effect on the polyadenylation process in different ways. First, AUX USEs and AUX DSEs can serve as binding sites for auxiliary proteins which influence the polyadenylation reaction positively, as for example, hnRNP H/H' [49], or negatively, as U1 A (N-terminal RBD) [15].

Secondly, they can be binding sites for general polyadenylation factors. Two kinds of these sites are known at present. Some of them are the same as the core elements of poly(A) signal, while the others represent the distinct sequences to which a factor binds through the second RNA binding domain. In the first case, the auxiliary element may be a trap for general polyadenylation factor (for example, for CstF [37]). In the second case, the general factor (for example, CPSF [25]) may interact simultaneously with both binding sites, that stabilizes its interaction with the poly(A) signal. Besides, the distinct sequence may be an entry site for the general factor [25].

Third, AUX USEs and DSEs could play a structural role, for example, providing favorable folding of pre-mRNA region covering poly(A) signal

where the core elements are well-exposed to the general polyadenylation factors [46]. Besides, the contacts between general polyadenylation factors and auxiliary proteins may be provided through dimerization of PTB molecules bound to pyrimidine tracts located in the vicinity of these protein binding sites [57].

Chen and Wilusz [59] suggested other possible function of auxiliary elements. They showed that pseudoknot from viral RNA could functionally substitute the unidentified Ad L3 AUX DSE. The authors supposed that stable secondary structures formed by auxiliary elements (like pseudoknot) may prevent sliding of CstF (relatively weak bound to RNA) along the pre-mRNA, and thereby limiting the region of interaction between the factor and the transcript mainly to the core poly(A) signal.

We suppose that some functions of auxiliary elements of poly(A) signals can be performed by four-stranded RNA structures, such as G-quadruplexes or i-motifs. We discuss below whether such structures may be formed in the region surrounding mammalian poly(A) signals and how they may influence the polyadenylation process.

Four-stranded RNA structures as possible auxiliary elements of poly(A) signals. Biologists begin to take an interest in four-stranded structures of polynucleotides during last 15 years when it has been discovered that G-quadruplexes can be involved in various important cellular events such as meiotic synapsis and recombination [61, 62], immunoglobulin class switch recombination [63], transcription regulation [64] and some others [65]. Most researches were focused on G-quadruplexes formed by DNA sequences, while a very little information was reported on RNA four-stranded structures. In particular, it was shown that the interstranded G-quadruplexes mediate the dimerization of HIV-1 RNA *in vitro* [66, 67]. Oliver et al. supposed that the gene 5 protein of filamentous bacteriophage fd may repress gene 2 mRNA translation through stabilizing the G-quadruplex formed by the operator sites of four mRNAs molecules [68]. As for cellular RNAs, Christansen et al. [69] showed that G-quadruplex is formed in the 3' untranslated region (UTR) of insulin-like growth factor II mRNA (IGF-II). This four-stranded structure was supposed to play a structural role ensuring that the site of endonucleolytic cleavage of IGF-II mRNA is accessible to interacting macromolecules and not sequestered in stable structures. Possible role of G-quadruplexes in RNA turnover was proposed by Bashkurov et al. [70] who purified mouse protein mXRN1p (homolog of the *Saccharomyces cerevisiae* Xrn1p exoribonuclease) which exhibits preference for

