
REVIEW

Bidirectional Promoters in the Transcription of Mammalian Genomes

A. S. Orekhova and P. M. Rubtsov*

*Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, ul. Vavilova 32,
119991 Moscow, Russia; fax: (499) 135-1405; E-mail: rubtsov@eimb.ru*

Received October 26, 2012

Revision received December 10, 2012

Abstract—In the genomes of humans and other mammals a large number of closely spaced pairs of genes that are transcribed in opposite directions were revealed. Their transcription is directed by so-called bidirectional promoters. This review is devoted to the characteristics of bidirectional promoters and features of their structure. The composition of “core” promoter elements in conventional unidirectional and bidirectional promoters is compared. Data on binding sites of transcription factors that are primarily specific for bidirectional promoters are discussed. The examples of promoters that share protein-coding genes transcribed by RNA polymerase II and the non-coding RNA genes transcribed by RNA polymerase III are described. Data obtained from global transcriptome analysis about the existence of short noncoding antisense RNA associated with the promoters in the context of the hypothesis of bidirectional transcription initiation as an inherent property of eukaryotic promoters are discussed.

DOI: 10.1134/S0006297913040020

Key words: genome structure, bidirectional promoters, divergent transcription, transcriptome, promoter core-elements, transcription initiation, CpG islands, transcription factors, short noncoding RNAs

The genomes of higher eukaryotes are excessive, and an important place in them is engaged by non-coding sequences. Many protein-coding genes have very extended 5' regulatory regions that contain alternate promoters, often separated by tens of thousands of nucleotides. Despite this redundancy in genomes of humans and other mammals there are a large number of closely spaced, oriented head-to-head and transcribed in opposite directions pairs of genes. Such gene pairs are separated by small intergenic region of less than one thousand base pairs (kb) length, which have a bidirectional promoter activity. This review covers the current understanding of the prevalence of bidirectional promoters in the genomes of mammals, peculiar properties of their organization and functioning, and of the transcription factors, binding sites of which are overrepresented in bidirectional promoters. In the final part of the review we discuss the hypothesis that the ability of bidirectional transcription initiation is an inherent property of eukaryotic promoters.

PREVALENCE OF BIDIRECTIONAL PROMOTERS IN MAMMALIAN GENOMES

Decoding of the nucleotide sequence of the human genome has opened the possibility for global analysis of gene organization. Already in early such studies, it was shown that tightly linked pairs of genes oriented head to head make up a very substantial fraction of the protein-coding genes [1, 2]. In particular, it was found that 31 of the 144 genes of chromosome 21, i.e. 22%, are located in the head-to-head orientation and are separated by spacers less than 1 kb [1]. A similar ratio (56 genes with such orientation of 319 (18%)) was found in chromosome 22 [1]. The search for such pairs of genes based on full-length cDNA mapping of the human genome showed slightly lower figures. Analysis of the distances between 23,752 genes showed that more than 10% of protein-coding genes of humans are transcribed from bidirectional promoters [2, 3]. In papers of Koyanagi et al. [4] and Frank et al. [5] it was shown that the enrichment of closely located bidirectional pairs of genes characteristic for the

* To whom correspondence should be addressed.

mammalian genome is less pronounced in other vertebrates. Prediction of bidirectional promoters in the genomes of humans and mammals was successfully performed using bioinformatic methods [6-8]. The study of features of mouse gene expression that are transcribed from bidirectional promoters by hybridization to microarrays showed that very often the genes transcribed in opposite directions from one bidirectional promoter are coordinately regulated [9]. The activities of bidirectional promoters in four different cell lines were defined, and it was found that the efficiency of transcription in opposite directions can be controlled to some extent depending on the cell type [2]. Thus, mammalian genomes contain a large number of bidirectional promoters directing transcription of closely oriented head-to-head gene pairs.

STRUCTURAL CHARACTERISTICS OF BIDIRECTIONAL PROMOTERS

In connection with the identification of multiple bidirectional promoters, a question about the features of their structure arose. Is it possible to identify consistent patterns that distinguish bidirectional and unidirectional promoters of protein-coding genes? In several studies the "core" elements in human bidirectional promoters were analyzed, in particular mapped TATA box, INR (initiator element), BRE (element recognizable by transcription factor IIB), DPE (downstream element) motifs, and the association of the promoters with CpG islands has been evaluated [1, 2, 10]. The arrangement of some core promoter elements of TATA-box lacking promoters is shown in Fig. 1.

