The EMBO Journal vol.10 no.9 pp.2595 – 2603, 1991

# Identification of transcription factors required for the expression of mammalian U6 genes *in vitro*

**Rainer Waldschmidt, Inawati Wanandi and Klaus H.Seifart**

Institut für Molekularbiologie und Tumorforschung, Karl-von-Frisch-Strasse, D-3550 Marburg, FRG

Communicated by I.W.Mattaj

**Transcription factors, required for the basal expression of the mouse U6 gene were identified in extracts from HeLa cells. This gene is transcribed at least four times more efficiently than its human counterpart in extracts from mouse or HeLa cells and hence provides an excellent *in vitro* system for the identification of transcription factors involved in the basal expression of mammalian U6 genes. At least four separate protein components were found to be required in addition to RNA polymerase III for correct synthesis of U6 RNA *in vitro*. These correspond to: (i) TFIIIB; (ii) a heat labile activity contained in a protein fraction enriched in TFIID; (iii) an, as yet, uncharacterized component contained in the flow-through upon rechromatography on phosphocellulose, and finally; (iv) a protein specifically binding to the mouse U6 gene promoter and transactivating its expression. Transcription factors IIIA and IIIC are not involved in mammalian U6 transcription *in vitro*. The U6-specific transcription factor has a molecular mass of ~90 ± 10 kDa. It specifically binds to the U6 gene from bp −42 to −78 on the coding and from bp −37 to −79 on the non-coding strand thereby centrally encompassing the PSE motif of the mouse U6 promoter. The binding activity of this protein is correlated with the efficiency with which the U6 gene is transcribed *in vitro*, thereby indicating a crucial role of the PSE-binding protein for U6 transcription.**

*Key words: in vitro* transcription/mammalian U6 RNA/RNA polymerase III/TFIIIB/U6 specific transcription factors

## Introduction

U6 small nuclear RNA (snRNA) is an essential component required in conjunction with other UsnRNA species for spliceosome mediated processing of pre-messenger RNA (for review see Guthrie and Patterson, 1988). Whereas U6RNA is the only snRNA transcribed by RNA polymerase III (pol III; for review see Kunkel, 1991; Kunkel *et al.*, 1986), the remaining UsnRNAs are thought to be synthesized by RNA polymerase II (pol II; Dahlberg and Lund, 1988). Specific expression in all cases is governed by the complex interaction of different *cis*-regulatory sequence elements with *trans*-acting transcription factors (TF). While gene expression by pol I and pol II is predominantly controlled from 5′ flanking regions (for review see Myers *et al.*, 1986), pol III genes are classically controlled by intragenic, discontinuous promoter sequences (for review see Geiduschek and Tocchini-Valentini, 1988) of the AC-(ribosomal 5S RNA) or AB-type (e.g. tRNA, 7 SL, virus-associated VA RNA, and EBER genes). These intragenic control regions (ICR) represent the primary binding sites for TFIIIA (Engelke *et al.*, 1980; Seifart *et al.*, 1989) and/or TFIIIC (Geiduschek and Tocchini-Valentini, 1988; Schneider *et al.*, 1989) which are considered to represent assembly factors for the formation of pol III preinitiation complexes. The subsequent incorporation of transcription factor IIIB (Waldschmidt *et al.*, 1988), thought to be the initiation factor proper for all known pol III genes (Kassavetis *et al.*, 1990) into these preinitiation complexes, generates stable transcription complexes recognized by pol III.

In contrast to the aforementioned 'classical' pol III genes, U6 snRNA genes from vertebrate species lack an apparent ICR and require no intragenic sequences for transcription *in vivo* and *in vitro* (Das *et al.*, 1988; Lobo and Hernandez, 1989; Kunkel and Pederson, 1988; Mattaj *et al.*, 1988). All vertebrate U6 snRNA genes so far examined contain three conserved and well characterized *cis*-acting upstream sequence elements essential for U6 promoter function: the distant sequence element (DSE, responsible for the binding of transcription factor Oct 1; Bark *et al.*, 1987; Carbon *et al.*, 1987; Das *et al.*, 1988; Kunkel and Pederson, 1988), the proximal sequence element (PSE; Carbon *et al.*, 1987; Kunkel and Pederson, 1988; Mattaj *et al.*, 1988; Lobo and Hernandez, 1989) near position −60 and an AT rich pol II TATA-like sequence element near position −30 (Mattai *et al.*, 1988; Lobo and Hernandez, 1989). The DSE and PSE were shown to be functionally interchangeable with comparable sequences upstream of the U2 snRNA gene transcribed by RNA polymerase II (Bark *et al.*, 1987; Kunkel and Pederson, 1988; Mattaj *et al.*, 1988), indicating a common function of these sequence elements in the expression of UsnRNA genes by either RNA polymerase II or III. The TATA-like sequence element, unique to U6 snRNA genes, could be shown to specifically direct pol III to this gene (Mattaj, *et al.*, 1988; Lobo and Hernandez, 1989).

Expression of vertebrate U6 genes *in vivo* (Kunkel and Pederson, 1988; Mattaj *et al.*, 1988; Lobo and Hernandez, 1989) and to a lesser extent *in vitro* (Das *et al.*, 1988) requires both the TATA-like and the PSE element, implicating that the 'basal' U6 RNA promoter in vertebrates consists of both sequence elements.

In contrast to the extensively and well characterized *cis*-regulatory promoter elements, less is known about the *trans*-acting protein components involved in the basal transcription of vertebrate U6 genes by RNA polymerase III. As was to be expected, the 5S rRNA specific transcription factor TFIIIA is dispensable for mouse U6 RNA expression *in vitro* (Reddy, 1988; and data reported in this paper). The involvement of other pol III transcription factors could not yet be unequivocally resolved. Experiments addressed at the requirement of TFIIIC in the expression of *Saccharomyces*

© Oxford University Press

*cerevisiae* U6 RNA led to contradictory results depending on the 3′ flanking sequence of the U6 snRNA gene used (Brow and Guthrie, 1990; Moenne *et al.*, 1990). Partially purified TFIIIB was reported to be required for yeast U6 transcription in the absence of TFIIIC (Moenne *et al.*, 1990) but direct involvement of TFIIIB in vertebrate U6 gene expression has not yet been demonstrated. However, the central role of TFIIIB, as the transcription initiation factor proper for RNA polymerase III, makes it a very likely candidate possibly interacting indirectly with the U6 gene TATA-like element through other transcription factor(s) binding to it. Transcription factors directly binding to the 'basal' promoter element of vertebrate U6 genes also remain to be characterized and only a partially purified protein fraction necessary for U6 transcription has previously been reported in this connection (Reddy, 1988).

