Conserved gene structure and transcription factor sites in the human and mouse deoxycytidine kinase genes[1]

Magnus Johansson, Ameli Norda, Anna Karlsson*

Division of Clinical Virology, Karolinska Institute, Huddinge University Hospital, S-141 86 Stockholm, Sweden

Received 10 August 2000; revised 28 November 2000; accepted 28 November 2000

First published online 8 December 2000

Edited by Takashi Gojobori

Abstract Deoxycytidine kinase (dCK) phosphorylates several anti-cancer and anti-viral nucleoside analogs. The enzyme is predominantly expressed in lymphoid tissues regulated by an unknown mechanism. We have cloned and sequenced the 20 kbp mouse dCK gene and ≈1.7 kbp of the 5′ flanking regions of both the human and mouse dCK genes. Five major inter-species conserved motifs were identified in the 5′ region including the transcription initiation region, an SP1 site and two closely located putative octamer transcription factor sites. Luciferase reporter experiments showed that the human dCK 5′ region efficiently initiated transcription but no tissue regulatory element could be identified.

Key words: Nucleoside analog; Nucleoside kinase; Transcription regulation; Nucleoside metabolism

1. Introduction

Human deoxycytidine kinase (dCK) phosphorylates several clinically important anti-cancer and anti-viral nucleoside analogs such as 1-β-D-arabinofuranosylcytosine and 2-chloro-2′-deoxyadenosine used in treatment of hematological malignancies and the anti-HIV compounds 2′,3′-dideoxycytidine and 2′-deoxy-3′-thiacytidine [1,2]. The nucleoside analogs are inactive prodrugs that are dependent on phosphorylation for pharmacological activity. dCK is constitutively expressed throughout the cell cycle and low levels of the protein are present in most tissues investigated [3–6]. dCK is, however, predominantly expressed in lymphoid cells with particular high levels in immature T-lymphoblasts [7,8]. The lymphoblasts are highly sensitive to purine deoxyribonucleosides and nucleoside analogs phosphorylated by dCK. The high levels of dCK in these cells have been implicated as the mechanism of lymphocyte depletion caused by disorders of purine metabolism and as a mechanism of tissue targeted cytotoxicity of dCK phosphorylated anti-leukemic nucleoside analogs [8,9]. In accordance with this hypothesis, acquired resistance to nucleoside analogs has been associated with decreased dCK expression [10,11] or mutation in the coding region of the dCK gene [12,13].

Parts of the human dCK gene are cloned, and a 0.5 kbp DNA fragment located upstream of the translation start site has been investigated for promoter activity [14,15]. The dCK gene lacks a TATA-box but contains a transcription initiator region located adjacent to the major transcription start site at bp −146 relative to the start of translation. The initiation region contains an imperfect E2F binding site [14]. The human dCK promoter has a high GC content similar to several other TATA-less promoters, and there are at least two SP1 sites that regulate transcription of the gene [15]. Reporter gene studies on the human dCK promoter show higher levels of reporter gene activity in T-lymphoblast than in B-lymphoblast cell lines [14]. However, further studies have failed to show that the 0.5 kbp region regulates tissue specific transcription of the human dCK gene [15].

We decided to clone the mouse dCK gene to further study the transcription regulation of dCK. We report in the present study the DNA sequencing of the 20 kbp mouse dCK gene and ≈1.7 kbp of the 5′ flanking regions of both the human and mouse dCK genes. We have used the sequence information to identify sequence conserved motifs that may regulate transcription of the genes.

2. Materials and methods

2.1. Cloning and sequencing of human and mouse dCK genomic DNA

The open reading frame of the mouse dCK cDNA [17] was labeled with [α-^{32}P]dCTP (3000 Ci/mmol, Amersham) (Prime-A-Gene, Promega). 10^6 plaques of a mouse (strain 129SVJ) liver genomic DNA library constructed in λ-FIXII vector (Stratagene) were replicated on Colony/Plaques screen membranes (DuPont NEN). Hybridization was carried out at 42°C in 4×SSC (SSC is 150 mM NaCl, 15 mM sodium citrate pH 7)/50 mM NaH_2PO_4/5×Denhardt's solution (Denhardt's is 0.02% Ficoll/0.02% polyvinylpyrrolidone/0.02% bovine serum albumin)/1% sodium dodecyl sulfate (SDS)/10% dextran sulfate/20% formamide/50 μg/μl denatured salmon sperm DNA. Filters were washed three times at 42°C in 2×SSC/0.1% SDS for 30 min and autoradiographed for 24–72 h. Positive plaques were harvested and re-screened with the cDNA probe to isolate single λ clones.