The proximal promoter region of bidirectional promoters (~500 bp before and 100 bp after the transcription start site) may contain such elements as TATA- and CCAAT-boxes, INR, BRE, DPE, which together with CpG islands are responsible for basal promoter activity. The TATA motif located at position -30 relative to the transcription initiation point both in unidirectional promoters and bidirectional promoters [10]. In the absence of the TATA-box, the downstream promoter element DPE is required for binding of the common transcription factor

TFIID. The DPE feature is conservative in different eukaryotes – from *Drosophila* to humans – and is revealed not only in promoters lacking the TATA box. DPE acts together with INR and is usually located at position +30 with respect to the A₊₁ nucleotide in the INR motif [11]. The consensus motif of DPE – DSWYVY [12] – was found in 46.6% of bidirectional promoters compared with 50.6% of unidirectional promoters [10]. The typical DPE-dependent promoter contains INR and DPE. Thus, DPE and INR act together [11]. TBP-associated factors (TAFs) [13], in particular TAF6 and TAF9 [14], are responsible for the recognition of DPE.

The INR element of DNA sequence overlaps the point of transcription initiation and is sufficient to determine the position of this point in a promoter that does not contain a TATA box; moreover, it increases the strength of TATA-containing promoter [15]. INR is functionally similar to the TATA box and can function independently from it [11], and together with the TATA motif or DPE [10]. For accurate transcription initiation *in vitro* and *in vivo*, the sequence of INR between -3 and +5 is necessary and sufficient. Research on the role of the distance between the TATA box and INR has shown that the two elements work synergistically when they are separated by 25-30 bp, and they operate independently if the distance between them is more than 30 bp. The consensus sequence of INR is YYANWYY [11]. It was shown that this motif is contained in 25.3% of bidirectional promoters and 30.8% of unidirectional promoters [10].

The BRE element (transcription factor IIB recognizing element) is located directly in front of the TATA box in some TATA-containing promoters and affects the ability of transcription factor IIB to be comprised in transcription complexes and support the initiation of transcription [13]. The incidence of this element is 16.5 and 11.1% in bidirectional and unidirectional promoters, respectively [10]. The consensus sequence of BRE is SSRCGCC [13]. The TFIIB-BRE interaction may play a dominant role in the assembly of pre-initiating complex and transcription initiation at promoters lacking the TATA box [13]. The CCAAT motif is a consensus sequence that is located at a distance of 75-80 bases before the transcription start site. Its occurrence in bidi-

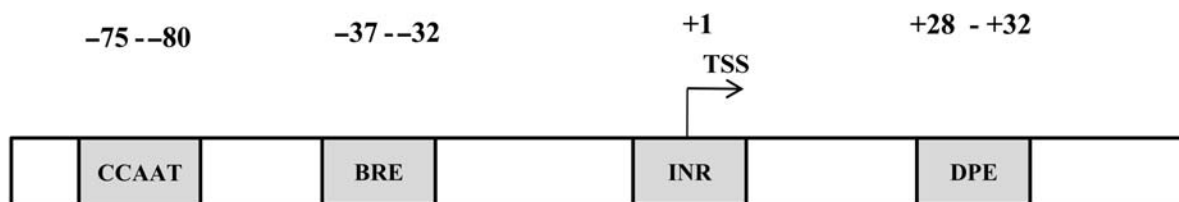


Fig. 1. Scheme of relative position of promoter elements in typical TATA-box lacking promoters. The positions of the CCAAT-box, BRE (TFIIB Recognition Element), INR (Initiator element), and DPE (Downstream Promoter Element) are shown relative to the transcription start site (TSS). Consensus sequences of elements: CCAAT – GGCCAATCT; BRE – SSRCGCC; INR – YYANWYY; DPE – DSWYVY(T). Designations of degenerate nucleotides correspond to IUPAC nomenclature.

rectional promoters is 12.9% and in unidirectional promoters 6.9% [10].

Comparison of the structure of uni- and bidirectional promoters has shown that bidirectional promoters more frequently than unidirectional ones lack TATA boxes. Bidirectional promoters more often contain BRE and CCAAT elements. However, promoters of these two types differ barely in the content of DPE and INR. In part, the difference in the set of "core" promoter elements can be explained by the fact that bidirectional promoters are enriched in GC-pairs.

