Oct-1 promoter region contains octamer sites and TAAT motifs recognized by Oct proteins

E.V. Pankratova*, O.L. Polanovasky

Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Vavilov str. 32, Moscow, 117984, Russia

Received 9 February 1998; revised version received 3 March 1998

Abstract The 5′-upstream region (1.3 kb) of the gene encoding the POU domain transcription factor Oct-1 was cloned and sequenced. CAT reporter gene analysis of this region has detected a functionally active promoter. This region contains 24 TAAT-core sites, arranged in five clusters (four to six sites in one cluster); two octamer sites (ATGCAAAT) are located in the first and second clusters; in the second one the CCAAT-box adjacent to the octamer overlaps with the TAAT-core site. As shown by gel retardation assay, Oct-1, Oct-2, and some unknown proteins from myeloma cell line NS/0 interact with the TAAT-core sites of these clusters. The results suggest autoregulation of *Oct-1* gene expression that may also be controlled by other POU proteins, homeodomain proteins and CCAAT *trans*-action factors.

Key words: *Oct-1* gene; Octamer site; Homeospecific site; Autoregulation

1. Introduction

Oct proteins belong to the POU domain transcription factor family. The Oct-1 protein is a universal transcription factor expressed in all dividing cells [1,2]. It is involved in regulation of gene expression, and is therefore essential for proliferation, differentiation and other key cell processes. Target genes controlled by Oct-1 include those encoding snRNAs [3], histone H2B [4], interleukin (IL) 3, IL-5, granulocyte/macrophage colony stimulating factor [5], IL-8 [6], gonadotropin-releasing hormone (GnRH) [7], human thyrotropin-β (hTSHβ) [8], thyroid transcription factor 1 (TTF-1) [9], Pit-1 [10], and also the α/immediate early genes of herpes simplex virus [11]. There are two Oct-1 forms in the nucleus, soluble and insoluble; the latter forms a complex with the nuclear matrix [8]. Oct-1 binding to DNA may either stimulate or inhibit transcription. The octamer sequence ATGCAAAT is involved in high-affinity Oct-1 binding. Some regulatory elements have a modified sequence also recognized by Oct-1, for example ATGATAATGAG and TAATGARAT elements [12–14]. In vitro experiments have shown Oct-1 and Oct-2 proteins binding with many octamer-related binding sites and with TAAT-core-containing homeospecific sequences [15–19].

POU proteins share a highly conservative DNA binding domain (POU domain) which may be subdivided into two subdomains termed POU specific (POUs) and POU homeo (POUh) domain [1]. As shown by our and others' studies of Oct protein interactions with TAAT-core sites in vitro, only POUh domain binds to these sites [16–19]. However, the entire POU domain is involved in the interaction with the canonical oct site ATGCAAAT. Therefore, the affinity of Oct proteins towards the TAAT-core sites is lower than their affinity towards the canonical oct sites. Interaction of Oct-2 protein with TAAT-core sites depends strongly on the two 3′-flanking nucleotides in 5′-TAATNN-3′ [17,19]. This is true also for homeoproteins [20].

Interaction of Oct protein with other transcription factors and coregulators provides fine regulation of gene expression [21–23]. The DNA binding Oct-1 protein activity in vivo is controlled by a cell cycle-dependent kinase: phosphorylation of Ser^{385} in Oct-1 during mitosis prevents Oct-1 binding to an octamer element [24].

The human OTF-1 gene is encoded by 16 exons spanning over 150 kb [25]. Human *Oct-1* cDNA was cloned and sequenced [4], its high homology with murine *Oct-1* cDNA was shown [4].

The aim of this work was to analyze the structure and functioning of the untranscribed 5′-region of the *Oct-1* gene. This appeared important taking into account that *cis*- and *trans*-factors regulating *Oct-1* expression are not yet known. The results allowed us to suggest autoregulation of *Oct-1* gene expression.