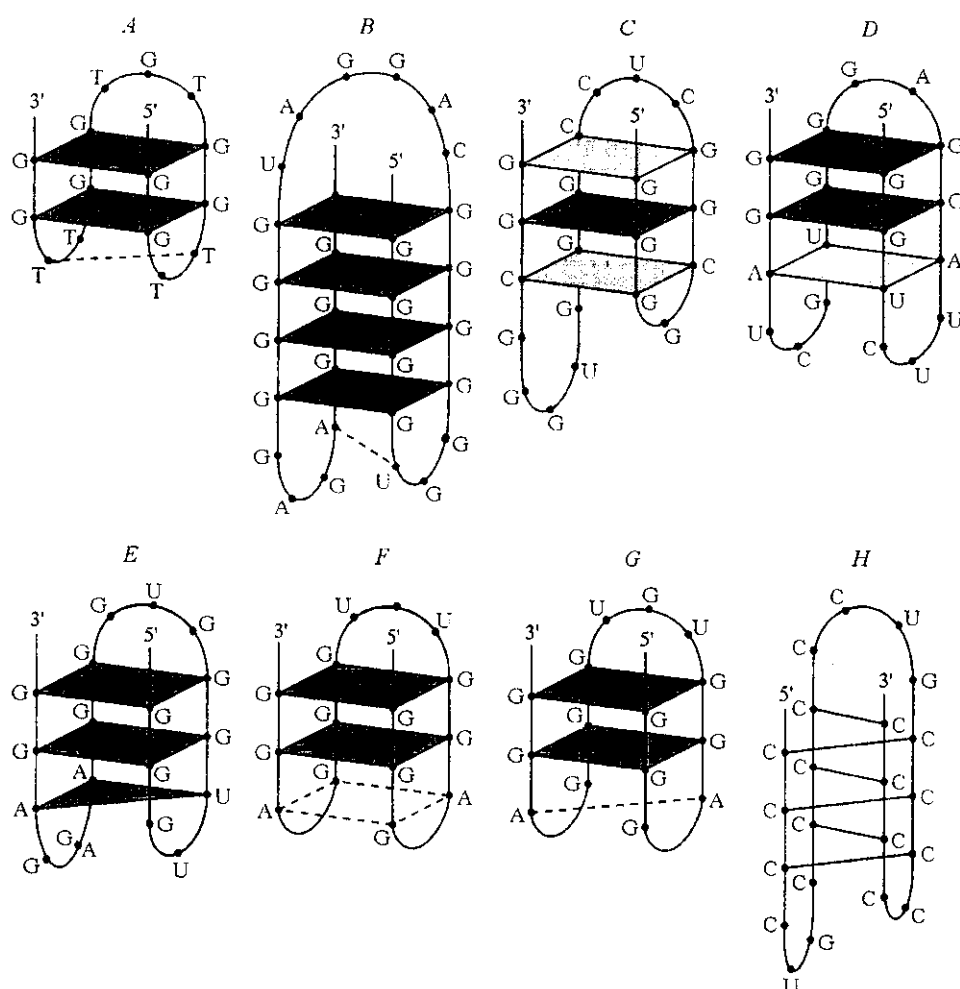


Fig. 3. G-quadruplex of thrombin-binding aptamer [75] (A); the patterns of pre-mRNA fragments folding into four-stranded structures: secretin (B), CD97 antigen (C), stromal interaction molecule 1 (D), dual specificity phosphatase (E), SV40 L (F and G), growth factor, augmentor of liver regeneration (H)

RNA G-quadruplex-containing substrate. These findings substantially confirm the suggestion that RNA four-stranded structures occur *in vivo*.

We were the first to suppose that G-quadruplexes can play role in the polyadenylation process. We found that many mammalian and eukaryotic viral pre-mRNAs contain clusters ( $n \geq 3$ ) of G-repeats in the region immediately downstream of core poly(A) signal and proposed that some of these G-rich sequences may form four-stranded structures [3]. Screening of our database of human poly(A) signals confirmed frequent occurrence of GRSs (with a potential to form G-quadruplexes) in the region downstream of core poly(A) signal [5]. Besides, we revealed that quadruplex-forming sequences also occur in the upstream region.

G-quadruplexes are composed of stacked G-tetrads which consist of four hydrogen-bonded copla-

nar guanines [71–74]. These structures can be formed by association of one, two or four polynucleotide molecules. Here, we examine mainly unimolecular quadruplexes. The quadruplexes with three or more G-tetrads are considerably more stable than those with two tetrads. For example, G-quadruplex formed by the thrombin-binding aptamer GGTTGG-TGTGGTTGG has  $T_m = 46.4^\circ\text{C}$  in the presence of 25 mM KCl [75], its folding pattern is shown in Fig. 3, A. The thrombin aptamer modified only in the number of tetrads (three tetrads instead of two) melts at  $64.5^\circ\text{C}$ . Thermodynamic stability of G-quadruplexes with two G-tetrads greatly depends on the sequence and the size of loops which connect G-tetrads [75–77]. In the case of the thrombin aptamer, the central loop shortening from TGT to TT significantly decreases the quadruplex stability ( $T_m = 21.0^\circ\text{C}$ ) [75]. Elongation of this loop to TTTT also

decreases the aptamer stability, but less effectively ( $T_m = 34.0$  °C), while the substitution of the central loop by the highly stable tetraloop GTAA leads only to slight change in melting temperature ( $T_m = 45.4$  °C). Marathias and Bolton [76] showed that the analog of thrombin aptamer GGTGGTGTGGT-GG is not folded into a unimolecular quadruplex, indicating that in this case one residue is not enough to make the external loops.

The central loop of unimolecular quadruplex was shown to connect guanine tracts either diagonally or edgewise [71–74]. It is of the edgewise type in the thrombin aptamer (Fig. 3, A), while increase in number of G-tetrads or elongation of the loops leads to formation of quadruplex with the diagonal central loop [76]. The other two types of G-quadruplex loops were identified ([78] and refs. therein) but we do not view them in this article.

Guanine tetrad core of quadruplex can be capped by base pairs, mismatches, triples or triads, which adds to its stability [71, 73]. The following triads and triple stacked over G-tetrads have been identified up to date: A·(T-A) [71], G·(C-A) [79], G·(T-T) [80], T·(A-A) [81] triads, and T·T·T triple [80]. Moreover, quadruplexes with G-tetrads can involve in the core structure other tetrads such as G·C·G·C [71], A·T·A·T [82], A·A·A·A [83], C·C·C·C [84], T·T·T·T [85], U·U·U·U [86], and T·G·T·G [87]. Besides, A·G·G·G·G·A hexad [88] and A·G·G·G·G pentad [78] were also identified in the quadruplex structures.

Thus, certain trends of formation of quadruplexes with G-tetrads could be summarized from experimental data reported. Though these trends were determined mainly for quadruplexes formed by G-rich DNA sequences, one may suppose that they are generally true for quadruplexes formed by G-rich RNA sequences too. Quadruplexes formed by one, two or four RNA molecules were reported in literature [66–69]. Moreover, in some cases G-quadruplex formed by RNA sequence is more stable than the equivalent DNA structure [89]. Based on the above-mentioned information, we have searched our database for the pre-mRNAs sequences with a potential to form quadruplexes. Approximately 27 % of pre-mRNAs studied appeared to contain such sequences. As many as 16 pre-mRNAs contain the sequences which may form quadruplexes with 3 or 4 tetrads. As an example, the model of secretin pre-mRNA fragment folding into G-quadruplex is presented in Fig. 3, B, guanine tetrads in this structure are flanked by A·U base pair. We found 13 pre-mRNAs with a potential to involve the A·U·A·U, G·C·G·C or U·G·U·G tetrads into core quadruplex structure. The

possible schematic structures of quadruplexes with the G·C·G·C tetrads or the A·U·A·U tetrad are depicted in Fig. 3, C and D, respectively. Some pre-mRNAs studied using our database are supposed to form quadruplexes capped by triads, the schematic drawing of G-quadruplex flanked by U·(A-A) triad is shown in Fig. 3, E. Four sequences found in our database have the potential to involve the A·A·A·A or U·U·U·U tetrads in the core quadruplex structure, but still there is no information on formation of such tetrads in unimolecular quadruplexes. These tetrads were shown to be involved in parallel stranded quadruplexes formed by four polynucleotide molecules [83, 86].