Connection between bidirectional promoters and CpG islands. In full-genome computer analysis, it was found that more than 70.8% of the human bidirectional promoters are GC rich (>60%) and enriched with binding sites for transcription factor Sp1 [10]. Bidirectional promoters are closely associated with CpG islands and enriched binding sites for several other transcription factors, such as GABPA, MYC, E2F1, E2F4, Nrf1, YY1, NF-Y [3, 16].

The CpG islands are found in the genomes of many vertebrates, and they represent regions of DNA that are enriched in CpG dinucleotides and unmethylated [17]. They have high G+C-content (more than 50%) and high-density of CpG dinucleotides [18]. The human genome contains ~29,000 CpG islands. The size of the islands is usually 0.5-2 kb in length. In mammals, CpG islands are associated with approximately 60% of the promoters of protein-coding genes [19]. The CpG islands are characterized by multiple binding sites for the transcription factor Sp1 [11]. The CpG islands are usually located in the 5' regions of housekeeping genes and many tissue-specific genes [18, 19], for example, genes that are expressed in the brain and nervous system.

Bidirectional promoters have an average G+C content of 66%, while for unidirectional promoters it is 53%. Moreover, 77% of bidirectional promoters are located within CpG islands compared with 38% of unidirectional promoters [2]; according to other sources ~90 and 45% of bidirectional and unidirectional promoters, respectively, are closely linked with CpG islands [10]. Promoters associated with CpG are more resistant to methylation. Hypermethylation of a CpG island in the promoter region usually leads to suppression of gene expression, and the promoters of some tumor suppressor genes are hypermethylated in cancer. Shu et al. [20] found that bidirectional promoters constitute more than 25% of all promoters that are hypermethylated and located in CpG islands in cancer cells. By analyzing three pairs of genes (*WNT9A/CD558500*, *CTDSPL/BC040563*, and *KCNK15/BF195580*) it was shown that the degree of methylation is inversely correlated with the content of the mRNA of these genes in various cancer cell lines. Treatment of cells with 5-aza-2'-deoxycytidine, which causes demethylation of CpG dinucleotides, leads to increased expression of these genes. Thus, hypermethylation of bidirectional

promoters associated with CpG islands is accompanied by silencing of two genes simultaneously.

In most cases, a CpG island covers partially or completely a region of the first exon (and sometimes more than one exon) of each gene. CpG island size, which is determined by the number of CpG dinucleotides, however, is not a determinant of bidirectional transcription [1] because no differences in the length and sequence features of CpG islands associated with bidirectional and unidirectional promoters was shown [2].

Binding sites of transcription factors in bidirectional promoters. As already mentioned, bidirectional promoters are enriched with specific binding sites for transcription factors such as GABPA, MYC, E2F1, E2F4, Nrf1, YY1, NF-Y, and SP1 [3, 16]. In many cases the role of the binding of these transcription factors in the functioning of the promoters was confirmed experimentally. For example, it was shown that the bidirectional promoter of two genes for the mitochondrial translation apparatus in mice is regulated by a CCAAT box that interacts with transcription factor NF-Y [21]. The genes of the mitochondrial ribosomal protein S12 (*Mrps12*) and mitochondrial seryl-tRNA ligase in the genomes of both humans and mice are transcribed in opposite directions from a conservative promoter of less than 200 bp. Using a reporter vector, the authors identified a set of four CCAAT boxes required for efficient transcription of two genes in mouse 3T3 cell culture, as well as for enhancement of the transcription of the *Mrps12* gene. By gel mobility shift assay and *in vivo* footprinting, it was confirmed that these elements that act as binding sites for the transcription factor NF-Y are important. This protein was found in the promoter in all tissues examined by chromatin immunoprecipitation (liver, brain, heart, kidneys, and NIH3T3 cells). Bidirectional activity characteristic of NF-Y makes it an especially suitable regulation factor of this class of promoters that direct the expression of genes associated with cell proliferation. A similar organization and regulation of expression with CCAAT boxes and transcription factor NF-Y is characteristic of the bidirectional promoter of orthologous human genes [22]. Genome-wide analysis of the distribution of NF-Y binding sites CCAAT in uni- and bidirectional promoters of the mouse and human genomes revealed that bidirectional promoters have special allocation of these boxes and are highly conserved in their nucleotide sequences [23].