We used the Human Promoter Finder DNA Walking kit (Clontech) to clone the 5′ flanking region of the human dCK gene. Oligonucleotide primers (5′-TGGGCGTAGTTGCTTTTAGAGGTAGCTTCCC and 5′-CGGTGGGGCGGGTTCTCTTCGAAGGGCA) were designed based on the DNA sequence of the human dCK promoter [14]. The Expand Long Template PCR system (Boehringer Mannheim) was used for the PCR as described in the Clontech protocol. All DNA sequences were determined with the Automatic Laser Fluorescence DNA sequencer (Pharmacia Biotech Inc.) or the ABI310 automated DNA sequencer (Perkin-Elmer, Applied Biosystems). Pu-

*Corresponding author. Fax: (46)-8-58587933.
E-mail: anna.karlsson@mbb.ki.se

[1] The nucleotide sequences reported in this paper have been submitted to the GenBank/EBI Data Bank with accession numbers: AF260315 and AF260316.

Abbreviations: dCK, deoxycytidine kinase

0014-5793/00/$20.00
PII: S0014-5793(00)02347-4

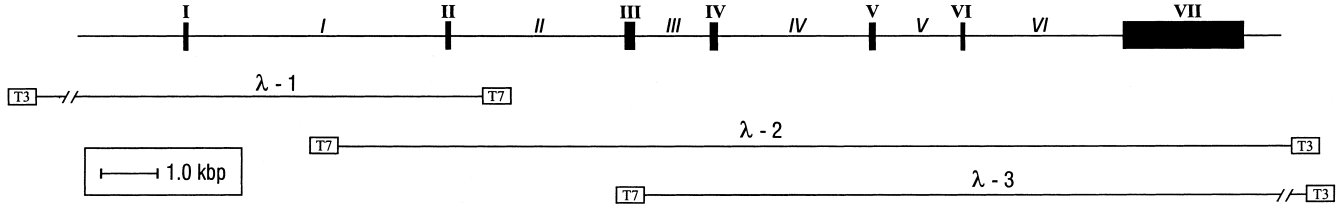

Fig. 1. Structure of the mouse dCK gene from the isolated λ clones (λ-1, λ-2, λ-3). The coding sequence is distributed in seven exons (black boxes).

tative transcription factor binding sites were identified using the TransFac data base [16].

2.2. Construction of reporter gene plasmids

An oligonucleotide primer located at bp −22 to −1 in the human dCK gene with a 5′ *Bam*HI restriction enzyme site was common to all reporter plasmids constructed (5′-CCCGGATCCTTAGTCTT-GTGGCGGCCCAGA). Oligonucleotide primers at bp −1630, −1000, −800 or −550 of the human dCK gene were designed with *Kpn*I restriction enzyme sites (5′-CTGGTACCATCCGGTTTAT-TAGTG, 5′-TTTGGTACCTATGTGCCAGGTGTCGG, 5′-ACA-GGTACCGGTATGGAGACTGGAAAGG, and 5′-GTGGGTACC-TTAAGTCTATCCAGTTCTGTCC). The PCR-amplified DNA fragments were cloned into the *Bam*HI–*Kpn*I sites of the pGL3-basic plasmid vector. The constructed plasmids were purified with the midi-prep kit (Quiagen).

2.3. Cell culture, transfection and reporter gene assays

Molt-4 and Jurkat human T-lymphoblast cells lines were gifts from Dr. J. Balzarini. The human epitheloid cervical carcinoma cell line HeLa was obtained from the American Type Culture Collection. Molt-4 and Jurkat cells were cultured in Eagle RPMI 1940 medium and HeLa cells were cultured in Dulbecco's modified Eagles medium. All cell culture medium was supplemented with 10% fetal calf serum (Gibco BRL), 100 U/ml penicillin and 0.1 mg/ml streptomycin. The cells were transfected with DMRIE-C lipofectin reagent (Gibco) according to the Gibco protocol. The cells were harvested 24 h after transfection and the luciferase activity was determined with the luciferase assay kit (Promega). β-Galactosidase activity was determined with the β-gal assay kit (Promega).