In this paper we report for the first time the identification and partial purification of a transcription factor, specifically binding to the PSE motif of the mouse U6 gene and transactivating its expression *in vitro*. The involvement of transcription factors IIIA, IIIB and IIIC in U6 RNA *in vitro* expression by RNA polymerase III was tested and only the participation of TFIIIB could be demonstrated. Furthermore, an as yet unidentified component(s), contained in the flow-through from phosphocellulose, was found to be essential. Finally, at least one other heat labile factor, contained in a fraction enriched in TFIID and not involved in the expression of AB- or AC-type pol III genes, could be identified to be necessary for U6 RNA transcription *in vitro*.

## Results

### *Transcription of mammalian U6 genes in vitro*

Mouse and human U6 genes were transcribed in a HeLa cell extract (S100) in comparison to various classical pol III genes, of the AB- or AC-type. The results in Figure 1 demonstrate that both mammalian U6 genes are faithfully transcribed, albeit with an at least 10-fold lower efficiency than VA, 5S or tRNA genes (lanes 1–3). Reactions were conducted at DNA concentrations optimized in pilot experiments for U6 templates and the fidelity of the U6 transcription was verified by primer extension analysis (data not shown). The significant points are: (i) that the mouse U6 gene (lane 6) is transcribed *in vitro* much more efficiently than the human gene (lane 4), and (ii) that deletion of the DSE from the human gene (compare lanes 4 and 5) depresses the corresponding U6 transcription. Interestingly this effect could not be observed for the mouse DSE, in which case little if any reduction was observed upon its deletion (compare lanes 6 and 7; Das *et al.*, 1988).

In order to show that the more efficient transcription of the mouse gene was not due to conceivable artefacts related to the heterologous extract employed, the mouse and human U6 genes as well as their DSE deletion mutants were also transcribed in a mouse extract. As was observed for the HeLa cell extract, the results (Figure 1, lanes 8–11) show that the mouse U6 gene is also transcribed more productively than the human U6 gene in the homologous extract, thus proving that this property is inherently related to the mouse U6 promoter and does not depend on the source of the mammalian extract used. Because of its at least four-fold better transcription *in vitro*, we employed the mouse instead of the human U6 gene for subsequent experiments aimed at the identification of factors essential for U6 transcription *in vitro*.

Experiments with α-amanitin (Figure 2, lanes 2–4) and tagetitoxin (lanes 5–7; Steinberg *et al.*, 1990) indicated that *in vitro* synthesis of U6 and tRNA is catalyzed by RNA polymerase III, although differential sensitivities against these inhibitors were observed in both reactions. 50 μg/ml α-amanitin slightly depress U6 transcription without markedly influencing tRNA synthesis (lane 3). Conversely 1.5 μM tagetitoxin inhibit tRNA synthesis to a greater extent than U6 expression (lane 6), suggesting that the composition of transcription complexes or RNA polymerase III acting on these two genes may be different.

### *Identification of components required for U6 transcription*

Fractionation of crude cell extracts by ion exchange and affinity chromatography show that at least three protein fractions are required for transcription of the U6 gene. As shown in lane 15, these include phosphocellulose fractions AA, B (containing among other proteins TFIIIB and RNA polymerase III, Waldschmidt *et al.*, 1988) and C (containing among other proteins TFIIIC and RNA polymerase III, Schneider *et al.*, 1989). The latter fraction was further purified by chromatography on DEAE–Sephadex. The $DS_{0.25}$ fraction is required in conjunction with phosphocellulose fractions AA and B for transcription (lane 17) and a protein contained in it binds to and can selectively be eluted with 0.5 M KCl from a DNA-affinity column containing the PSE sequence (lane 21). Fractionation of $DS_{0.25}$ by chromatography on FPLC Mono-Q leads to a comparable purification of the PSE-binding activity in the Mono $Q_{0.30}$ fraction which is also transcriptionally active in conjunction with fractions AA and B (lane, 19). Phosphocellulose fraction B contains two components required for U6 transcription as will be shown subsequently.

### *Neither TFIIIA nor TFIIIC but TFIIIB in conjunction with a fraction enriched in TFIID is required for transcription of the mouse U6 gene in vitro*

Although not directly shown, it was previously concluded on the basis of experiments in which metal cations were depleted by chelation with EDTA that TFIIIA was not

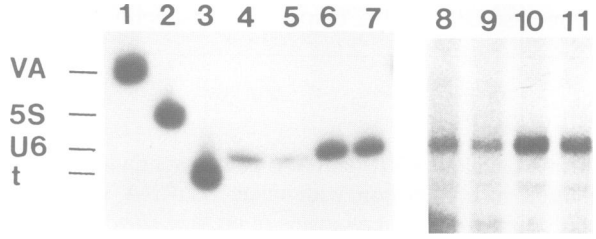


**Fig. 1.** Comparative transcription of human and mouse U6 templates in relation to classical pol III genes in a HeLa cell extract. The following gene constructs were employed at the indicated DNA concentrations, optimal for human and mouse U6 RNA *in vitro* transcription. Incubation with 20 μl HeLa (lanes 1–7) or mouse F9 (lanes 8–11) S100 was as described in Materials and methods. Autoradiography was at −80°C for 36 h (lanes 1–7) or 96 h (lanes 8–11) with intensifying screen. Lane 1: $VA_I$ gene (0.2 μg, pUVAI); lane 2: 5S gene (0.2 μg, pUh5S); lane 3: tRNA gene (0.2 μg, $pUht^{Met}$); lane 4 and 8: human wild-type U6 gene (3 μg, pUhU6); lane 5 and 9: human U6 gene, lacking the DSE (3 μg, $pUhU6_{0.35}$); lane 6 and 10: mouse wild-type U6 gene (3 μg, pUmU6); lane 7 and 11: mouse U6 gene, lacking the DSE (3 μg, $pUmU6_{0.34}$).

involved in the transcription of the mouse U6 gene (Reddy, 1988). This conclusion is supported by experiments in Figure 2, showing that U6 transcription could be reconstituted by phosphocellulose fractions AA, B and C, (lane 15) none of which contain hTFIIIA, since this protein elutes with 1 M KCl upon rechromatography on phosphocellulose (fraction AD; Seifart *et al.*, 1989).