In one of the first studies analyzing the differences in the repertoires of transcription factors that bind to bidirectional compared with unidirectional promoters, a special role of transcription factor GABP from the ets-family was established [3]. The binding of GABP was verified with 121 and 291 randomly selected bidirectional and unidirectional promoters in three different human cell lines (HeLa, Jurkat, and K562) by chromatin immunoprecipitation. It was shown that GABP binds to 86.6% of bidirectional and only 30.6% of unidirectional promoters

in at least one of the three cell lines [3]. The GABP binding correlated well with bidirectional promoter activity in the reporter gene expression system of luciferase. Moreover, integration of the consensus sequence of the GABP binding site in randomly selected unidirectional promoters in 67% of cases caused an oppositely directed activity, so that a unidirectional promoter was transformed into a bidirectional promoter [3].

Recently, using bioinformatics and biochemical techniques, it was shown that another transcription factor, hStaf/Znf143, is involved in the control of gene expression by bidirectional promoters [24]. Staf was originally discovered as a transcriptional activator of the selenocysteine tRNA gene in *Xenopus laevis* [25]. Later it was found that this protein with zinc fingers is involved in the regulation of gene expression of many small nuclear and small nucleolar RNAs and protein-coding genes, i.e. it is an activator of genes that are transcribed by RNA polymerase II and III [26]. Staf homologs in the human genome are encoded by genes *ZNF143* and *ZNF76* [27]. Later, hStaf/ZNF143-binding sites in the promoters of individual genes have been mapped [26, 28, 29]. Anno et al. [24] confirmed the fact that bidirectional promoters are enriched in functional binding sites of hStaf/ZNF143. They chose genes that transcribed in opposite directions with non-overlapping 5' ends separated by intergenic regions of less than 1000 bp from the human genome. In that analyses set containing 1678 genes (839 bidirectional promoters), binding sites of hStaf/ZNF143 were over-represented in comparison with unidirectional promoters. Direct mapping of the transcription factor by chromatin immunoprecipitation confirmed that most of the predicted binding sites in bidirectional promoters were really bound by this protein *in vivo*. Thus, in a random sample of 87 bidirectional promoters that separate the protein-coding genes, 84 contain functional binding sites for hStaf/ZNF143 [24]. In experiments on gene overexpression and knockdown of *hStaf/ZNF143*, the participation of this factor in the activation of transcription from bidirectional promoters was confirmed.

Thus, transcription from bidirectional promoters is regulated by a specific set of protein factors, in which proteins GABP and hStaf/ZNF143 play a special role.

BIDIRECTIONAL PROMOTERS AND NONCODING RNA GENES

Due to the fact that the Staf is an activator of transcription of both RNA polymerase II and RNA polymerase III [26], there is special interest in the characterization of bidirectional promoters associated with pairs of noncoding RNA genes, especially with couples of protein-coding gene and noncoding RNA genes. The number of such promoters in the human genome is small. In the already mentioned earlier representative sample of

bidirectional promoters, their number is 12 and 26, respectively [24]. All 12 promoters from the first of these groups separate pairs of tRNA genes, apparently transcribed by RNA polymerase III. The bidirectional promoter from which the one-way RNA polymerase III transcribes noncoding RNA H1 (RNA component of ribozyme RNase P) and the other way RNA polymerase II transcribes mRNA of poly(ADP-ribose)-polymerase 2 (PARP2) was previously identified in the mouse genome and described in detail [30]. The distance between the initiation points of gene transcription of RNA H1 and PARP2 is only 114 bp. A similar organization of these genes is found in the genomes of other mammals including humans. Divergently oriented pairs with protein-coding genes in the human genome constitute another 25 small noncoding RNA genes: genes of the RNA component of signal-recognition particle (SRP), RNA component of ribonucleoprotein MRP involved in initiation of mtDNA replication and processing of precursor rRNA in the nucleus, RNA 7SK, RNA of Cajal bodies scaRNA17 (U91), small nuclear RNA U6.2, U6.9, and U12, small nucleolar RNA U13, and 17 different tRNAs [24].

The bidirectional promoter that separates genes PARP2 and RNA H1 was used to create a vector with simultaneous expression of protein-coding genes and genes of short hairpin RNA (shRNA) – the precursors of small interfering RNA (siRNA) [31]. The effectiveness of these vectors was confirmed by simultaneous expression of the green fluorescent protein gene and shRNA specific to luciferase mRNA. It was also shown that the bidirectional expression from this promoter can regulate the expression of functionally related genes in the cell. Thus, the authors provided an overexpression of cloned into vector p53 tumor suppressor gene and also suppressed the expression of the cellular gene of antiapoptotic protein Bcl-2 by specific shRNA transcribed from the same bidirectional promoter [31]. This approach may be of interest for developing methods of gene therapy to overcome the action of dominant-negative mutations.