3. Results

We screened a mouse genomic DNA λ library with a mouse dCK cDNA probe. Five positive λ clones were isolated from the 10^6 plaques screened. Southern blot analysis with oligonucleotide probes derived from different parts of the mouse dCK cDNA sequence were used to confirm that the λ clones contained the mouse dCK gene (data not shown). Three of the five clones showed unique restriction patterns when the purified λ DNA was restriction-digested with *Eco*RI and *Hin*dIII (data not shown). All λ clones contained DNA inserts of ≈20 kbp. We designed oligonucleotide primers to PCR amplify the regions in the mouse dCK gene that corresponded to the intron regions of the human gene. These experiments showed that the three unique clones (λ-1, λ-2 and λ-3) covered the complete coding sequence of the mouse dCK gene as well as the 5′ and 3′ flanking regions (Fig. 1). The clones were sequenced and a complete DNA sequence of 20 250 bp was obtained, including 1750 bp located upstream of the translation start codon (GenBank accession number AF260315). The mouse dCK gene consisted of seven exons and the intron/exon junctions were located at the same positions in both the human and mouse dCK genes. The DNA sequence that corresponded to the 5′ untranslated region of the mRNA and the translation initiation codon were located in the first exon. The large 3′ untranslated region was located in exon seven.

We decided to clone and compare the 5′ flanking sequences of the human and mouse dCK genes to identify inter-species

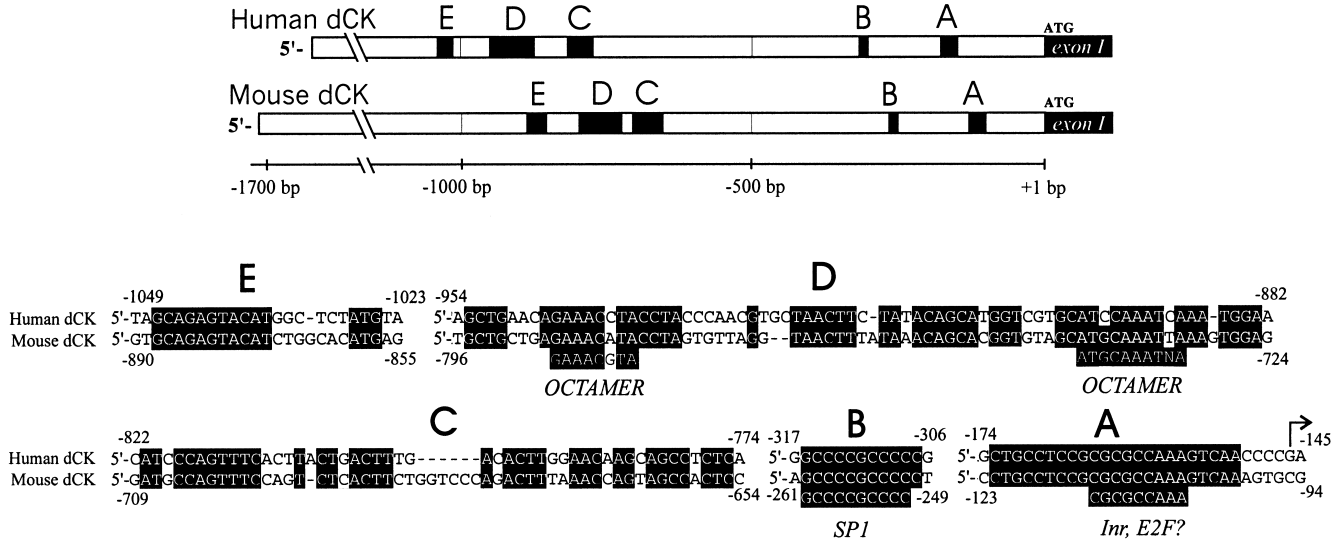

Fig. 2. Alignment of the human and mouse dCK gene 5′ flanking regions. Five major sequence conserved regions were identified (A–E). The transcription start site of the human dCK gene is at bp −146 (arrow). The numbers indicate the nucleotide positions relative to the translation start codon.