Searching our database, 9 pre-mRNAs appeared to contain the sequences which could be folded into quadruplexes with A·G·A·G tetrads. However, by means of *ab initio* quantum chemical study, Gu and Leszczynski found that compared to the two separated A·G pairs, the A·G·A·G tetrad has only 3.5 kcal/mol of stabilizing energy and suggested this tetrad may not be important in the quadruplexes structures [90]. Results of studies on quadruplex formation by telomeric repeat sequences,  $(T_2AG_3)_4$  [91] and  $(T_4AG_3)_4$  [92], are in agreement with this suggestion. The authors proposed two possible unimolecular quadruplex structures for these sequences: the quadruplex with two G-tetrads flanked by two A·G·A·G tetrads and the quadruplex with three G-tetrads (possibly flanked by one A·T·A·T tetrad [91]). Only the second model was supported by experimental observations, including chemical probing data [91, 92]. However, these results do not except the possibility of A·G·A·G tetrad formation in quadruplexes, because the intramolecular quadruplex with three highly stable G-tetrads in the case of  $(T_2AG_3)_4$  and  $(T_4AG_3)_4$  is very likely to be more energetically favorable than those with two G-tetrads capped by two low stable A·G·A·G tetrads. The stabilization energy relative to the isolated bases is 66.5, 29.4 and 32.7/30.4 kcal/mol for G·G·G·G, A·G·A·G and A·T·A·T (two forms), respectively [90]. Even if A and G residues in external loops of G-quadruplexes do not form A·G·A·G tetrads, these quadruplexes could be stabilized by formation of G·A or A·A mismatches.

The G-rich sequence of SV40 L pre-mRNA can also be folded into quadruplexes with GA external loops. Two variants of such structure are presented in Fig. 3, F and G. It should be noted that the putative quadruplex of SV40 L pre-mRNA GRS (Fig. 3, G) differs from that of thrombin-binding aptamer (Fig. 3, A) exclusively in the sequence of the external loops, possessing the same stable UGU central loop.

In view of the fact that the interaction between GA loops can stabilize SV40 L pre-mRNA GRS quadruplex, we consider the formation of four-stranded structure in the G-rich region of SV40 L poly(A) signal to be very likely.

However, the results of Hans and Alwine [93] who examined the secondary structure of SV40 L pre-mRNA in the region of poly(A) signal by nuclease sensitivity analysis techniques are not in agreement with formation of G-quadruplexes, since some guanine residues which we suppose to form G-tetrads are highly sensitive to T1 nuclease attack. This discrepancy could be due to the following reason. The nuclease sensitivity analysis was performed in the presence of 10 mM MgCl<sub>2</sub> [93]. Though divalent cations are known to stabilize G-quadruplexes at millimolar concentration, but at the concentration about 10 mM they can exert destabilizing effect [74]. So, the further investigations are needed to clarify the point whether or not G-quadruplexes are formed by SV40 L pre-mRNA GRS.

Taking into account the fact that hnRNP proteins (H, H' and F) specifically bind to poly(G) at 2 M NaCl [94], when poly(G) is a completely four-stranded macromolecule under equilibrium conditions [95], we recently supposed [5] that these proteins can specifically bind to quadruplexes formed by G-rich sequences of different pre-mRNAs. This binding may directly influence the cleavage complex assembly. Besides, G-quadruplexes may play a structural role, better exposing poly(A) signal elements to the cleavage factors. Also, G-quadruplexes may prevent sliding of general polyadenylation factors along RNA, which better tethers them to the binding sites.

In conclusion we briefly discuss four-stranded structures formed by C-rich polynucleotide sequences, that are i-motifs in which two cytidine stretches form a parallel-stranded duplex and two such duplexes are associated head-to-tail by base-pair intercalation into a quadruplex [96]. To form this structure, the C-stretches must be protonated. Poly(dC) protonation occurs under physiological pH, while poly(C) is ionized under the acidic pH ([97] and refs. therein). However, protonation of C-rich regions of RNAs inside the cell may be performed by proteins (see short review in [98]). Pre-mRNAs containing the sequences with a potential to form intercalated structures were found in our database ~1.4 times less often than ones which are capable to form G-quadruplexes. The sequences with the potential to form G-quadruplexes were shown to occur approximately 2-fold frequently in the downstream region than in the upstream region. This is completely contrast to the i-motifs occurrence. Analyzing -

70/+70 region, we found that 26 pre-mRNA sequences with the potential to form G-quadruplexes and 4 sequences with i-motifs are located downstream of the cleavage site, while 5 sequences with G-quadruplexes and 6 sequences with i-motifs are located upstream. The possible schematic structure of i-motif in GFER pre-mRNA is shown in Fig. 3, H.

Based on the fact that i-motifs of RNA sequences are much less stable than those of their DNA equivalents [99], we suppose the intercalation structures are very likely to participate indirectly in the polyadenylation process (at DNA level). RNA-polymerase II (Pol II) is known to be an essential cleavage factor in the polyadenylation reaction [1-3]. Yonaha and Proudfoot [100] showed that G-rich sequences ...TGGCCTTGGGGGAGGGGGAGGC... (which are the binding sites for the transcription factor MAZ) located downstream of polyadenylation signal in synthetic DNA template pause Pol II. This pausing leads to the stimulation of 3'-end processing of Pol II transcript. On the other hand, the pausing induced by a mutant form of EcoRI protein (which is defective in cleavage function but retains high affinity for the wild-type recognition sequence) does not activate polyadenylation when its binding site is inserted in DNA template. In fact, as seen from Fig. 5 shown in [100], the slight activation takes place in this case. In view of the fact that the GRS of the synthetic template may be folded into G-quadruplex, we propose the following mechanism of polyadenylation activation by both G-quadruplexes and i-motifs. Any four-stranded structure will pause Pol II and somewhat stimulate 3'-end processing. Specific proteins may stabilize the quadruplexes, thereby increasing the pausing as it probably occurs in the case of MAZ. Also, they may facilitate Pol II function in the cleavage reaction.