2. Materials and methods

2.1. Genomic library: construction and screening

Enriched genomic library was constructed using a 1.5–3.0 kb fraction of PTF-02 cell DNA digested with *Eco*RI. Total DNA (about 80 μg) from the murine myeloma cell line PTF-02 was digested with *Eco*RI. The fragment of interest was detected by Southern blot hybridization (20 μg DNA-*Eco*RI) with a 176 bp *Oct-1* cDNA fragment. The DNA probe for Southern blotting was prepared from *Oct-1* cDNA isolated earlier [26]. Untranslated 176 bp cDNA fragment was obtained by *Eco*RI-*Xba*I digestion and ^{32}P-labelled by nick-translation. Only one 2.2 kb band was identified. This band was isolated from the remaining DNA (60 μg) after electrophoresis under the same conditions using phenol/chloroform extraction and then ligated with the *Eco*RI arms of λgt10 vector (Promega). The library was screened using Lambda gt10 vector kit (Promega) according to the manufacturer's instruction with the same DNA probe as for Southern blotting. Subcloning and sequencing were carried out in plasmid pUC18.

DNA was sequenced using T7 DNA polymerase (Fermentas) and universal primers.

2.2. Oligonucleotides

Double-stranded oligonucleotides containing the oct sequence or TAAT-core sequence were used as probes for electromobility shift assay. Flanking sequences of these probes were identical and contained no additional TAAT or ATGC motifs (Table 1).

2.3. Electromobility shift assay

Myeloma cell lysate (line NS/0) was prepared as described earlier [27]. Electromobility shift assay was performed as described [28] with 0.3 ng of ^{32}P-labelled oligonucleotide, 0.5 μg of poly(dGdC)po-

*Corresponding author. Fax: (7) (095) 135 14 05.
E-mail: pank@genome.eimb.rssi.ru

PII S0014-5793(98)00316-0

ly(dGdC) and some experiments with 15 ng of 'cold' oct probe. Experimental conditions were the same for all TAAT-core probes.

2.4. Cell transfection and CAT reporter gene analysis

The 1.3 kb *Pst*I-*Eco*RI 5′-upstream fragment with part of the first exon of the *Oct-1* gene (Fig. 1) was cloned on pCAT Enhancer Vector (Promega) (pCAT.*Oct-1*). Transfection was performed using transfectam reagent (Promega) in NB41A3 and HepII cells with recombinant pCAT.*Oct-1* vector and pCAT Control vector (Promega) in six cell dishes at about 60% cell density (1 μg DNA; transfectam reagent:DNA=4:1). CAT analysis was run with Promega kit as recommended by the manufacturer.

3. Results

3.1. Murine Oct-1 5′-upstream region contains multiple TAAT-core sites and oct sites

By screening the enriched genomic library we found a clone with a 2.2 kb insert. As shown by restriction analysis and sequencing, this insert contained the 1.3 kb 5′-upstream region, a 27 base first exon of the *Oct-1* gene, and a 0.87 kb region of the first intron. Two canonical oct sequences ATGCAAAT of opposite orientation were found in the 5′-region, −412 bp, −535 bp upstream from the translation start site (Fig. 1A). Three oct-related sequences were also found in this region: two with single substitutions, TTGCAAAT at −99 bp and ATCCAAAT at −822 bp and one with a nucleotide insert between the POUs and POUh recognition sites, ATGCcAAAT at −1168 bp. The region from −293 to −1261 bp contained five AT-rich clusters separated by 80–170 bp spacers. The A/T to G/C ratio in these clusters exceeds the same ratio in spacers about two times. Twenty-four homeospecific-like sites with TAAT-core sequence are located in these clusters (Fig. 1). The first cluster contained four TAAT-core sites and an oct sequence. The second cluster had six TAAT-core sites, an octamer of reversed orientation, and also two CCAAT sites of opposite orientation. The latter may be a target for CCAAT binding proteins. This cluster had two overlaps of the *cis*-elements: four crossed sequences (oct, CCAAT and two TAAT) and three overlapping TAAT-core sites. The third cluster had five TAAT-core sites and two oct-related sequences. The fourth cluster had three TAAT-core sites. One more TAAT site was found between the third and fourth clusters. The fifth cluster had five TAAT-core sites and an oct-related sequence.