Human TFIIIC is likewise not required for transcription of the mouse U6 promoter. Fraction PSE 0.5, obtained by affinity chromatography of phosphocellulose fraction C, while still capable of reconstituting U6 transcription (Figure 2, lane 21), does not support tRNA synthesis in conjunction with phosphocellulose fraction B (data not shown). More importantly, transcription of the U6 gene is not competed by a double-stranded synthetic DNA oligonucleotide containing the B-box sequence of the $VA_I$ gene (Figure 3A, lane 3) to which TFIIIC binds in a primary fashion (Schneider *et al.*, 1990). The same oligonucleotide completely eliminates transcription of a tRNA or $VA_I$ RNA gene (Figure 3A lane 7, Schneider *et al.*, 1990). Conversely, the PSE sequence largely eliminates U6 transcription (Figure 3A; lanes 2 and 4) but leaves unaffected the synthesis of tRNA (lane 6). Neither U6 nor tRNA transcription is affected by an irrelevant oligonucleotide (Figure 3A; lanes 5 and 8). These results show unequivocally that TFIIIC is not involved in the transcription of the mouse U6 gene *in vitro*.

To study the possible participation of hTFIIIB in the expression of mammalian U6 RNA a fraction containing hTFIIIB (PC-B), which was able to fully reconstitute tRNA (Figure 3B, lane 1) and U6 (lane 6) *in vitro* transcription, was inactivated by heat treatment for 20 min at 47°C. This fraction was designated as fraction B° and the complete loss of contained hTFIIIB activity was demonstrated by the failure to support tRNA transcription (Figure 3B, lane 2). By the readdition of purified hTFIIIB (Cibacron Blue, 2 M KCl/ 5 M urea fraction; Waldschmidt *et al.*, 1988) tRNA transcription could successfully be reconstituted (Figure 3B, lane 3). Addition of a protein fraction enriched in TFIID did not reconstitute tRNA synthesis (lane 4) indicating that it neither contained any detectable hTFIIIB activity nor a potential inhibitor of transcription (lane 5). When the same experiment was performed with the U6 template (Figure 3B, lanes 6–10) a very surprising effect was observed. As was the case for tRNA transcription, heating of the PC-B fraction for 20 min at 47°C completely inactivated U6 transcription (lane 7). This activity could not be restored by the addition of purified hTFIIIB alone (lane 8), thus contrasting the observation for tRNA synthesis (lane 3). U6 expression could also not be reconstituted by the mere addition of a fraction containing TFIID (lane 9). This latter fraction was shown to fully reconstitute the pol II-mediated expression of the histone H5 gene by a HeLa whole-cell extract (data not shown) in which TFIID activity was specifically depleted by heat treatment of the extract for 20 min at 47°C as previously described (Nakajima *et al.*, 1988). However, the joint addition of purified hTFIIIB and the fraction containing TFIID led to transcription of the U6 gene, albeit with a reduced efficiency (compare U6 signals in lanes 6 and 10). In agreement with and extension of results obtained in one of the two published yeast systems (Margottin *et al.*, 1991), these data indicate that apart from pol III and TFIIIB, at least one additional heat labile component contained in fraction IID is essentially required in conjunction with a protein binding to the PSE sequence for expression of the mouse U6 gene *in vitro*.

### *DNA-binding studies employing the PSE sequence*

Transcriptionally active protein fractions were subsequently employed in DNA-binding studies. Electrophoretic mobility shift experiments (Figure 4) show that a protein in fraction $DS_{0.25}$ (and to a comparable extent in fractions $PSE_{0.5}$ and $MQ_{0.3}$; data not shown) specifically forms a complex with a DNA fragment (*Eco*RI–*Hin*dIII fragment from pUmU6–0.155), which contains the transcriptionally active upstream promoter of the mouse U6 gene (Figure 4, lane 1) including the PSE and TATA sequence. Identical results were obtained with a fragment containing only the PSE motif (*Eco*RI–*Hin*dIII fragment from pUmU6 PSE, data not shown). This DNA–protein complex is specific for the PSE sequence since competition with plasmid DNA containing either the mouse U6 promoter (lane 2) or only the PSE (lane 3), completely abolishes formation of the complex whereas DNA containing other pol III genes (5S, tRNA, VA RNA; lanes 4–6) or a *bona fide* pol II-TATA box (Ad2MLP; lane 7) does not serve as a competitor.

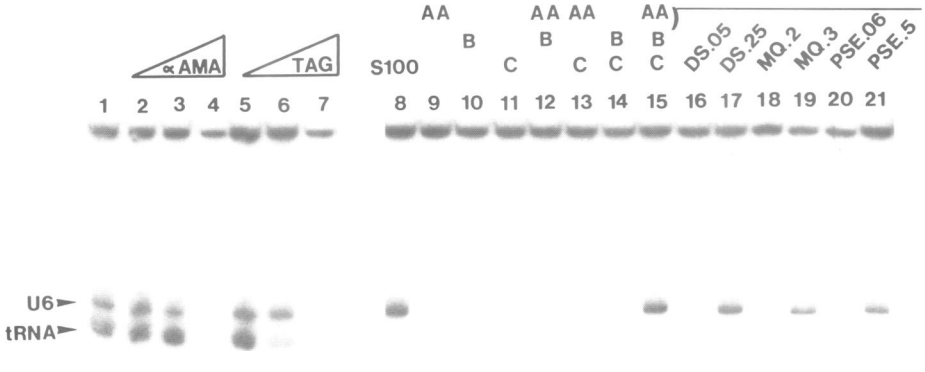


**Fig. 2.** *In vitro* transcription of the mouse U6 gene by RNA polymerase III and additional protein fractions from HeLa cell extracts. Lanes 1–7: U6 DNA ($pUmU6_{0.34}$; 3 μg) and a human tRNA gene, serving as a control ($pUht^{Met}_i$; 0.2 μg), were simultaneously transcribed by 30 μl of HeLa extract (S100) in a reaction volume of 50 μl as described in Materials and methods. Transcription was inhibited either by 1 (lane 2), 50 (lane 3) and 200 μg/ml (lane 4) α-amanitin or by 0.15 (lane 5) 1.5 (lane 6) or 15 μM (lane 7) tagetitoxin. Lanes 8–21: The template $pUmU6_{0.34}$ (3 μg) was transcribed as appropriately indicated by reconstitution of various protein fractions (25 μl each) obtained from HeLa S100 as described in Materials and methods.