Bidirectional promoters of protein-coding genes can be used to create expression vectors. We have successfully used one of these promoters controlling the expression of genes of hypothetical human proteins with unknown function – CCDC142 (Coiled-Coil Domain Containing) and TTC31 (TetraTricopeptide repeat Containing) for the simultaneous expression of genes in the same cell of two fluorescent proteins – green (EGFP) and red (DsRed2) [32].

BIDIRECTIONAL TRANSCRIPTION INITIATION AS AN INHERENT PROPERTY OF EUKARYOTIC PROMOTERS

In recent years, due to the development and application of methods for global transcriptome analysis, there is a lot of new information requiring the revision of some of

New classes of RNA associated with promoters and transcription initiation sites of mammalian genes identified in global transcriptome analysis

Short note	Full name	Length	Characterization	Reference
TSSa-RNA	transcription start site-associated anti-sense RNAs	20-90 nt	short noncoding RNAs formed by two-way activity of RNA polymerase	[37, 38]
PASR	promoter-associated short RNAs	20-200 nt	short RNAs that overlap the area of "core" promoters	[39, 40]
NRO-RNA	RNA detected by nuclear "run-on" transcription	20-50 nt	short RNAs revealed by global sequencing of nuclear "run-on" transcripts; correspond to areas adjoined to points of transcription initiation	[41]
tiRNA	transcription initiation RNAs	~18 nt	RNAs corresponding to the ~20 bp long region downstream of the transcription initiation point	[42]
PROMPT	promoter upstream transcripts	0.5-2 knt	unstable transcripts mapped in the area upstream the transcription initiation sites	[43-45]

the concepts of the transcription of mammalian genomes. It was found that a much larger share of genomes is transcribed than previously thought. By some estimates, more than 70% of nucleotide sequences of the human genome are transcribed, of which only 2% are protein-coding transcripts [33]. A considerable part of the noncoding transcripts is mapped to the boundaries of genes, including the promoter regions. In addition it was found that the promoters of protein-coding genes are areas in which there is constant (pervasive) transcription, and the initiation of transcription occurs in both directions [34-36]. This conclusion is based on the detection and detailed characterization of several new classes of RNAs that are associated with transcription initiation sites and promoters of actively transcribed genes (table). Let us dwell on the characteristics of these noncoding RNAs.

Short noncoding RNAs that are associated with sites of transcription initiation, TSSa-RNAs, were originally discovered in the analysis of libraries of cDNA synthesized on the template of short RNAs from mouse embryonic stem cells [37]. They represent RNA of 20-90 nucleotides in length. Most of them are mapped to genome segments that cover transcription start sites (TSS) of RNA polymerase II. These RNAs correspond to sequences of both the sense strand and the antisense strand. The first mapped predominantly in the range of +1 to +50 with respect to the TSS, and the second in the region between -100 and -300 [37, 38]. It was also found that the localization of sense and antisense TSSa-RNAs in chromatin coincides with the occupation of these sites by RNA polymerase II that initiates transcription and by trimethylated at Lys4 histone H3 (3met-H3K4) – markers of transcription initiation [37]. It is noteworthy that TSSa-RNAs were also found in embryonic stem cells lacking RNase Dicer,

that is, they are not products of the processing of longer RNA by the RNA interference mechanism [37].

Total analysis of nuclear and cytoplasmic RNAs of less than 200 nucleotides in human cell lines using high-resolution hybridization on microarrays and sequencing showed that a significant portion of these RNAs mapped to the 5' and 3' flanks of protein-coding genes [39, 40]. They were named PASR (promoter-associated RNA) and TASR (termini-associated RNA), respectively. PASR length varies from 22 to 200 nt, and their content in the cell is correlated with the levels of expression of the corresponding genes. Sites complementary to PASR coincide in the genomes of different human cell lines, in particular in the genomes of lines HepG2 and HeLa. The regions in the genome corresponding to the PASR, as well as TSSa-RNA, are associated with markers of active transcription (presence of RNA polymerase II, acetylation of histones H3 and H4, presence of 3met-H3K4, and sensitivity to DNase I) [40].

Similar results were obtained using a different approach, the global sequencing of nuclear "run-on"-transcripts (GRO-seq) [41]. This approach allowed mapping the genome-wide location and orientation of RNA polymerase. It turned out that most of the promoters contain RNA polymerase II bound to the upstream region in the orientation opposite to the main direction of transcription of the gene. It was concluded that the possibility of productive transcription and its direction is determined by the interaction of RNA polymerase and various regulators on the relatively extensive promoter region.