3.2. The 5′-upstream region (EcoRI-PstI) of the Oct-1 gene has promoter activity

CAT reporter gene analysis of the 5′-upstream region revealed a functionally active promoter. Promoter activity of the

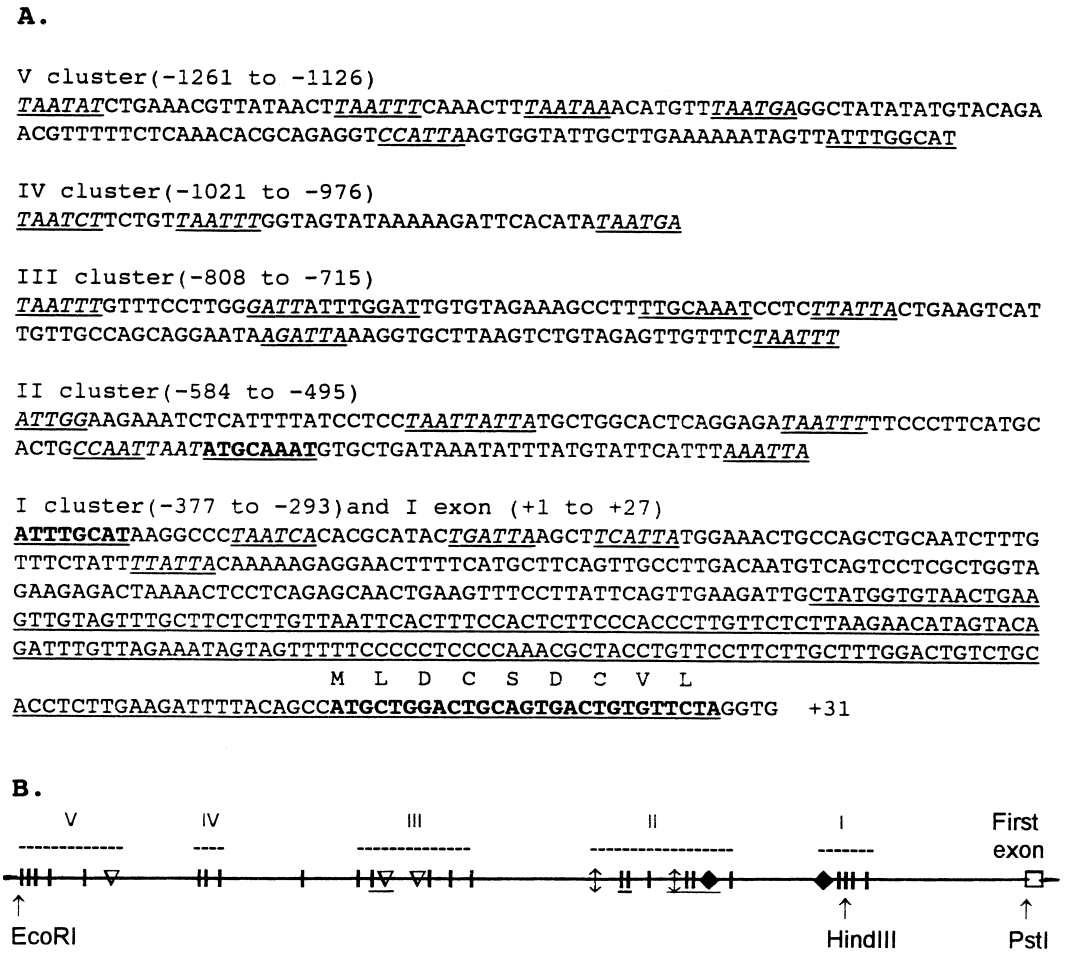

Fig. 1. Murine *Oct-1* promoter. A: Nucleotide sequences of AT-rich clusters containing TAAT-core sequences; oct sites, oct-related sites, TAAT-core sites and CCAAT sites are underlined. The amino acid sequence of the first exon is indicated. Untranslated region of cDNA and first exon are underlined. B: Schematic cluster organization showing the TAAT-core sites (|), oct (♦), oct-related sites (▽) and CCAAT sites (↕); overlapping sites are underlined.

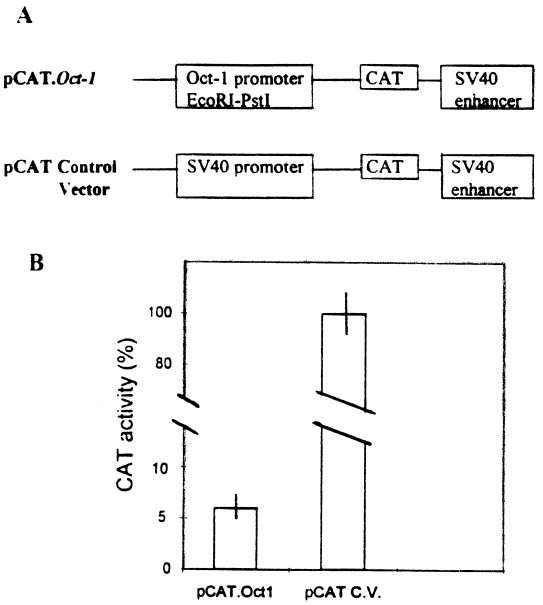

Fig. 2. CAT analysis of the *Oct-1* gene 5′-upstream region. A: Schematic map of the pCAT.*Oct-1* construct and pCAT control vector. B: Comparison of *Oct-1* promoter activity and SV40 promoter activity. Cells were harvested 48 h after transfection and equal amounts assayed for CAT activity.