The exact sequence to which the transactivating protein contained in fraction $DS_{0.25}$ binds was analyzed in greater detail by DNase I protection experiments described in Figure 5. The results show that the sequence from −42 to −78 of the coding strand of the mouse U6 gene, centrally encompassing the PSE sequence from −49 to −68, is completely protected by increasing amounts of the PSE-binding protein (Figure 5A, lanes 2−7). A hypersensitive site, appearing downstream of the TATA box in the polylinker region, was associated with a limited DNase I digestion caused by high concentrations of protein (Figure 5A, lanes 1−7) or salt (Figure 5B, lanes 1−6). The footprint is abolished upon competition with unlabeled DNA containing the PSE sequence ($pUmU6_{0.155}$, $pUmU6_{PSE}$; lanes 8 and 9) but neither by DNA containing other pol III genes (5S, $VA_I$; lanes 10 and 11), nor the mouse U6 RNA coding sequence or pUC18 vector DNA (data not shown). The noncoding strand of the mouse U6 gene was footprinted likewise (data not shown) and the results from these experiments are summarized in the lower part of Figure 5. They clearly show that the PSE-binding protein protects a region which centrally encompasses the PSE sequence (boxed part in the lower part of Figure 5) and extends 10−11 bp toward its 5′ and 7−12 bp to its 3′ flanking sequence respectively, depending on whether the footprint is analyzed on the coding or non coding strand of the gene (solid bars in the lower part of Figure 5). The footprint shown delineates a single binding site within the 5′ flanking sequence of the mouse U6 gene. Data from methylation inference assays reveal that the two G-residues at positions −63 and −64 on the non-coding strand of the mouse U6 gene (indicated as G° in the lower part of Figure 5) play a particular role for binding of the PBP.

As demonstrated in Figure 5B the footprint of the PSE-binding protein is stable to dissociation to at least 240 mM KCl, indicating a comparatively tight binding of the protein,

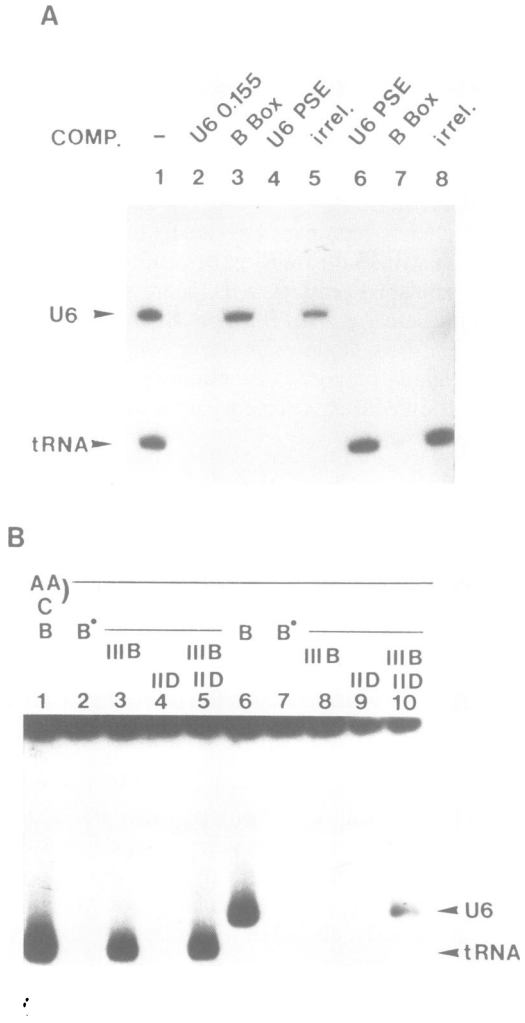


**Fig. 3. Panel A**. Competition analyses show that TFIIIC is not required for U6 transcription. Lanes 1−8 contained 10 μl of HeLa S100 and were preincubated for 30 min at 30°C with 500 ng of $U6_{0.155}$ *Eco*RI−*Hin*dIII fragment (lane 2), or one of the following synthetic double-stranded DNA oligomer (250 ng each): pol III B-box sequence (lanes 3 and 7), the mouse U6 PSE sequence (lanes 4 and 6) or an irrelevant oligonucleotide (lanes 5 and 8). Reactions were then incubated with appropriate template DNAs (lanes 1−5: 3 μg $pUmU6_{0.34}$; lanes 1, 6−8: 0.2 μg $pUht^{met}$) and further processed as described in Materials and methods. **Panel B**. TFIIIB is required in conjunction with a fraction enriched in TFIID for U6 transcription. Lanes 1−10 contained 25 μl each of phosphocellulose fractions AA and C as appropriately indicated. Lanes 1 and 6 additionally contained 25 μl of phosphocellulose fraction B. To identify heat labile components in this fraction, proteins were incubated for 20 min at 47°C, centrifuged and the supernatant adjusted to 5 mM DTT as described (Nakajima *et al.*, 1988). This fraction was designated as B° and was either tested alone (lanes 2 and 7), with 25 μl purified TFIIIB (IIIB; fraction $CB_{2/5}$, 0.085 mg/ml, Waldschmidt *et al.*, 1988; lanes 3 and 8), with 25 μl of a fraction enriched in TFIID (IID; fraction eluted from phosphocellulose with 1 M KCl, lanes 4 and 9) or in conjunction with both IIIB and IID (lanes 5 and 10). Reactions either contained 3 μg $pUmU6_{0.34}$ (lanes 6−10) or 0.2 μg $pUht^{met}$ (lanes 1−5) as template DNA.

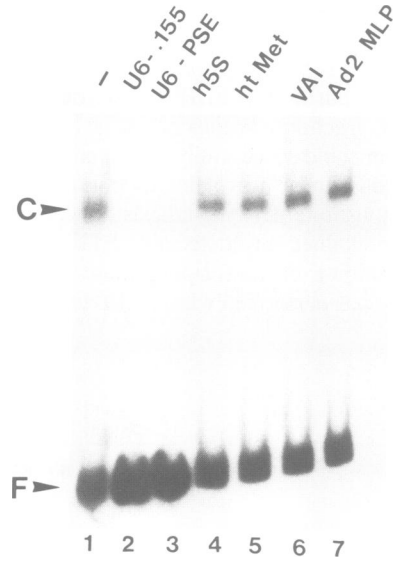


**Fig. 4**. Electrophoretic mobility shift analyses identify a protein component in fraction $DS_{0.25}$, specifically interacting with the PSE sequence element of the mouse U6 gene. The *Eco*RI−*Hin*dIII fragment of $pUmU6_{0.155}$ was employed in electrophoretic mobility shifts on 4% polyacrylamide gels. The mobility of the free DNA fragment which contains the 5′ flanking sequence of the mouse U6 gene (bp −150 to +5), is appropriately depicted (F). Lanes 2−7 show reactions in which 15.0 μl of fraction $DS_{0.25}$ were pre-incubated (30 min at 30°C) either in the absence (lane 1) or the presence of 1 μg of the following competitor plasmid DNAs before adding the radioactively labeled $U6_{0.155}$ fragment: lane 2: $pUmU6_{0.155}$; lane 3: $pUmU6_{PSE}$; lane 4: pUh5S; lane 5: $pUht^{Met}{}_I$; lane 6: $pUVA_I$; lane 7: pUAd2 MLP.

which was also exploited in the purification of the protein by affinity chromatography (Figure 2).