Thus, we suppose that four-stranded structures, particularly G-quadruplexes, may participate not only in such important cellular events, as recombination, transcription and others [65], but in the polyadenylation process as well.

**Acknowledgements.** We thank Prof. Anna V. El'skaya for advice and helpful comments on the article. We are also grateful to Dr. Olena Yo. Cherepenko for assistance and helpful discussions.

*М. І. Зарудна, А. Л. Потягайло, І. М. Коломісць, Д. М. Говорун*

*Допоміжні елементи полі(А)-сигналів пре-мРНК ссавців*

*Резюме*

*Поліаденілування — один з клітинних процесів, на рівні яких відбувається регуляція експресії генів. Допоміжні елементи полі(А)-сигналів пре-мРНК ссавців стимулюють або пригнічують реакцію поліаденілування, що спрямовується кореляційними сигналами. У цій роботі ми обговорюємо структу-*

ру і функції відомих в літературі допоміжних елементів та проводимо їхній пошук у створеній нами базі даних полі(A)-сигналів про-мРНК людини, яка містить 244 послідовності про-мРНК в області місця розщеплення. Аналіз літератури і створеної бази даних свідчить про широку поширеність у клітинних про-мРНК допоміжних елементів, зокрема послідовностей, що зв'язують U1 мРНК-специфічний U1 A білок. Окрім того, в створеній базі даних здійснено пошук послідовностей про-мРНК, які можуть утворювати чотирихланцюгові структури (G-квадруплєкси або i-мотиви). Ми припустили, що такі структури здатні виконувати деякі функції допоміжних елементів. У роботі наведено моделі квадруплєксів для фрагментів деяких про-мРНК людини, які ґрунтуються на аналізі літератури, присвяченій чотирихланцюговим структурам, а також моделі G-квадруплєксів для допоміжного G-багатого елемента полі(A)-сигналу про-мРНК SV40 L.

М. І. Зарудная, А. Л. Потягайло, И. Н. Коломиец,  
Д. Н. Говорун

Дополнительные элементы поли(A)-сигналов пре-мРНК  
млекопитающих

#### Резюме

Полиаденилирование — один из клеточных процессов, на уровне которых происходит регулирование экспрессии генов. Дополнительные элементы поли(A)-сигналов пре-мРНК млекопитающих стимулируют или ингибируют реакцию полиаденилирования, направляемую кор-элементами. В этой работе мы обсуждаем структуру и функции известных в литературе дополнительных элементов и проводим их поиск в созданной нами базе данных поли(A)-сигналов пре-мРНК человека, содержащей 244 последовательности про-мРНК в области места расщепления. Анализ литературы и созданной базы данных указывает на широкое распространение в клеточных про-мРНК дополнительных элементов, в частности последовательностей, связывающих U1 мРНК-специфичный U1 A белок. Кроме того, в созданной базе данных проведен поиск последовательностей про-мРНК, способных образовывать четырехцепочечные структуры (G-квадруплєкси или i-мотиви). Мы предположили, что такие структуры могут выполнять некоторые функции дополнительных элементов. В работе приведены модели квадруплєксов для фрагментов некоторых про-мРНК человека, основанные на анализе литературы, посвященной четырехспиральным структурам, а также модели G-квадруплєксов для дополнительного G-багатого элемента поли(A)-сигнала про-мРНК SV40 L.