Oct-1 5′-upstream region was about 20 times lower than that of the pCAT control vector containing SV40 virus promoter (Fig. 2).

3.3. *The TAAT-core sites from the 5′-upstream region of the Oct-1 gene bind Oct proteins and some other nuclear proteins*

Electromobility shift assay has shown varying affinity of Oct-1 and Oct-2 proteins from the NS/0 myeloma cell extract towards the TAAT-core sites depending on the two 3′-flanking nucleotides of the probe 5′-TAATNN-3′ (Fig. 3A,B). Some unidentified proteins also interacted with the TAAT-core probes (Fig. 3A,B). High affinity to Oct proteins was shown for the TAATTT, TAATGA, and TAATTA sites, confirming the data on the high affinity of the Oct-2 POU domain to these sites obtained earlier (Table 2) [19]. The interaction with TAATTT and TAATGA was completely inhibited by the 'cold' oct probe. Other TAAT-core sites from the 5′-upstream region (TAATCA, TAATTG, TAATTA, TAATAT, TAATCT, TAATGG) interacted with Oct proteins showing different affinity and they also bound unidentified nuclear proteins (interaction was not affected or was not completely inhibited by the 'cold' oct probe).

Table 1
Probes

Oct	5′-AGGTACCTGAG**ATGCAAAT**GAGACTGTCTCTC-TAGAG-3′
TAATTA	5′-AGGTACCTGAGTTGA**TAATTA**CTGTCTGTCTC-TAGAG-3′
TAATCT	5′-AGGTACCTGAGTTGA**TAATTC**CTGTCTGTCTC-TAGAG-3′
TAATGA	5′-AGGTACCTGAGTTGA**TAATGA**CTGTCTGTCTC-TAGAG-3′
TAATCC	5′-AGGTACCTGAGTTGA**TAATCC**CTGTCTGTCTC-TAGAG-3′
TAATAT	5′-AGGTACCTGAGTTGA**TAATAT**CTGTCTGTCTC-TAGAG-3′
TAATTT	5′-AGGTACCTGAGTTGA**TAATTT**CTGTCTGTCTC-TAGAG-3′
TAATCA	5′-AGGTACCTGAGTTGA**TAATCA**CTGTCTGTCTC-TAGAG-3′
TAATTG	5′-AGGTACCTGAGTTGA**TAATTG**CTGTCTGTCTC-TAGAG-3′
TAATGG	5′-AGGTACCTGAGTTGA**TAATGG**CTGTCTGTCTC-TAGAG-3′

4. Discussion

The data available on the structure of the promoter and enhancer regions in Oct-1-regulated genes suggest that the sites involved in regulation of gene expression by Oct-1 in vivo form three groups (Table 3): (1) the canonical oct site ATGCAAAT; (2) non-canonical oct sites with single nucleotide substitutions (oct-related sites) and (3) TAAT-core sites.

The canonical oct site ATGCAAAT binds the Oct-1 protein with the highest affinity. This sequence is found in promoter and enhancer regions of the actively transcribed Oct-1-activated genes, for example in the genes encoding H2B histone