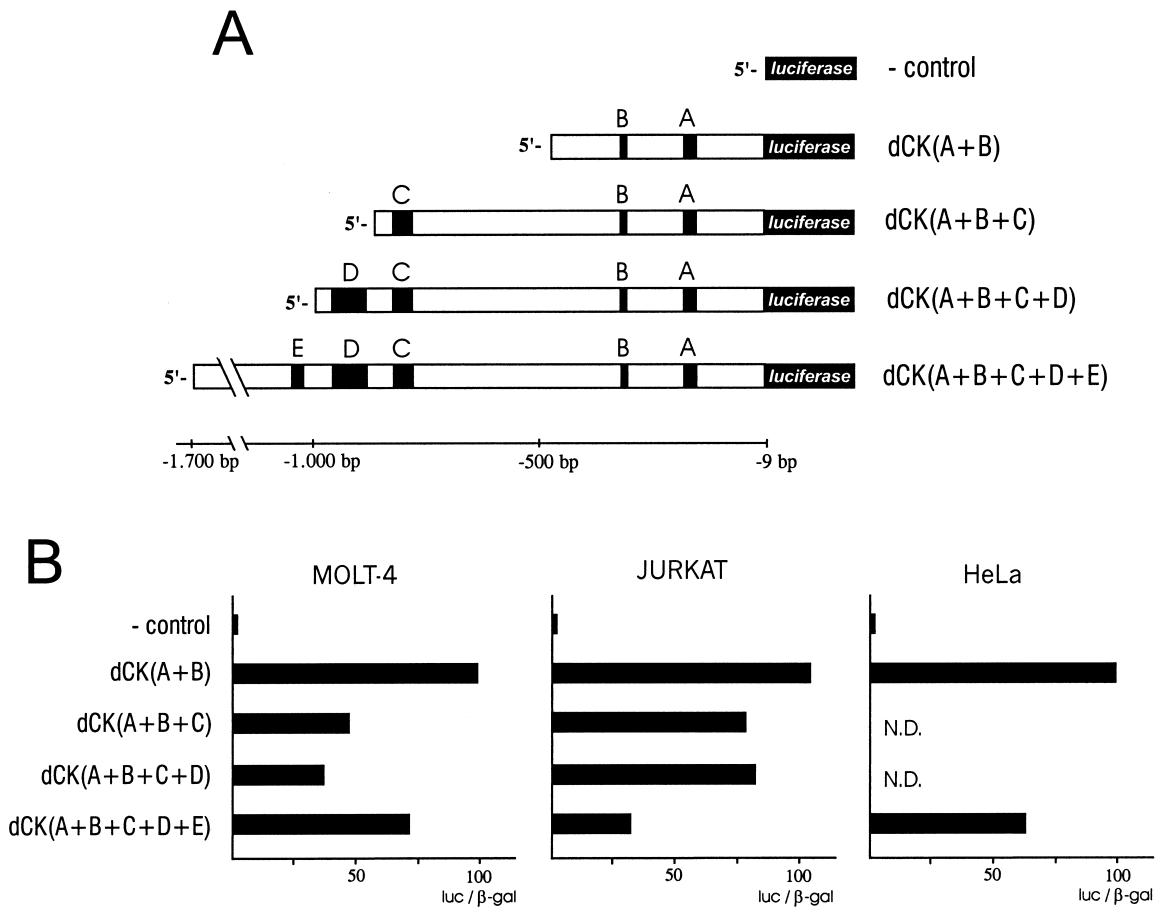

Fig. 3. Reporter gene analysis of the transcription activation by the human dCK gene 5′ region. (A) Parts of the human dCK 5′ region were cloned upstream of the luciferase reporter gene. A–E indicate the conserved sequence motifs present in both human and mouse dCK genes (see Fig. 2). (B) Assays of luciferase activity in crude extracts of human cell lines transfected with the reporter gene plasmids. The level of luminescence is shown in relation to β-galactosidase activity (luc/β-gal). N.D. not determined.

sequence conserved regions that might regulate transcription of the dCK gene. Similar to the human dCK 5′ sequence, the mouse gene had a high GC content and the gene lacked a TATA-box. The immediate 0.5 kbp region upstream of the translation initiation codon of the human dCK gene is cloned and characterized [14,15]. We cloned a larger part of the 5′ flanking sequence of the human dCK gene in order to compare the human DNA sequence with the 1.7 kbp 5′ sequence of the mouse gene. We used a PCR-based method to clone a 1.6 kbp DNA fragment 5′ of the human dCK gene (GenBank accession number AF260316). Alignment of the 5′ DNA sequences of the human and mouse dCK genes showed five major regions at best alignment that were conserved in the genes of both species (Fig. 2). Region A was located at bp −150 to −174 in the human gene and at bp −99 to −123 in the mouse gene relative to the translation start codons. The region was located adjacent to the transcription start of the human dCK gene and the partial binding site for E2F identified in the human gene was also present in the mouse dCK gene. Northern blot analysis of mouse dCK mRNA shows that the mRNA is 3.4 kbp whereas the translated region of the cloned cDNA and the 3′ untranslated region are 2.8 kbp [17]. Transcription initiation of the mouse dCK gene occurs accordingly within 0.6 kbp upstream of the translation start, but the exact position cannot be predicted because the length of the 3′ polyadenylated tail of the mRNA is not known. However, these data are in accordance with the hypothesis that transcription of the human and mouse dCK is initiated at region A.