### *Characterization of the molecular mass of the PSE-binding protein*

The molecular mass of the PSE-binding protein was estimated by centrifugation through glycerol density gradients and subsequent assay of individual fractions by electrophoretic mobility shift analysis. As shown in Figure 6, the PSE-binding activity sediments as a single peak with a molecular mass of $\sim 90 \pm 10$ kDa. Comparable results could be arrived at by *in vitro* transcriptional analysis of glycerol gradient fractions reconstituted with phosphocellulose fractions AA and B, exhibiting the same native molecular mass of PBP (data not shown).

## Discussion

Although *cis* regulatory elements of mammalian U6 genes are well characterized (Kunkel, 1991), little is known about *trans*-acting protein components involved in their expression by RNA polymerase III. Our aim was therefore to identify and characterize transcription factors from HeLa cells involved in the basal expression of mammalian U6 genes.

### *Transcription of mammalian U6 genes in vitro*

Although mouse and human U6 genes are faithfully transcribed in cytoplasmic extracts from HeLa cells, the efficiency of this process is at least 10-fold lower than that observed for pol III genes of the AB- or AC-type. Comparative transcriptions of U6 RNA in whole cell, nuclear or cytoplasmic extracts gave no indication for the particular requirement of additional nuclear factors for basal U6 transcription. Interestingly, comparatively high template concentrations ($\sim 40-80$ μg DNA/ml) and extended incubation times (>90 min) were needed for efficient transcription *in vitro*, indicating a complex interaction and/or a weak affinity of required transcription components to their target sequences.

As demonstrated in Figure 1, *in vitro* expression of the mouse U6 gene in extracts from human or mouse cells is about four-fold more efficient than that of the human gene. In view of the identity of the coding sequence in both cases, this difference must reside in variable 5′ flanking regions of the promoters. As has previously been reported (Das *et al.*, 1988) and is born out by results in Figure 1, expression of the mouse U6 gene *in vitro* was virtually independent from the upstream enhancer (distal sequence element, DSE). This is in contrast to the human U6 gene, in which case deletion of the DSE reduced its expression to $\sim 25-35\%$ of the wild-

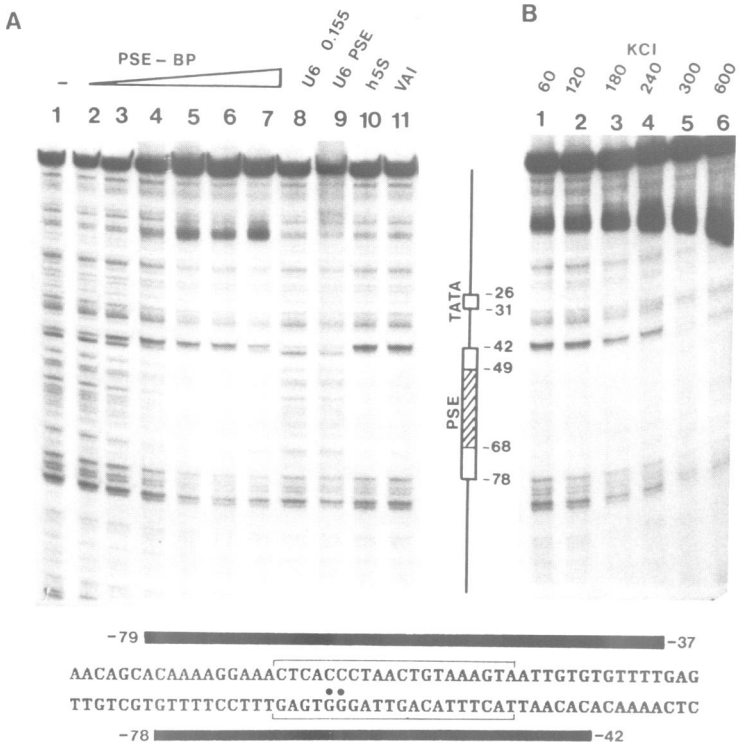


**Fig. 5.** DNase I protection of the U6 promoter by the PSE-protein containing $DS_{0.25}$ fraction. **Panel A**. The protein fraction $DS_{0.25}$ was preincubated for 30 min at 30°C with 0.1 μg/μl each of pUC18 and poly(dIdC) before incubation with the labeled 210 bp *Eco*RI–*Hin*dIII fragment ($\sim 5000$ c.p.m.) of $pUmU6_{0.155}$ containing the mouse U6 promoter. Treatment with 20 ng DNaseI was for 60 s at room temperature. Lane 1 contained no protein. Lanes 2–7 contained 0.5, 1, 2, 4, 8 or 16 μl whereas reactions in lanes 8–11 contained 8 μl of fraction $DS_{0.25}$ and were preincubated for another 30 min at 30°C with 1 μg of plasmid DNA containing single copies of the respective genes indicated at the top of the figure. **Panel B**. Salt stability of DNA protein complexes formed between the PSE-binding protein and the mouse U6 promoter. The binding reaction containing 16 μl of fraction $DS_{0.25}$ and the $U6_{0.155}$ *Eco*RI–*Hin*dIII fragment was conducted exactly as described in panel A. After addition of the labeled fragment (30 min, 30°C) the KCl concentrations (mM) of invidual reactions were adjusted as appropriately indicated in lanes 1–6, subsequently incubated for 10 min at 30°C and then treated for 60 s with 20, 30, 40, 60, 80 or 120 ng DNase I (lanes 1–6). **Lower panel**: Schematic representation of sequences protected from DNase I digestion (solid bars) in both strands of the 5′ flanking region of the mouse U6 promoter. The PSE consensus sequence is boxed by thin bars. G residues at positions –63 and –64 on the coding strand were identified by methylation interference assays to be essential for PBP binding and are appropriately marked with a dot.

type level (Figure 1). In view of its more efficient transcription we employed the heterologous mouse U6 gene to identify transcription factors in extracts from HeLa cells essential for the basal expression of mammalian U6 genes. It could subsequently be shown that all the protein components identified for the rodent U6 RNA synthesis, are also necessary for transcription of the human U6 gene, indicating that an identical set of transcription factors is required for both mammalian genes (data not shown).

In agreement with previously reported results (Steinberg *et al.*, 1990) we found that transcription of the tRNA and U6 genes shows slightly different sensitivities toward the RNA polymerase III inhibitors $\alpha$-amanitin and tagetitoxin (Figure 2). These differences can be interpreted either by: (i) a differential composition or assembly of pol III transcription complexes which could be related to the architecture of the respective promoter elements or, (ii) the possible involvement of different subspecies of RNA polymerase III (pol IIIA and pol IIIB) previously identified by chromatographic separation (Seifart and Benecke, 1975; Schwartz *et al.*, 1974) which could possibly be responsible for the selective expression of internally and externally positioned pol III promoters.