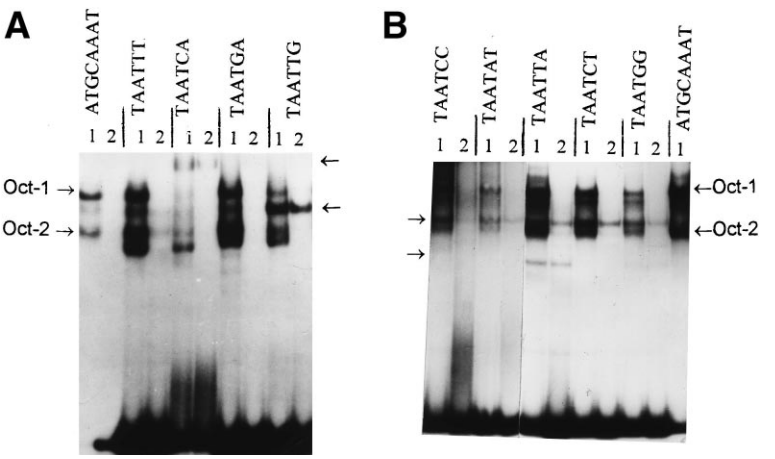

Fig. 3. Electromobility shift assay analysis of the TAAT-core elements in the murine *Oct-1* promoter. Lanes 1, 0.3 μg ^{32}P-labelled probe; lanes 2, reaction in the presence of 1.5 μg 'cold' oct probe; arrows (→) indicate complexes not affected or incompletely inhibited by the 'cold' oct probe.

[4] and U1, U2, U6, 7SK snRNAs [3,28–32]. The oct site is also present in the human Pit-1 gene promoter, however transcriptional activity of this gene is negatively regulated by Oct-1 protein [10].

The promoter and enhancer regions of many genes controlled by Oct-1 protein contain Oct-1-specific sites of lower affinity: oct-related sites and TAAT-core sites (Table 3). The TAAT-core sites in the regulatory gene regions appear important for Oct-1 protein functioning. TAAT-core sites may be involved in either gene repression or activation, or may be essential for replication and chromatin loop formation.

The Oct-1 protein activates transcription of TTF-1. The promoter of the TTF-1 gene has two functionally active TAAT-core sites, TAATTA and TAATTG, interacting with the Oct-1 protein [9]. The prolactin gene promoter has two sites binding Oct-1 and Pit-1 proteins [33]. These transcription factors appear to interact simultaneously with this region with a synergistic effect on prolactin gene transcription. The Oct-1 binding site contains a TAATCA sequence.

As shown for some genes, Oct-1 may also act as a repressor. Oct-1 strongly represses transcriptional activity of the IL-8 promoter by binding to the non-canonical oct site TTGCAAAT [6]. Oct-1 is possibly involved in hormone-dependent transcriptional repression of GnRH. Oct-1 in complex with a glucocorticoid receptor interacts with the distal promoter of the GnRH gene. The Oct-1 binding site contains a TAATGA sequence [7]. At least in some cases, the repressive function of the Oct-1 protein may be related to chromatin loop formation. It was shown that Oct-1 is involved in nuclear matrix formation. An AT-rich fragment with numerous overlapping TAATAA and TAATAT sites was found in the promoter of the hTSHβ gene. Binding of Oct-1 protein by this region is important for silencing of the hTSHβ promoter. A significant fraction of Oct-1 was shown to be associated with the nuclear matrix, suggesting a possible role of the Oct-1-hTSHβ silencer interaction in chromatin organization [8].

Oct-1 acts also as a replication factor, at least for viral genomes. In human adenoviruses the origin of replication contains sequences interacting with Oct-1. The canonical oct sequence as well as oct-related sequences and TAAT-core sequences were found in the origin of replication of human adenoviruses. The effect of Oct-1 depends on its affinity towards the binding site; an origin containing a high affinity oct site is stimulated to a higher level than a low affinity site [34,35].

Table 2
Number of copies of the TAAT-core sites in the *Oct-1* gene 5′-upstream region and relative K_d values [19] for the Oct-2 POU domain

Site	Number of copies	Rel K_d
TAATTT	6	6.4
TAATAA	5	16.4
TAATGA	3	5.8
TAATTA	1	5.6
TAATAT	2	nd
TAATCT	2	nd
TAATCA	2	10.2
TAATTG	1	nd
TAATCC	1	22.0
TAATGG	1	18.0
ATGCAAAT	2	1 max affinity

nd, not determined.