Region B corresponded to a GC-rich cluster that binds SP1 in the human dCK gene, and reporter gene experiments suggest that the site is important for transcriptional activity of the human gene [15]. Regions C, D, E are all located upstream of the previously characterized human dCK promoter and there is no information available about their importance for transcription of the dCK gene. We searched the TransFac data base to identify transcription factor binding sites in the inter-species conserved regions C, D, and E [16]. The best matches found were two octamer transcription factor binding sites in the region D (Fig. 2).

We used luciferase gene reporter assays to study the transcription activating properties of regions C, D and E identified in the human and mouse dCK genes. Parts of the human dCK 5′ flanking region were inserted upstream of the luciferase reporter gene (Fig. 3A). The plasmids were transiently transfected into HeLa epithelium carcinoma cells, T-lymphoblast Molt-4, and Jurkat cells. A plasmid that encoded the β-galactosidase gene was used as a control to monitor the transfection efficiency. The plasmids that contained the previously characterized proximal 0.5 kbp fragment of the human dCK gene (dCK (A+B), Fig. 3A) increased luciferase activity ≈80-fold as compared to the negative control (Fig. 3B). Addition

of regions C, D or E to the reporter gene construct did not significantly alter the levels of reporter gene activity as compared to the plasmid that contained only regions A and B in the investigated cell lines.

4. Discussion

The expression of dCK in normal and malignant cells has been studied by several groups [1]. Predominant expression of dCK in lymphoid cells and tissues has been shown both by enzymatic assays [3,6,7] and by analysis of the levels of dCK protein with anti-dCK antibodies [5]. Steady-state levels of dCK mRNA are high in lymphoid tissues, but high levels of dCK mRNA are also present in muscle and brain [17,18]. This is in contrast to studies on crude protein extract of human muscle and brain that contain very low levels of dCK [5,6]. The reason for the discrepancy is not known, but it has been suggested that expression of dCK may be regulated at the post-transcriptional level in certain cell types [19]. It is presently not known whether the lymphoblast predominant expression of dCK is regulated at the transcriptional or post-transcriptional level and further studies will be required to clarify this issue.

We have extended previous studies on the transcriptional regulation of the dCK gene by cloning 1.7 kbp and 1.6 kbp regions upstream of the translation start of both the human and mouse dCK genes. Comparison of the DNA sequences showed that a sequence motif located adjacent to the transcription start of the human dCK genes was conserved in the mouse dCK gene. The initiation region binds the E2F transcription factor in vitro [15]. Members of the E2F family may initiate transcription in other TATA-less promoters such as the dihydrofolate reductase promoter [20]. However, E2F is involved in cell cycle regulation and many of its target genes are differentially expressed in the phases of the cell cycle [20]. The mechanism of E2F-mediated transcription initiation in cell cycle constitutively expressed genes, such as dCK, is not known. It is possible that an E2F-like factor is constitutively expressed and functions as an initiator, similar to the constitutively expressed TATA binding proteins. In addition to the initiation region, we identified a SP1 binding site that was present in the dCK genes of both species. SP1 sites are common in genes with TATA-less promoters and SP1 has been suggested to activate transcription in cooperation with the initiation protein [20]. There is an E-box motif adjacent to the SP1 site in the human dCK gene [14]. However, we were not able to identify a similar motif in the mouse dCK gene and its physiological importance for regulation of dCK transcription remains unclear. Among the other three regions that were conserved in the 5′ sequences of the human and mouse dCK genes, one region contained two closely located imperfect binding sites for octamer transcription factors. Octamer motifs regulate expression of B-cell specific genes such as the immunoglobulins [21]. The octamer binding transcription factors oct-1 and oct-2 are however present in most hemopoietic and lymphoid tissues [22] and the octamer transcription factors are involved in transcription regulation of T-lymphocyte specific genes [23]. Closely located octamer sites, such as those found in the dCK gene, allow cooperative binding of oct-2 proteins [24]. It is tempting to speculate that the octamer motifs in the dCK 5′ region participate in the regulation of dCK transcription. There is, however, so far no experimental evidence that either regions C, D, or E are important for the expression of the dCK gene.