### *Identification of transcription factors involved in the basal expression of mammalian U6 genes*

*Neither TFIIIA nor TFIIIC are required for the synthesis of vertebrate U6 RNA.* Hitherto transcription factors required for the basal expression of vertebrate U6 RNA have neither been purified nor characterized. In the present study we describe the fractionation of transcriptionally active extracts from HeLa cells and show for the first time that at least four different protein components, contained in three phosphocellulose fractions, are required for the faithful transcription from a basal mammalian U6 promoter. The characterization of protein component(s) contained in the flow-through upon rechromatography on phosphocellulose (fraction AA) was not pursued further in this report, but is currently being investigated and will be documented elsewhere.

As can be deduced from Figure 2, the 5S gene specific transcription factor IIIA was dispensable for U6 transcription *in vitro* which proves the conclusion previously suggested by indirect results from chelation experiments with EDTA (Reddy, 1988). Transcription factor IIIC is similarly not required in our transcription system, since U6 RNA synthesis could not be competed by an excess of a double-stranded synthetic B-box oligonucleotide which completely eliminated tRNA (Figure 3) or VA RNA synthesis (Schneider *et al.*, 1990) *in vitro*. This result is in good agreement with the lack of TFIIIC binding sites within mammalian U6 genes or their flanking regions. Moreover we found no specific interaction of purified hTFIIIA or hTFIIIC with a labeled U6 promoter fragment in mobility shift analyses (data not shown).

Results from *in vitro* transcription analyses of a truncated U6 gene from *Saccharomyces cerevisiae*, which lacked the B-box element in its 3′ flanking region and hence partially resembled the vertebrate U6 genes, show that TFIIIC (tau) was dispensable for *in vitro* transcription (Moenne *et al.*, 1990). Using the yeast wild-type U6 gene it could be shown, however, that a DNA sequence 120 bp downstream from the coding region and encompassing a potential B-box promoter element, was required for U6 expression *in vivo* and *in vitro*, thereby suggesting an additional requirement for TFIIIC (tau) in *S.cerevisiae* U6 transcription (Brow and Guthrie, 1990). These differences in promoter structure and

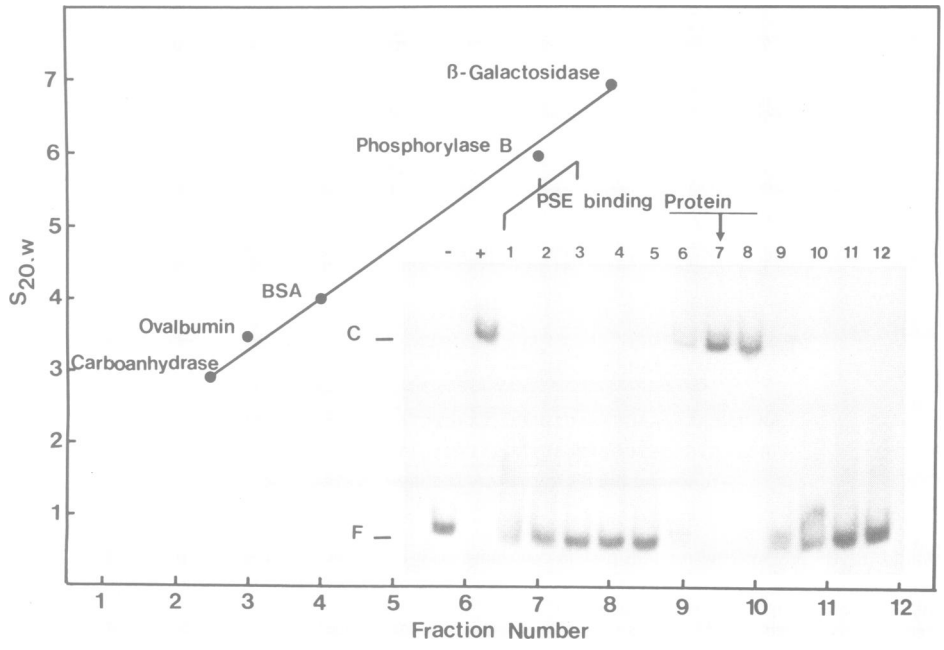


**Fig. 6.** Glycerol gradient centrifugation of the PSE-binding protein. Fraction $DS_{0.25}$ was concentrated by ammonium sulfate precipitation and contained 10 $\mu$g of protein/$\mu$l. 200 $\mu$l of the fraction was layered on a linear 4 ml glycerol gradient (12.5% to 30% in buffer 3 containing 100 mM KCl), centrifugation was for 12 h at 50 000 r.p.m. (Beckman SW60 rotor) and 4°C. Fractions of 360 $\mu$l were collected from the top and assayed for DNA-binding capacity to the mouse U6 promoter PSE sequence by electrophoretic mobility shift (insert). Complexed (C) and free (F) DNA are appropriately indicated. Individual marker proteins of known molecular mass were simultaneously run in parallel and their sedimentation was analyzed by SDS–PAGE of individual fractions.

transcription factor requirement between yeast and mammalian U6 RNA expression could possibly reflect an evolution of pol III mediated U6 transcription from an AB-like (yeast) to the pol II Usn-like (vertebrate) type, using either the classical pol III (TFIIIC) or additional pol II transcription factors (Oct 1, PBP and IID-like activity) in conjunction with the core transcription apparatus (pol III, IIIB).

*Transcription factor IIIB and an activity possibly related to TFIID are involved in the expression of mammalian U6 genes.* The direct involvement of TFIIIB in U6 transcription has hitherto only been described for the yeast system employing a truncated U6 gene (Moenne *et al.*, 1990; Margottin *et al.*, 1991). In the mammalian system we describe here for the first time that after heat inactivation of a protein fraction (fraction B from phosphocellulose), which is essential for the expression of the mouse U6 gene, two fractions are simultaneously required to reconstitute U6 transcription *in vitro*. These are: (i) a highly purified fraction containing TFIIIB (Waldschmidt *et al.*, 1988) and (ii) a fraction enriched in TFIID. These fractions were individually capable to fully reconstitute transcription of the tRNA (Figure 3) and histone H5 gene (data not shown) in appropriate transcription systems for pol III and pol II respectively. After this manuscript had been completed, our attention was drawn to a recent publication, showing that recombinant human TFIID was required for the expression of vertebrate U6 RNA (Simmen *et al.*, 1991), thereby complementing in an ideal way the conclusions drawn in this paper.