Table 3
Oct-1 recognized sites in promoter and enhancer regions of the Oct-1 regulated genes

Gene	Sequence
H-2B	CTT**ATGCAAAT**AAGGT
U2 snRNA	GCT**ATGCAAAT**AGGGT
U6 snRNA	GGC**ATGCAAAT**TCGAA
Pit-1 (human)	GTC**ATGCAAAT**CTCAC
IL-8	CAG**TTGCAAAT**CGTGG
Ad2 NFIII	AATATGA**TAATGA**GGG
ICP4 (HSV IE)	GGGCGG**TAATGA**GAT
Hox2.3α	CTCCCAG**TAATGA**GGA
Hox2.3β	TGATCAA**TAATGA**ATG
TTF-1 (BS-2)	TGTAAGCTCT
TTF-1 (BS-1)	TCCTCC**TAATTG**GCT
Prolactin	ATATATA**TAATCA**GG
GnRH(−216 to −201)	AAGATTT**TAATGA**CCT
hTSHβ	**TTAATATAATAAATAA** **TTTATTAATATTAAT**
IL-3 (−157 to −145)	GGATGAA**TAATTA**
IL-5 (−46 to −59)	AAATGAA**TAATTT**
GM-CSF (−38 to −50)	AAA**TGATTAATGG**

Oct-1 plays multiple roles in the cell, acting as a positive or negative regulator of gene transcription, DNA replication, and possibly anchoring chromatin loops to the nuclear matrix. The pleiotropic effect of Oct-1 in cells is based on the following. First, Oct-1 protein interacts with DNA as a monomer or forms heterodimers with tissue-specific transcription factors, or forms complexes with coactivators depending on the binding site. For example, the VP16 protein interacts with Oct-1 only when it binds to the TAATGARAT site [11] and a complex with Oca-B (OBF-1/Bob1) is formed only if Oct-1 or Oct-2 bind to the canonical oct site ATGCAAAT [21–23]. Second, in vivo Oct-1 acts via the DNA sites of three types, interacting with them with different affinity.

Regulation of the *Oct-1* gene expression probably depends on all three types of the Oct-1 binding elements (Tables 2 and 3). As shown in this work, the 5′-region of the *Oct-1* gene contains two canonical oct sites, three oct-related sites, and 24 TAAT-core sites, arranged in five clusters (Fig. 1). These sites vary in their affinity towards Oct-1 binding. Moreover, the TAAT-core sites may bind some other proteins, so the interaction of the TAAT-core site with nuclear protein depends on the concentration of proteins in a cell. TAAT-core sites, even if transcriptionally inactive, may bind the Oct proteins increasing the local concentration of the Oct proteins at this region of the chromosome.

The CCAAT sites may also be important in *Oct-1* gene regulation. It was shown [6] that the CCAAT enhancer binding protein (C/EBP) and Oct-1 act as antagonists for regulation of IL-8 gene transcription; Oct-1 strongly represses the transcriptional activity of the IL-8 promoter by binding independently to an element overlapping with the C/EBP binding site [6]. The second cluster of the 5′-upstream region of the *Oct-1* gene contains two CCAAT sequences of opposite orientation (Fig. 1A). One of them, four bases upstream of the oct sequence, overlaps with the TAATTG site, which interacts in NS/0 cells with an unknown protein, probably a homeoprotein (Fig. 3A). This region contains four overlapping sites: an oct site, two TAAT sites, and a CCAAT site (Fig. 1A).

This suggests a tight interaction of the Oct, homeo, and CCAAT binding proteins for regulation of *Oct-1* gene expression. The cooperativity and complex pattern of this interac-

tion are confirmed by clustering of the sites. Also we suggest *Oct-1* gene autoregulation via oct sites and TAAT-core sites. Autoregulation is involved in the control of other genes of the POU family: murine and human genes encoding Pit-1 [10,35], and the human Brn-4 gene [36].

Acknowledgements: We thank Dr. Alexander Stepchenko for supplying DNA probes. This work was supported by the Russian Human Genome Program and Russian Fund for Basic Research (Grant 96-04-494997).