The physiological role of the constitutively expressed deoxyribonucleoside kinases is suggested to be in providing deoxyribonucleotides for DNA repair and replication of mitochondrial DNA in resting and terminally differentiated G0 cells, when the de novo pathway of deoxyribonucleotide synthesis is not active. This model does, however, not explain why dCK is predominantly expressed in lymphoid tissues or why very low levels of the enzyme are present in tissues that mainly consist of resting or terminally differentiated cells. The cloning and sequencing of the mouse dCK gene reported in this paper will be the basis for future studies on the physiological importance of dCK by gene targeting experiments in transgenic mice.

Acknowledgements: This work was supported by grants from the Swedish Medical Research Council, the Swedish Cancer Foundation, the Swedish Foundation for Strategic Research, and the Medical Faculty of the Karolinska Institute.

References

[1] Arnér, E.S.J. (1995) Pharmacol. Ther. 67, 155–186.
[2] Plunkett, W. and Ghandi, V. (1996) Hematol. Cell Ther. 38, S67–S74.
[3] Durham, J.P. and Ives, D.H. (1969) Mol. Pharmacol. 5, 358–375.
[4] Arnér, E.S.J., Flygar, M., Bohman, C., Wallström, B. and Eriksson, S. (1988) Exp. Cell Res. 178, 335–342.
[5] Kawasaki, H., Carrera, C.J. and Carson, D.A. (1992) Anal. Biochem. 207, 193–196.
[6] Spasokoukotskaja, T., Arnér, E.S.J., Brosjö, O., Gunvén, P., Juliussion, G., Liliemark, J. and Eriksson, S. (1995) Eur. J. Cancer 31A, 202–208.
[7] Arnér, E.S.J., Spasokoukotskaja, T. and Eriksson, S. (1992) Biochem. Biophys. Res. Commun. 188, 712–718.
[8] Cohen, A., Lee, J.W.W., Dosch, H.-M. and Gelfand, E.W. (1980) J. Immunol. 125, 1578–1582.
[9] Carson, D.A., Kaye, J. and Seegmiller, J.E. (1977) Proc. Natl. Acad. Sci. USA 74, 5677–5681.
[10] Kobayashi, T., Kakihara, T., Uchiyama, M., Fukuda, T., Kishi, K. and Shibata, A. (1994) Leuk. Lymphoma 15, 503–505.
[11] Kakihara, T., Fukuda, T., Tanaka, A., Emura, I., Kishi, K., Asami, K. and Uchiyama, M. (1998) Leuk. Lymphoma 31, 405–409.
[12] Owens, J.K., Shewach, D.S., Ullman, B. and Mitchell, B.S. (1992) Cancer Res. 52, 2389–2393.
[13] Flasshove, M., Strumberg, D., Ayscue, L., Mitchell, B.S., Tirier, C., Heit, W., Seeber, S. and Schutte, J. (1994) Leukemia 8, 780–785.
[14] Song, J.J., Walker, S., Chen, E., Johnson II, E.E., Spychala, J., Gribbin, T. and Mitchell, B.S. (1993) Proc. Natl. Acad. Sci. USA 90, 431–434.
[15] Chen, E.H., Johnson II, E.E., Vetter, S.M. and Mitchell, B.S. (1995) J. Clin. Invest. 95, 1660–1668.
[16] Wingender, E., Kel, A.E., Kel, O.V., Karas, H., Heinemeyer, T., Dietze, P., Knuepple, R., Romaschenko, A.G. and Kolchanov, N.A. (1997) Nucleic Acids Res. 25, 265–268.
[17] Karlsson, A., Johansson, M. and Eriksson, S. (1994) J. Biol. Chem. 269, 24374–24378.
[18] Johansson, M. and Karlsson, A. (1996) Proc. Natl. Acad. Sci. USA 93, 7258–7262.
[19] Hengstschläger, M., Denk, C. and Wawra, E. (1993) FEBS Lett. 321, 237–240.
[20] Azizkhan, J.C., Jensen, D.E., Pierce, A.J. and Wade, M. (1993) Crit. Rev. Eukaryot. Gene Exp. 3, 229–254.
[21] Wirth, T., Pfisterer, P., Annweiler, A., Zwilling, S. and König, H. (1995) Immunobiology 193, 161–170.
[22] Cockerill, P.N. and Klinken, S.P. (1990) Mol. Cell. Biol. 10, 1293–1296.
[23] Graef, I.A. and Crabtree, G.R. (1997) Science 277, 193–194.
[24] LeBowitz, J.H., Clerc, R.G., Brenowitz, M. and Sharp, P.A. (1989) Genes Dev. 3, 1625–1638.