#### REFERENCES

1. Wahle E., Rügsegger U. 3'-End processing of pre-mRNA in eukaryotes // *FEMS Microbiol. Rev.*—1999.—23, N 3.—P. 277—295.
2. Zhao J., Hyman L., Moore C. Formation of mRNA 3' ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis // *Microbiol. Mol. Biol. Rev.*—1999.—63, N 2.—P. 405—445.
3. Zarudnaya M. I. mRNA polyadenylation. 1. 3'-end formation of vertebrates' mRNAs // *Biopolimery i kletka.*—2001.—17, N 2.—P. 93—108.
4. Edwards-Gilbert G., Veraldi K. L., Milcarek C. Alternative poly(A) site selection in complex transcription units: means to an end? // *Nucl. Acids Res.*—1997.—25, N 13.—P. 2547—2561.
5. Zarudnaya M. I., Kolomiets I. M., Potyahaylo A. L., Hovorun D. M. Downstream elements of mammalian pre-mRNA polyadenylation signals: primary, secondary and higher-order structures // *Nucl. Acids Res.*—2003.—31.—(submitted).
6. Beaulieu E., Freier S., Wyatt J. R., Claverie J.-M., Gautheret D. Patterns of variant polyadenylation signal usage in human genes // *Genome Res.*—2000.—10, N 7.—P. 1001—1010.
7. Graber J. H., Cantor C. R., Mohr S. C., Smith T. F. In silico detection of control signals: mRNA 3'-end-processing sequences in diverse species // *Proc. Nat. Acad. Sci. USA.*—1999.—96, N 24.—P. 14055—14060.
8. MacDonald C. C., Redondo J.-L. Reexamining the polyadenylation signal: were we wrong about AAUAAA? // *Mol. Cell. Endocrinol.*—2002.—190, N 1—2.—P. 1—8.
9. Sheets M. D., Ogg S. C., Wickens M. P. Point mutations in AAUAAA and the poly(A) addition site: effects on the accuracy and efficiency of cleavage and polyadenylation in vitro // *Nucl. Acids Res.*—1990.—18, N 19.—P. 5799—5805.
10. Proudfoot N. Poly(A) signals // *Cell.*—1991.—64, N 4.—P. 671—674.
11. Chen F., MacDonald C. C., Wilusz J. Cleavage site determinants in the mammalian polyadenylation signal // *Nucl. Acids Res.*—1995.—23, N 14.—P. 2614—2620.
12. Gunderson S. I., Vagner S., Polycarpou-Schwarz M., Mattaj J. W. Involvement of the carboxyl terminus of vertebrate poly(A) polymerase in U1 A autoregulation and in the coupling of splicing and polyadenylation // *Genes and Develop.*—1997.—11, N 6.—P. 761—773.
13. Gunderson S. I., Polycarpou-Schwarz M., Mattaj J. W. U1 snRNP inhibits pre-mRNA polyadenylation through a direct interaction between U1 70K and poly(A) polymerase // *Mol. Cell.*—1998.—1, N 2.—P. 255—264.
14. Ko B., Gunderson S. I. Identification of new poly(A) polymerase-inhibitory proteins capable of regulating pre-mRNA polyadenylation // *J. Mol. Biol.*—2002.—318, N 5.—P. 1189—1206.
15. van Gelder C. W. G., Gunderson S. I., Jansen E. J. R., Boelens W. C., Polycarpou-Schwarz M., Mattaj J. W., van Venrooij W. J. A complex secondary structure in U1 A pre-mRNA that binds two molecules of U1 A protein is required for regulation of polyadenylation // *EMBO J.*—1993.—12, N 13.—P. 5191—5200.
16. Phillips C., Jung S., Gunderson S. I. Regulation of nuclear poly(A) addition controls the expression of immunoglobulin M secretory mRNA // *EMBO J.*—2001.—20, N 22.—P. 6443—6452.
17. Dietrich-Goetz W., Kennedy I. M., Levins B., Stanley M. A., Clements J. B. A cellular 65-kDa protein recognizes the negative regulatory element of human papillomavirus late mRNA // *Proc. Nat. Acad. Sci. USA.*—1997.—94, N 1.—P. 163—168.
18. Lutz C. S., Murthy K. G. K., Schek N., O'Connor J. P., Manley J. L., Alwine J. C. Interaction between the U1 snRNP-A protein and the 160-kD subunit of cleavage-polyadenylation specificity factor increases polyadenylation efficiency in vitro // *Genes and Develop.*—1996.—10, N 3.—P. 325—337.
19. Lutz C. S., Alwine J. C. Direct interaction of the U1 snRNP-A protein with the upstream efficiency element of the SV40 late polyadenylation signal // *Genes and Develop.*—1994.—8, N 5.—P. 576—586.
20. Russnak R. H. Regulation of polyadenylation in hepatitis B viruses: stimulation by the upstream activating signal PS1 is orientation-dependent, distance-independent, and additive // *Nucl. Acids Res.*—1991.—19, N 23.—P. 6449—6456.
21. Cherrington J., Russnak R., Ganem D. Upstream sequences and cap proximity in the regulation of polyadenylation in ground squirrel hepatitis virus // *J. Virol.*—1992.—66, N 12.—P. 7589—7596.