Although U6 transcription could be reconstituted in an appropriate *in vitro* system by the simultaneous addition of TFIIIB and a fraction containing TFIID (Figure 3), the efficiency of transcription was low. As discussed in the preceding paragraph, this effect was not due to a conceivable inactivation of the individual protein components. Apart from the conceivable possibility of functionally related but different forms of TFIID, the poor reconstitution of U6 synthesis could either be due to: (i) an additional component in fraction PCB, also inactivated by heat treatment and not contained in either of the two fractions added back or; (ii) partial masking by complexation of TFIID with other pol II transcription factors, thereby rendering this component only partially accessible for the interaction with pol III specific components (e.g. TFIIIB). The latter alternative is conceivable because of the unusual elution point (0.35 M KCl instead of 1 M KCl from phosphocellulose) of the U6 transcription factor activity resembling TFIID. Moreover, this latter activity in fraction PCB could only incompletely reconstitute histone H5 transcription *in vitro* (data not shown), thereby reinforcing the conclusions drawn above.

*Identification of a protein specifically binding to the PSE sequence of the mouse U6 gene and transactivating its expression.* Purification of HeLa cell extracts provided positive evidence for the existence of a U6 specific transcription factor. Fractionation on phosphocellulose identified at least three fractions (AA, B and C) which were required. Transcriptional activity contained in fraction C could further be purified by ion exchange and affinity chromatography employing the PSE motif covalently coupled to Sepharose (Figure 2). The specific binding of this protein (PBP) to the PSE sequence of the mouse U6 promoter was further verified by electrophoretic mobility shift analyses (Figure 4) and DNase I protection experiments (Figure 5). These experiments show unequivocally that the PBP protects a region centrally encompassing the PSE consensus sequence (bp −49 to −68 Dahlberg and Lund, 1988) of the mouse U6 promoter, but extending about 11 bp to the 5′ (bp −79) and 12 bp to the 3′ flanking sequence (bp −37). Competition experiments, employing either other pol III genes of the AB- or AC-type or a typical pol II promoter containing a *bona fide* TATA element, show that the PBP interacts very specifically (Figures 4 and 5A) and stably (Figure 5B) with its target sequence. It should be stressed, that the binding activity of this fraction was correlated with transcriptional activation of the U6 gene (Figures 2 and 3).

Based on the differential expression rates from basal mammalian U6 promoters (Figure 1 compare lanes 5 and 7) and on the correlation between binding and transactivating activity of fractions containing the purified PBP, we are currently analyzing in greater detail the role which the PSE-binding protein and its target sequence play in the regulation of vertebrate U6 transcription *in vitro*. Results from electrophoretic mobility shift analyses show that the mouse PSE served as an ~10-fold more potent competitor than the human PSE for PBP binding and this difference was strictly correlated with a correspondingly higher transcription from the mouse compared to the basal human U6 promoter. These results will be presented elsewhere .

To characterize the PBP from HeLa cells further, its native molecular mass was estimated by glycerol density gradient centrifugation and it was found to be ~90 ± 10 kDa (Figure 6). It was previously described that a protein related to human autoantigen Ku, which had a molecular mass of 86 and 70 kDa (Reeves and Sthoeger, 1989; Yaneva *et al.*, 1989), bound to the PSE motif of the human $U_I$ gene and was shown by immunodepletion and re-addition experiments to activate transcription (Knuth *et al.*, 1990).The Ku-protein is believed to either be a structural protein (Yaneva *et al.*, 1985) or to be involved in gene regulation (Reeves and Sthoeger, 1989) or DNA repair (Mimori and Hardin, 1986). To analyze the possible relationship between PBP and the Ku-autoantigen further, we probed different protein fractions with an α-Ku antibody (kindly provided by Dr W.van Venrooij, Amsterdam) and found it to prominently react with two protein components of the described molecular dimensions contained in our cytoplasmic HeLa cell extract (data not shown). However, these polypeptides were also recognized to a comparable extent in phosphocellulose fractions AA, B and C—of which the former two fractions contained no PBP activity. Moreover the Ku protein was shown to avidly bind to termini of DNA fragments (Mimori and Hardin, 1986), an activity which we have never observed for purified PBP. On account of these observations, it is very likely that the proteins binding to the PSE motif of the U1 and U6 promoters respectively are different. However, we can obviously not exclude at this stage of investigation that the PBP could be a member of a family of Ku related polypeptides, as has previously been suggested for the $PSE_1$ protein (Knuth *et al.*, 1990).

## Materials and methods

### *Plasmids and synthetic double-stranded DNA oligomers*

The plasmid pmU6−52 BE (Oshima *et al.*, 1981) containing the mouse U6 gene was kindly provided by Dr Y.Oshima and the 2.3 kb U6 gene insert was subcloned into the *Hin*cII site of pUC18 (pUmU6). The mouse

U6 gene, containing 340 bp from −150 to +190, thus lacking the DSE sequence element, was excised from pUmU6 with *Alu*I and subcloned into the *Hinc*II site of pUC18 ($pUmU6_{0.34}$). 155 bp of the 5′ flanking sequence extending from −150 to +5, containing the TATA and PSE sequences of the U6 gene, were excised from pUmU6 with *Alu*I and *Hgi*AI and subcloned into the *Hinc*II site of pUC18 ($pUmU6_{0.155}$). A synthetic blunt-ended 40 bp double-stranded DNA oligomer, corresponding to the sequence from bp −40 to −79 of the mouse U6 gene:

5′-ACAAAAGGAAACTCACCCTAACTGTAAAGTAATTGTGTGT
TGTTTTCCTTTGAGTGGGATTGACATTTCATTAACACACA

encompassing the proximal sequence element (PSE) from −49 to −68, was cloned into the *Hinc*II site of pUC18 ($pUmU6_{PSE}$).

A plasmid containing the human U6 gene was kindly donated by Dr G.Kunkel (Kunkel *et al.*, 1986) and the 800 bp *Ava*I fragment containing the human U6 gene was subcloned into the *Hinc*II site of pUC18 (pUhU6). A DSE sequence element deletion mutant of the human U6 gene extending from −149 to +189 was constructed by restriction of pUhU6 with *Rsa*I and *Sno*I (pUhU6 0.35). The plasmid pUyU6 was kindly provided by Dr D.Brow (p-53H6; Brow and Guthrie, 1990) and it contained 1169 bp of *S.cerevisiae* genomic DNA encompassing the complete yeast U6 snRNA gene. The plasmids $pUVA_1$, $pUht^{Met}$ and pUh5S were as previously described (Schneider *et al.*, 1989; Seifart *et al.*, 1989) and contained single copies of the genes coding for $VA_1$-RNA, human $tRNA^{Met}$ and human 5S rRNA respectively. Plasmid pUAd2MLP contained the Adeno major late promoter subcloned into pUC 18 (Weingart *et al.*, 1989). Synthetic B-box and 'irrelevant' dsDNA oligomers were as previously described (Schneider *et al.*, 1990).