Another two classes of short RNAs mapped near the TSS are tiRNA [42] and PROMPT [43-45]. The tiRNAs correspond to the short segments of the sense strand following immediately after the TSS, and the PROMPT

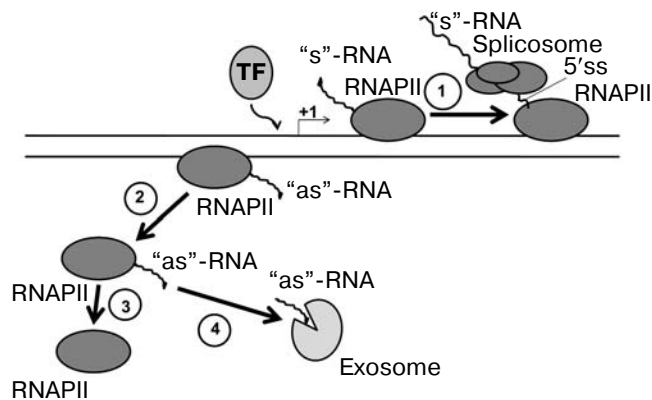


Fig. 2. Scheme of bidirectional transcription initiation by RNA polymerase II. RNA polymerase II (RNAPII) with the participation of transcription factors (TFs) binds to a site close to the transcription initiation site (+1) and initiates the synthesis of short sense and antisense transcripts ("s"- and "as"-RNAs). After that, the transcription stops in both directions. Then, in the case of unidirectional promoters there is a transition to productive elongation of transcription (1) in one direction and coupled with transcription pre-mRNA splicing. The RNAPII complex with "as"-RNA, in turn dissociates from the promoter (2), RNAPII is released (3), and the "as"-RNA is degraded by exosomes (4). In the case of bidirectional promoters, productive elongation of transcription goes in both directions. The 5'ss is the donor splice site.

(PROMoter uPstream Transcripts) are mapped to the TSS and with both the sense and antisense strands. PROMPT are extremely unstable, but their content is increased during suppression by RNA interference of the system of 3'-5' degradation of eukaryotic mRNA – RNA exosomes [43]. It was shown that PROMPT is polyadenylated at the 3'-termini and is 5'-capped [45]. Moreover, it was found that these RNAs are also synthesized from promoters that are actively transcribed by RNA polymerases I and III [45].

The active synthesis of short transcripts associated with the promoter regions of genes and the corresponding to both strands of DNA is confirmed by recent publications by participants of the ENCODE project (*Encyclopedia of DNA Elements*) aimed at detailed mapping of transcription sites and other functional regions of the human genome [46-48].

The data presented in this part of the review suggest that the ability to initiate transcription in opposite directions is an inherent property of eukaryotic promoters. In the case of unidirectional promoters, productive transcription elongation is possible only in one direction, and the short transcripts initiated from the antisense strand do not elongate and are degraded. This scenario is shown schematically in Fig. 2. The factors governing the choice of the direction of productive transcription initiation after initial two-way initiation remained poorly understood until recently. But now it has been shown that the adopting of the loop conformation by the gene contributes to

the transition to a productive synthesis of mRNA [49]. Formation of such a conformation depends on the promoter-associated transcription factors and polyadenylation complex factors bound to the 3' end of the gene (*pAC*), which interact with each other through the mediation of protein factor Ssu72 [49]. Thus, the Ssu72-dependent convergence of the 5' and 3' ends of the gene determines the direction of transcription from the unidirectional promoter. For productive transcription in both directions, the promoter probably should contain additional elements, in particular functional binding sites of specific transcription factors such as GABP or hStaf/Znf143. It is unclear whether there will be in this case an interaction of the promoter region with the 3'-terminal regions of two divergently transcribed genes. A more detailed study of the factors and conditions that promote productive bidirectional transcription has significant relevance for future research in the field of eukaryotic transcription.

This work was supported by the Russian Foundation for Basic Research (project 09-04-01497a).