22. Natalizio B. J., Muniz L. C., Arhin G. K., Wilusz J., Lutz C. S. Upstream elements present in the 3' UTR of collagen genes influence the processing efficiency of overlapping polyadenylation signals // *J. Biol. Chem.*—2002. (epub ahead of print).
23. Gilmartin G. M., Fleming E. S., Oetjen J., Graveley B. R. CPSF recognition of an HIV-1 mRNA 3'-processing enhancer: multiple sequence contacts involved in poly(A) site definition // *Genes and Develop.*—1995.—9, N 1.—P. 72—83.
24. Brackenridge S., Proudfoot N. J. Recruitment of a basal polyadenylation factor by the upstream sequence element of the human lamin B2 polyadenylation signal // *Mol. and Cell. Biol.*—2000.—20, N 8.—P. 2660—2669.
25. Klasens B. I. F., Thiesen M., Virtanen A., Berkhout B. The ability of the HIV-1 AAUAAA signal to bind polyadenylation factors is controlled by local RNA structure // *Nucl. Acids Res.*—1999.—27, N 2.—P. 446—454.
26. Valsamakis A., Zeichner S., Carswell S., Alwine J. C. The human immunodeficiency virus type 1 polyadenylation signal: a 3' long terminal repeat element upstream of the AAUAAA necessary for efficient polyadenylation // *Proc. Nat. Acad. Sci. USA.*—1991.—88, N 6.—P. 2108—2112.
27. Graveley B. R., Gilmartin G. M. A common mechanism for the enhancement of mRNA 3' processing by U3 sequences in two distantly related lentiviruses // *J. Virol.*—1996.—70, N 3.—P. 1612—1617.
28. Prescott J., Falck-Pedersen E. Sequence elements upstream of the 3' cleavage site confer substrate strength to the adenovirus L1 and L3 polyadenylation sites // *Mol. and Cell. Biol.*—1994.—14, N 7.—P. 4682—4693.
29. Moreira A., Takagaki Y., Brackenridge S., Wollerton M., Manley J. L., Proudfoot N. J. The upstream sequence element of the C2 complement poly(A) signal activates mRNA 3' end formation by two distinct mechanisms // *Genes and Develop.*—1998.—12, N 16.—P. 2522—2534.
30. Zang W.-Q., Li B., Huang P.-Y., Lai M. M. C., Yen T. S. B. Role of polypyrimidine tract binding protein in the function of the hepatitis B virus posttranscriptional regulatory element // *J. Virol.*—2001.—75, N 22.—P. 10779—10786.
31. Huang Y., Wimler K. M., Carmichael G. G. Introless mRNA transport elements may affect multiple steps of pre-mRNA processing // *EMBO J.*—1999.—18, N 6.—P. 1642—1652.
32. Huang Y., Steitz J. A. Splicing factors SRp20 and 9G8 promote the nucleocytoplasmic export of mRNA // *Mol. Cell.*—2001.—7, N 4.—P. 899—905.
33. Gilmartin G. M., Hung S.-L., DeZazzo J. D. Sequences regulating poly(A) site selection within the adenovirus major late transcription unit influence the interaction of constitutive processing factors with the pre-mRNA // *J. Virol.*—1996.—70, N 3.—P. 1775—1783.
34. Key S. C. S., Pagano J. S. A noncanonical poly(A) signal, UAUAAA, and flanking elements in Epstein-Barr virus DNA polymerase mRNA function in cleavage and polyadenylation assays // *Virology.*—1997.—234, N 1.—P. 147—159.
35. Key S. C. S., Yoshizaki T., Pagano J. S. The Epstein-Barr virus (EBV) SM protein enhances pre-mRNA processing of the EBV DNA polymerase transcript // *J. Virol.*—1998.—72, N 11.—P. 8485—8492.
36. Aissouni Y., Perez C., Calmels B., Benech P. D. The cleavage/polyadenylation activity triggered by a U-rich motif sequence is differently required depending on the poly(A) site location at either the first or last 3' terminal exon of the 2'-5' oligo A synthetase gene // *J. Biol. Chem.*—2002.—277, N 39.—P. 35808—35814.
37. Cumming S. A., Repellin C. E., McPhillips M., Radford J. C., Clements J. B., Graham S. V. The human papillomavirus type 31 late 3' untranslated region contains a complex bipartite negative regulatory element // *J. Virol.*—2002.—76, N 12.—P. 5993—6003.
38. Koffa M. D., Graham S. V., Takagaki Y., Manley J., Clements J. B. The human papillomavirus type 16 negative regulatory RNA element interacts with three proteins that act at different posttranscriptional levels // *Proc. Nat. Acad. Sci. USA.*—2000.—97, N 9.—P. 4677—4682.
39. Vagner S., Vagner C., Mattaj J. W. The carboxyl terminus of vertebrate poly(A) polymerase interacts with U2AF 65 to couple 3'-end processing and splicing // *Genes and Develop.*—2000.—14, N 4.—P. 403—413.
40. Das Gupta J., Gu H., Chernokalskaya E., Gao X., Schoenberg D. R. Identification of two cis-acting elements that independently regulate the length of poly(A) on *Xenopus* albumin pre-mRNA // *RNA.*—1998.—4, N 7.—P. 766—776.
41. Gu H., Das Gupta J., Schoenberg D. R. The poly(A)-limiting element is a conserved cis-acting sequence that regulates poly(A) tail length on nuclear pre-mRNAs // *Proc. Nat. Acad. Sci. USA.*—1999.—96, N 16.—P. 8943—8948.
42. Das Gupta J., Gu H., Schoenberg D. R. Position and sequence requirements for poly(A) length regulation by the poly(A) limiting element // *RNA.*—2001.—7, N 7.—P. 1034—1042.
43. Bauren G., Belikov S., Wieslander L. Transcriptional termination in the *Balbani ring 1* gene is closely coupled to 3'-end formation and excision of the 3'-terminal intron // *Genes and Develop.*—1998.—12, N 17.—P. 2759—2769.
44. McCracken S., Lambermon M., Blencowe B. J. SRm160 splicing coactivator promotes transcript 3'-end cleavage // *Mol. and Cell. Biol.*—2002.—22, N 1.—P. 148—160.
45. Hir H. L., Izaurralde E., Maquat L. E., Moore M. J. The spliceosome deposits multiple proteins 20—24 nucleotides upstream of mRNA exon-exon junctions // *EMBO J.*—2000.—19, N 24.—P. 6860—6869.
46. Sittler A., Gallinaro H., Jacob M. The secondary structure of the adenovirus-2 L4 polyadenylation domain: evidence for a hairpin structure exposing the AAUAAA signal in its loop // *J. Mol. Biol.*—1995.—248, N 3.—P. 525—540.
47. Bagga P. S., Ford L. P., Chen F., Wilusz J. The G-rich auxiliary downstream element has distinct sequence and position requirements and mediates efficient 3' end pre-mRNA processing through a trans-acting factor // *Nucl. Acids Res.*—1995.—23, N 9.—P. 1625—1631.
48. Bagga P. S., Arhin G. K., Wilusz J. DSEF-1 is a member of the hnRNP H family of RNA-binding proteins and stimulates pre-mRNA cleavage and polyadenylation *in vitro* // *Nucl. Acids Res.*—1998.—26, N 23.—P. 5343—5350.
49. Arhin G. K., Boots M., Bagga P. S., Milcarek C., Wilusz J. Downstream sequence elements with different affinities for the hnRNP H/H' protein influence the processing efficiency of mammalian polyadenylation signals // *Nucl. Acids Res.*—2002.—30, N 8.—P. 1842—1850.
50. Caputi M., Zahler A. M. Determination of the RNA binding specificity of the heterogeneous nuclear ribonucleoprotein (hnRNP) H/H'/F/2H9 family // *J. Biol. Chem.*—2001.—276, N 47.—P. 43850—43859.
51. Fogel B. L., McNally L. M., McNally M. T. Efficient polyadenylation of Rous sarcoma virus RNA requires the negative regulator of splicing element // *Nucl. Acids Res.*—2002.—30, N 3.—P. 810—817.
52. Veraldi K. L., Arhin G. K., Martincic K., Chung-Ganster L.-H., Wilusz J., Milcarek C. hnRNP F influences binding of a 64-kilodalton subunit of cleavage stimulation factor to mRNA precursors in mouse B cells // *Mol. and Cell. Biol.*—2001.—21, N 4.—P. 1228—1238.
53. Ashe M. P., Pearson L. H., Proudfoot N. J. The HIV-1 5' LTR poly(A) site is inactivated by U 1 snRNP interaction with