References

[1] Verrijzer, C.P. and van der Vliet, P.C. (1993) Biochim. Biophys. Acta 1173, 1–21.
[2] Ryan, A.K. and Rosenfeld, M.G. (1997) Genes Dev. 11, 1207–1225.
[3] Mittal, V., Cleary, M.A., Herr, W. and Hernandez, N. (1996) Mol. Cell. Biol. 16, 1955–1965.
[4] Sturm, R.A., Das, G. and Herr, W. (1988) Genes Dev. 2, 1582–1599.
[5] Kaushansky, K., Shoemaker, S.G., O'Rork, C.A. and McCarty, J.M. (1994) J. Immunol. 152, 1812–1820.
[6] Wu, G.D., Lai, E.J., Huang, N. and Wen, X. (1997) J. Biol. Chem. 272, 2396–2403.
[7] Clark, M.E. and Mellon, P.L. (1995) Mol. Cell. Biol. 15, 6169–6177.
[8] Kim, M.K., Lesson-Wood, L.A., Weintraub, B.D. and King, J.H. (1996) Mol. Cell. Biol. 16, 4366–4377.
[9] Bingle, C.D. and Gowan, S. (1996) Biochem. J. 319, 669–674.
[10] Delhase, C., Castrillo, J., de la Hoya, M., Rajas, F. and Hooghe-Peters, E.L. (1996) J. Biol. Chem. 271, 32349–32358.
[11] Kristie, T.M., Lebowitz, J.H. and Sharp, P.A. (1989) EMBO J. 8, 4229–4238.
[12] Herr, W. and Cleary, M.A. (1995) Genes Dev. 9, 1679–1693.
[13] Baumruker, T.R., Sturm, R. and Herr, W. (1988) Genes Dev. 2, 1400–1413.
[14] Aprhis, C.M.J., Ciufo, D.M., O'Neil, E.A., Kelly, T.J. and Hayward, G.S. (1989) J. Virol. 63, 2798–2812.
[15] Stepchenko, A.G. (1992) Dokl. Akad. Nauk USSR 335, 175–179.
[16] Verrijzer, C.P., Alkema, M.J., van Weperen, W.W., van Leeuwen, H.C., Strating, M.J.J. and van der Vliet, P.C. (1992) EMBO J. 11, 4993–5003.
[17] Stepchenko, A.G. (1994) FEBS Lett. 337, 175–178.
[18] Stepchenko, A.G., Luchina, N.N. and Polanovsky, O.L. (1997) FEBS Lett. 412, 5–8.
[19] Stepchenko, A.G., Luchina, N.N. and Pankratova, E.V. (1997) Nucleic Acids Res. 25, 2847–2853.
[20] Persival-Smith, A., Muller, M., Affolter, M. and Gehring, W.J. (1990) EMBO J. 9, 3967–3974.
[21] Strubin, M., Newell, J.W. and Matthias, P. (1995) Cell 80, 497–506.
[22] Cepek, K.L., Chasman, D.I. and Sharp, P.A. (1996) Genes Dev. 10, 2079–2088.
[23] Gstaiger, M., Georgiev, O., van der Leeuven, H., van der Vliet, P.C. and Schaffner, W. (1996) EMBO J. 15, 2781–2790.
[24] Segil, N., Roberts, S.B. and Heintz, N. (1991) Science 254, 1814–1816.
[25] Sturm, R.A., Cassady, J.L., Das, G., Romo, A. and Evans, G.A. (1993) Genomics 16, 333–341.
[26] Stepchenko, A.G. (1992) Nucleic Acids Res. 20, 1419.
[27] Manley, J.L. (1984) in: Transcription and Translation. A Practical Approach (Hames, B.D. and Higgins, S.G., Eds.) pp. 89–110, IRL Press, Oxford.
[28] Carbon, P., Murgo, S., Ebel, J.P., Krol, A., Tebb, G. and Mattaj, L.W. (1987) Cell 51, 71–79.
[29] Murphy, S., Pierani, A., Scheideret, C., Melli, M. and Roeder, R.G. (1989) Cell 59, 1071–1080.
[30] Janson, L., Weller, P. and Pettersson, U.L.F. (1989) J. Mol. Biol. 205, 381–396.
[31] Das, G., Henning, D., Wright, D. and Reddy, R. (1988) EMBO J. 7, 503–512.
[32] Kunkel, G.R. and Pederson, T. (1988) Genes Dev. 2, 196–204.
[33] Voss, J.W., Wilson, L. and Rosenfeld, M.G. (1991) Genes Dev. 5, 1309–1320.
[34] Pruijn, G.J.M., van Miltenburg, R.T., Claessens, J.A.J. and van der Vliet, P.C. (1988) J. Virol. 62, 3092–3102.
[35] Rhodes, S.J., Chen, R., DiMattia, G.E., Scully, K.M., Kalla, K.A., Lin, S.-C., Yu, V.C. and Rosenfeld, M.G. (1993) Genes Dev. 7, 913–932.
[36] Malik, K.F., Kim, J., Hartman, A.L., Kim, P. and Young III, W.S. (1996) Mol. Brain Res. 38, 209–221.