REFERENCES

- Adachi, N., and Lieber, M. R. (2002) *Cell*, **109**, 807-809.
- Trinklein, N. D., Aldred, S. H., Hartman, S. J., Schroeder, D. I., Otilar, R. P., and Myers, R. M. (2004) *Genome Res.*, **14**, 62-66.
- Collins, P. J., Kobayashi, Y., Nguyen, L., Trinklein, N. D., and Myers, R. M. (2007) *PLoS Genet.*, **3**, e208.
- Koyanagi, K. O., Hagiwara, M., Itoh, T., Gojobori, T., and Imanishi, T. (2005) *Gene*, **353**, 169-176.
- Franck, E., Hulsen, T., Huynen, M. A., De Jong, W. W., Lubsen, N. H., and Madsen, O. (2008) *Mol. Biol. Evol.*, **25**, 1909-1921.
- Yang, M. Q., and Elnitski, L. L. (2007) *Lect. Notes Comput. Sci.*, **4463**, 361-371.
- Yang, M. Q., and Elnitski, L. L. (2008) *BMC Genom.*, **9** (Suppl. 1), S2.
- Piontkivska, H., Yang, M. Q., Larkin, D. M., Lewin, H. A., Reecy, J., and Elnitski, L. (2009) *BMC Genom.*, **10**, 189.
- Engstrom, P. G., Suzuki, H., Ninomiya, N., Akalin, A., Sessa, L., Lavorgna, G., Brozzi, A., Lizi, L., Tan, S. L., Yang, L., Kunarso, G., Ng, E. L., Batalov, S., Wahlestedt, C., Kai, C., Kawai, J., Carninci, P., Hayashizaki, Y., Wells, C., Bajic, V. B., Orlando, V., Reid, J. F., Lenhard, B., and Lipovich, L. (2006) *PLoS Genet.*, **2**, e47.
- Yang, M. Q., and Elnitski, L. L. (2008) *BMC Genom.*, **9** (Suppl. 2), S3.
- Smale, S. T., and Kadonaga, J. T. (2003) *Annu. Rev. Biochem.*, **72**, 449-479.
- Kutach, A. K., and Kadonaga, J. T. (2000) *Mol. Cell. Biol.*, **20**, 4754-4764.
- Lagrange, T., Kapanidis, A. N., Tang, H., Reinberg, D., and Ebricht, R. H. (1998) *Genes Dev.*, **12**, 34-44.
- Yang, C., Bolotin, E., Jiang, T., Sladek, F. M., and Martinez, E. (2007) *Gene*, **389**, 52-65.