- the downstream major splice donor site // *EMBO J.*—1997.—16, N 18.—P. 5752—5763.
54. Ashe M. P., Furger A., Proudfoot N. J. Stem-loop 1 of the U1 snRNP plays a critical role in the suppression of HIV-1 polyadenylation // *RNA*.—2000.—6, N 2.—P. 170—177.
  55. Furger A., Monks J., Proudfoot N. J. The retroviruses human immunodeficiency virus type 1 and Moloney murine leukemia virus adopt radically different strategies to regulate promoter-proximal polyadenylation // *J. Virol.*—2001.—75, N 23.—P. 11735—11746.
  56. Lou H., Neugebauer K. M., Gagel R. F., Berget S. M. Regulation of alternative polyadenylation by U1 snRNPs and SRp20 // *Mol. and Cell. Biol.*—1998.—18, N 9.—P. 4977—4985.
  57. Lou H., Helfman D. M., Gagel R. F., Berget S. M. Polypyrimidine tract-binding protein positively regulates inclusion of an alternative 3'-terminal exon // *Mol. and Cell. Biol.*—1999.—19, N 1. P. 75—85.
  58. Phillips C., Schimpl A., Dietrich-Goetz W., Clements J. B., Virtanen A. Inducible nuclear factors binding the IgM heavy chain pre-mRNA secretory poly(A) site // *Eur. J. Immunol.*—1996.—26, N 47.—P. 144—3152.
  59. Chen F., Wilusz J. Auxiliary downstream elements are required for efficient polyadenylation of mammalian pre-mRNAs // *Nucl. Acids Res.*—1998.—26, N 12.—P. 2891—2898.
  60. Kan J. L. C., Moran R. G. Intronic polyadenylation in the human glycylamide ribonucleotide formyltransferase gene // *Nucl. Acids Res.*—1997.—25, N 15.—P. 3118—3123.
  61. Sen D., Gilbert W. Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis // *Nature*.—1988.—334, N 6180.—P. 364—366.
  62. Muniyappa K., Anuradha S., Byers B. Yeast meiosis-specific protein Hop 1 binds to G4 DNA and promotes its formation // *Mol. Cell. Biol.*—2000.—20, N 4.—P. 1361—1369.
  63. Dempsey L. A., Sun H., Hanakahi L. A., Maizels N. G4 DNA binding by LRI and its subunits, nucleolin and hnRNP D, a role for G-G pairing in immunoglobulin switch recombination // *J. Biol. Chem.*—1999.—274, N 2.—P. 1066—1071.
  64. Simonsson T., Pecinka P., Kubista M. DNA tetraplex formation in the control region of *c-myc* // *Nucl. Acids Res.*—1998.—26, N 5.—P. 1167—1172.
  65. Shafer R. H., Smirnov I. Biological aspects of DNA/RNA quadruplexes // *Biopolymers*.—2001.—56, N 3.—P. 209—227.
  66. Sundquist W. I., Heaphy S. Evidence for interstrand quadruplex formation in the dimerization of human immunodeficiency virus 1 genomic RNA // *Proc. Nat. Acad. Sci. USA*.—1993.—90, N 8.—P. 3393—3397.
  67. Awang G., Sen D. Mode of dimerization of HIV-1 genomic RNA // *Biochemistry*.—1993.—32, N 42.—P. 11453—11457.
  68. Oliver A. W., Bogdarina I., Schroeder E., Taylor I. A., Kneale G. G. Preferential binding of fd gene 5 protein to tetraplex nucleic acid structures // *J. Mol. Biol.*—2000.—301, N 3.—P. 575—584.
  69. Christiansen J., Kofod M., Nielsen F. C. A guanosine quadruplex and two stable hairpins flank a major cleavage site in insulin-like growth factor II mRNA // *Nucl. Acids Res.*—1994.—22, N 25.—P. 5709—5716.
  70. Bashkurov V. I., Scherthan H., Solinger J. A., Buerstedt J.-M., Heyer W.-D. A mouse cytoplasmic exoribonuclease (mXRN 1p) with preference for G4 tetraplex substrates // *J. Cell. Biol.*—1997.—136, N 4.—P. 761—773.
  71. Patel D. J., Bouaziz S., Kettani A., Wang Y. Structures of guanine-rich and cytosine-rich quadruplexes formed *in vitro* by telomeric, centromeric, and triplet repeat disease DNA sequences // *Oxford handbook of nucleic acid structure* / Ed. S. Neidle.—Oxford: OXFORD University press, 1999.—P. 389—453.
  72. Keniry M. A. Quadruplex structures in nucleic acids // *Biopolymers*.—2001.—56, N 3.—P. 123—146.
  73. Simonsson T. G-quadruplex DNA structures—variations on a theme // *J. Biol. Chem.*—2001.—382, N 4.—P. 621—628.
  74. Hardin C. C., Perry A. G., White K. Thermodynamic and kinetic characterization of the dissociation and assembly of quadruplex nucleic acids // *Biopolymers*.—2001.—56, N 3.—P. 147—194.
  75. Smirnov I., Shafer R. H. Effect of loop sequence and size on DNA aptamer stability // *Biochemistry*.—2000.—39, N 6.—P. 1462—1468.
  76. Marathias V. M., Bolton P. H. Determinants of DNA quadruplex structural type: sequence and potassium binding // *Biochemistry*.—1999.—38, N 14.—P. 4355—4364.
  77. Jing N., Rando R. F., Pommier Y., Hogan M. E. Ion selective folding of loop domains in a potent anti-HIV oligonucleotide // *Biochemistry*.—1997.—36, N 41.—P. 12498—12505.
  78. Zhang N., Gorin A., Majumdar A., Kettani A., Chernichenko N., Skripkin E., Patel D. J. V-shaped scaffold: a non architectural motif identified in an A·(G·G·G·G) pentad-containing dimeric DNA quadruplex involving stacked G (anti)·G (anti)·G (anti)·G (syn) tetrads // *J. Mol. Biol.*—2001.—311, N 5.—P. 1063—1079.
  79. Al-Hashimi H. M., Majumdar A., Gorin A., Kettani A., Skripkin E., Patel D. J. Field- and phage-induced dipolar couplings in a homodimeric DNA quadruplex: relative orientation of G·(C-A) triad and G-tetrad motifs and direct determination of C2 symmetry axis orientation // *J. Amer. Chem. Soc.*—2001.—123, N 4.—P. 633—640.
  80. Kuryavii V., Majumdar A., Shallop A., Chernichenko N., Skripkin E., Jones R., Patel D. J. A double chain reversal loop and two diagonal loops define the architecture of a unimolecular DNA quadruplex containing a pair of stacked G (syn)·G (syn)·G (anti)·G (anti) tetrads flanked by a G·(T-T) triad and a T·T·T triple // *J. Mol. Biol.*—2001.—310, N 1.—P. 181—194.
  81. Kuryavii V., Kettani A., Wang W., Jones R., Patel D. J. A diamond-shaped zipper-like DNA architecture containing triads sandwiched between mismatches and tetrads // *J. Mol. Biol.*—2000.—295, N 3.—P. 455—469.
  82. Zhang N., Gorin A., Majumdar A., Kettani A., Chernichenko N., Skripkin E., Patel D. J. Dimeric DNA quadruplex containing major groove-aligned A·T·A·T and G·C·G·C tetrads stabilised by inter-subunit Watson-Crick A·T and G·C pairs // *J. Mol. Biol.*—2001.—312, N 5.—P. 1073—1088.
  83. Patel P. K., Koti A. S. R., Hosur R. V. NMR studies on truncated sequences of human telomeric DNA: observation of a novel A-tetrad // *Nucl. Acids Res.*—1999.—27, N 19.—P. 3836—3843.
  84. Patel P. K., Bhavesh N. S., Hosur R. V. NMR observation of a novel C-tetrad in the structure of the SV 40 repeat sequence GGGCGG // *Biochem. and Biophys. Res. Commun.*—2000.—270, N 3.—P. 967—971.
  85. Patel P. K., Hosur R. V. NMR observation of T-tetrads in a parallel stranded DNA quadruplex formed by *Saccharomyces cerevisiae* telomere repeats // *Nucl. Acids Res.*—1999.—27, N 12.—P. 2457—2464.
  86. Cheong C., Moore P. B. Solution structure of an unusually stable RNA tetraplex containing G- and U-quartet structures // *Biochemistry*.—1992.—31, N 36.—P. 8406—8414.
  87. Jing N., Hogan M. E. Structure-activity of tetrad-forming oligonucleotides as a potent anti-HIV therapeutic drug // *J. Biol. Chem.*—1998.—273, N 52.—P. 34992—34999.

88. Kettani A., Gorin A., Majumdar A., Hermann T., Skripkin E., Zhao H., Jones R., Patel D. J. A dimeric DNA interface stabilised by stacked A·(G·G·G·G)·A hexads and coordinated monovalent cations // *J. Mol. Biol.*—2000.—297, N 3.—P. 627—644.
89. Oliver A. W., Kneale G. G. Structural characterization of DNA and RNA sequences recognized by the gene 5 protein of bacteriophage fd // *Biochem. J.*—1999.—339, Pt 3.—P. 525—531.
90. Gu J., Leszczynski J. Structures and properties of mixed DNA bases tetrads: nonempirical *ab initio* HF and DFT studies // *J. Phys. Chem.*—2000.—104, N 9.—P. 1898—1904.
91. Balagurumoorthy P., Brahmachari S. K. Structure and stability of human telomeric sequence // *J. Biol. Chem.*—1994.—269, N 34.—P. 21858—21869.
92. Murchie A. I. H., Lilley D. M. J. Tetraplex folding of telomere sequences and the inclusion of adenine bases // *EMBO J.*—1994.—13, N 4.—P. 993—1001.
93. Hans H., Alwine J. C. Functionally significant secondary structure of the simian virus 40 late polyadenylation signal // *Mol. and Cell. Biol.*—2000.—20, N 8.—P. 2926—2932.
94. Matunis M. J., Xing J., Dreyfuss G. The hnRNP F protein: unique primary structure, nucleic acid-binding properties, and subcellular localization // *Nucl. Acids Res.*—1994.—22, N 6.—P. 1059—1067.
95. Souleil C., Panijel J. Immunochemistry of polyribonucleotides. Study of polyribinosinic and polyriboguanilyc acids // *Biochemistry.*—1968.—7, N 1.—P. 7—13.
96. Gueron M., Leroy J.-L. The i-motif in nucleic acids // *Curr. Opin. Struct. Biol.*—2000.—10, N 3.—P. 326—331.
97. Zarudnaya M. I., Potyahaylo A. L., Hovorun D. M. Conformational transition of poly(C) and poly(dC): study by the proton buffer capacity method // *Biopolymery i kletka.*—2000.—16, N 6.—P. 495—504.
98. Zarudnaya M. I., Hovorun D. M. Hypothetical double-helical poly(A) formation in a cell and its possible biological significance // *IUBMB Life.*—1999.—48, N 6.—P. 581—584.
99. Snoussi K., Nonin-Lecomte S., Leroy J. -L. The RNA i-motif // *J. Mol. Biol.*—2001.—309, N 1.—P. 139—153.
100. Yonaha M., Proudfoot N. J. Specific transcriptional pausing activates polyadenylation in a coupled *in vitro* system // *Mol. Cell.*—1999.—3, N 5.—P. 593—600.

УДК 577.21

Надійшла до редакції 20.06.02