15. Javahery, R., Khachi, A., Lo, K., Zenzie-Gregory, B., and Smale, S. T. (1994) *Mol. Cell. Biol.*, **14**, 116-127.
16. Lin, J. M., Collins, P. J., Trinklein, N. D., Fu, Y., Xi, H., Myers, R. M., and Weng, Z. (2007) *Genome Res.*, **17**, 818-827.
17. Bird, A. P. (1986) *Nature*, **321**, 209-213.
18. Gardiner-Garden, M., and Frommer, M. (1987) *J. Mol. Biol.*, **196**, 261-282.
19. Antequera, F. (2003) *Cell. Mol. Life Sci.*, **60**, 1647-1658.
20. Shu, J., Jelinek, J., Chang, H., Shen, L., Qin, T., Chung, W., Oki, Y., and Issa, J. P. (2006) *Cancer Res.*, **66**, 5077-5084.
21. Zanutto, E., Shah, Z. H., and Jacobs, H. T. (2007) *Nucleic Acids Res.*, **35**, 664-677.
22. Zanutto, E., Lehtonen, V., and Jacobs, H. T. (2008) *Biochim. Biophys. Acta*, **1783**, 2352-2362.
23. Hakkinen, A., Healy, S., Jacobs, H. T., and Ribeiro, A. S. (2011) *J. Theor. Biol.*, **281**, 74-83.
24. Anno, Y.-N., Myslinski, E., Ngondo-Mbongo, R. P., Krol, A., Poch, O., Lecompte, O., and Carbon, P. (2011) *Nucleic Acids Res.*, **39**, 3116-3127.
25. Schuster, C., Myslinski, E., Krol, A., and Carbon, P. (1995) *EMBO J.*, **14**, 3777-3787.
26. Schaub, M., Myslinski, E., Schuster, C., Krol, A., and Carbon, P. (1997) *EMBO J.*, **16**, 173-181.
27. Myslinski, E., Krol, A., and Carbon, P. (1998) *J. Biol. Chem.*, **273**, 21998-22006.
28. Schaub, M., Krol, A., and Carbon, P. (1999) *J. Biol. Chem.*, **274**, 24241-24249.
29. Myslinski, E., Gerard, M. A., Krol, A., and Carbon, P. (2006) *J. Biol. Chem.*, **281**, 39953-39962.
30. Ame, J.-C., Schreiber, V., Fraulob, V., Dolle, P., de Murcia, G., and Niedergang, C. P. (2001) *J. Biol. Chem.*, **276**, 11092-11099.
31. Hung, C. F., Cheng, T. L., Wu, R. H., Teng, C. F., and Chang, W. T. (2006) *Biochem. Biophys. Res. Commun.*, **339**, 1035-1042.
32. Orekhova, A. S., Sverdlova, P. S., Spirin, P. V., Leonova, O. G., Popenko, V. I., Prasolov, V. S., and Rubtsov, P. M. (2011) *Mol. Biol. (Moscow)*, **45**, 486-495.
33. The ENCODE Project Consortium (2007) *Nature*, **447**, 799-816.
34. Jacquier, A. (2009) *Nat. Rev. Genet.*, **10**, 833-844.
35. Carninci, P. (2010) *DNA Res.*, **17**, 51-59.
36. Wei, W., Pelechano, V., Jarvelin, A. I., and Steinmetz, L. M. (2011) *Trends Genet.*, **27**, 267-276.
37. Seila, A. C., Calabrese, J. M., Levine, S. S., Yeo, G. W., Rahl, P. B., Flynn, R. A., Young, R. A., and Sharp, P. A. (2008) *Science*, **322**, 1849-1851.
38. Seila, A. C., Core, L. J., Lis, J. T., and Sharp, P. A. (2009) *Cell. Cycle*, **8**, 2557-2564.
39. Kapranov, P., Cheng, J., Dike, S., Nix, D. A., Dutttagupta, R., Willingham, A. T., Stadler, P. F., Hertel, J., Hacker-muller, J., Hofacker, I. L., Bell, I., Cheung, E., Drenkow, J., Dumais, E., Patel, S., Helt, G., Ganesh, M., Ghosh, S., Piccolboni, A., Sementchenko, V., Tammana, H., and Gingeras, T. R. (2007) *Science*, **316**, 1484-1488.
40. Fejes-Toth, K., Sotirova, V., Sachidanandam, R., Assaf, G., Hannon, G. J., Kapranov, P., Foissac, S., Willingham, A. T., Dutttagupta, R., Dumais, E., and Gingeras, T. R. (2009) *Nature*, **457**, 1028-1032.
41. Core, L. J., Waterfall, J. J., and Lis, J. T. (2008) *Science*, **322**, 1845-1848.
42. Taft, R. J., Glazov, E. A., Cloonan, N., Simons, C., Stephen, S., Faulkner, G. J., Lassmann, T., Forrest, A. R., Grimmond, S. M., Schroder, K., Irvine, K., Arakawa, T., Nakamura, M., Kubosaki, A., Hayashida, K., Kawazu, C., Murata, M., Nishiyori, H., Fukuda, S., Kawai, J., Daub, C. O., Hume, D. A., Suzuki, H., Orlando, V., Carninci, P., Hayashizaki, Y., and Mattick, J. S. (2009) *Nat. Genet.*, **41**, 572-578.
43. Preker, P., Nielsen, J., Kammler, S., Lykke-Andersen, S., Christensen, M. S., Mapendano, C. K., Schierup, M. H., and Jensen, T. H. (2008) *Science*, **322**, 1851-1854.
44. Preker, P., Nielsen, J., Schierup, M. H., and Jensen, T. H. (2009) *Cell. Cycle*, **8**, 1106-1107.
45. Preker, P., Almvig, K., Christensen, M. S., Valen, E., Mapendano, C. K., Sandelin, A., and Jensen, T. H. (2011) *Nucleic Acids Res.*, **39**, 7179-7193.
46. The ENCODE Project Consortium (2012) *Nature*, **489**, 57-74.
47. Sanyal, A., Lajoie, B. R., Jain, G., and Dekker, J. (2012) *Nature*, **489**, 109-113.
48. Yip, K. I., Cheng, C., Bhardwaj, N., Brown, J. B., Leng, J., Kundaje, A., Rozowsky, J., Birney, E., Bickel, P., Snyder, M., and Gerstein, M. (2012) *Genome Biol.*, **13**, R48, 1-22.
49. Tan-Wong, S. M., Zaugg, J. B., Camblong, J., Xu, Z., Zhang, D. W., Mischo, H. E., Ansari, A. Z., Luscombe, N. M., Steinmetz, L. M., and Proudfoot, N. J. (2012) *Science*, **338**, 671-675.