### *Preparation from HeLa and mouse F9 cell extracts*

Cytoplasmic extracts (S100) from HeLa cells were prepared from several batches of 20 l suspension cultures with an index of $5\times10^5$ cells/ml as previously described (Waldschmidt *et al.*, 1988). Mouse embryonal carcinoma cells (F9 cells) were grown in monolayer culture and extracts (S 100) were prepared from $\sim2\times10^9$ confluent cells, and fractionated on phosphocellulose as described below. Extracts with a protein concentration of 15 mg/ml were stored at −80°C.

### *Purification of a protein binding to the PSE sequence of the U6 gene*

*Buffers*. buffer 1: 20 mM HEPES, pH 7.9, 20% (v/v) glycerol, 3 mM DTT, 0.2 mM PMSF; buffer 2: 20 mM HEPES, pH 7.9, 20% (v/v) glycerol, 5 mM $MgCl_2$, 3 mM DTT, 0.2 mM PMSF; buffer 3: 20 mM Tris−HCl, pH 7.9, 10% glycerol, 3 mM DTT, 5 mM $MgCl_2$, 0.2 mM PMSF; buffer 4: 40 mM Tricine, pH 7.9, 10 % glycerol, 5 mM $MgCl_2$, 3 mM DTT, 0.2 mM PMSF.

HeLa cell extract (~100 ml S100) was dialyzed against 100 vol of buffer 1 including 0.10 M KCl and applied to phosphocellulose at 10 mg of protein/ml bed volume. The flow-through (fraction PC A; 6 mg of protein/ml) was collected and the column was extensively washed with the same buffer. It was subsequently eluted with buffer 1 containing 0.35 M KCl (fraction PCB, 3 mg/ml) and 0.6 M KCl (fraction PCC; 1.3 mg of protein/ml). The PCA fraction was adjusted to 0.6 M KCl and rechromatographed on a second column of phosphocellulose as described (Seifart *et al.*, 1989) yielding hTFIIIA free fraction PCAA (4.3 mg/ml).The PCC fraction was dialysed against buffer 2 containing 0.05 M ammonium sulfate and applied to a column of DEAE−Sephadex A25 at a loading rate of 5 mg protein/ml bed volume. The flow-through was collected (fraction $DS_{0.05}$; 0.7 mg protein/ml). The column was washed with the same buffer and subsequently eluted with 0.25 M ammonium sulfate (fraction $DS_{0.25}$; 1 mg of protein/ml). This fraction was dialyzed against buffer 3, incubated for 30 min at 30°C with unspecific competitor DNA [0.1 mg/ml poly(dIdC) and pUC18] and applied to a PSE−DNA affinity column equilibrated in the same buffer. It contained ~100 μg/ml bed vol of multimerized PSE sequence of the mouse U6 gene covalently coupled to CNBr-activated Sepharose CL 6 B (Arndt-Jovin *et al.*, 1975). The flow-through was collected (fraction $PSE_{0.06}$; 0.7 mg protein/ml). The column was extensively washed and eluted with 0.5 M KCl (fraction $PSE_{0.5}$; 0.08 mg protein/ml). Alternatively, fraction $DS_{0.25}$ was dialyzed against buffer 4 with 0.2 M KCl and applied to a 1 ml Mono Q FPLC column equilibrated in the same buffer. After washing with buffer 4 containing 0.2 M KCl (fraction $MQ_{0.2}$, 0.2 mg protein/ml), the column was eluted with 0.30 M KCl (fraction $MQ_{0.3}$; 0.13 mg protein/ml).

A protein fraction enriched in TFIID but containing no detectable amounts of polIII or transcription factors IIIA, B or C (IID, phosphocellulose 1 M KCl eluate, 0.2 mg of protein/ml) was prepared by heparin−Sepharose and phosphocellulose chromatography of whole cell extracts and was kindly donated by Dr F.Düring (IMT, Marburg).

### *In vitro transcription*

The *in vitro* transcription reactions were performed as previously described (Jahn *et al.*, 1987) but with the following modifications: Incubation time was for 120 min at 30°C in a final vol of 50−100 μl. *In vitro* synthesized RNA products were electrophoretically separated on 7% denaturing urea sequencing gels and autoradiographed for at least 36 h at −80°C with an intensifying screen.

### *Electrophoretic mobility shift analyses of protein−DNA complexes*

The 210 bp *Eco*RI−*Hin*dIII fragment of $pUmU6_{0.155}$ was gel purified and labeled at its 3′ terminus with [α-$^{32}$P]dCTP and the Klenow fragment of DNA polymerase I. Protein fractions (15 μl) were preincubated in buffer 3 (30 min at 30°C) with 0.1 μg/μl each of pUC18 and poly(dIdC) before incubation with the labeled fragment (30 min at 30°C). Where appropriate, samples were incubated with competitor DNA for 30 min at 30°C as indicated in the legends of Figures 3 and 4. After incubation, the reaction mixtures were electrophoresed on 4% polyacrylamide gels and further processed as described (Schneider *et al.*, 1990).

### *DNase I footprinting*

The 210 bp fragment containing the PSE sequence of the mouse U6 gene was labeled on the coding strand at its 3′ terminus with [α-$^{32}$P]dCTP and the Klenow fragment of DNA polymerase I. The binding reaction with the protein fraction was conducted as described under the previous section. After assembly on ice and incubation at 30°C for 30 min, digestion was initiated by the addition of DNase I. After 60 s at room temperature the samples were processed as described (Seifart *et al.*, 1989).

### *Rate zonal centrifugation on glycerol gradients*

The fractions which were to be analyzed were dialyzed against buffer 3 and layered onto linear 12.5−30% glycerol gradients which were centrifuged and analyzed as described in Figure 6.

## Acknowledgements

We gratefully acknowledge the expert technical assistance of Ulla Kopiniak and Frauke Seifart, as well as financial support by the Deutsche Forschungsgemeinschaft and the Fonds der Chemischen Industrie. We also thank Andreas Blum, from our laboratory, for providing subcloned human U6 DNA and for data from experiments in the initial part of this investigation.

## References

Arndt-Jovin,P.J., Jovin,T.M., Bähr,W., Frischauf,A.M. and Marquardt,M. (1975) *Eur. J. Biochem.*, **54**, 411−418.
Bark,C., Wello,P., Zabielski,J., Janson,L. and Petterson,U. (1987) *Nature*, **328**, 356−359
Brow,D. and Guthrie,C. (1990) *Genes Dev.*, **4**, 1345−1356
Carbon,P., Murgo,S., Ebel,J., Krol,A., Tebb,G. and Mattaj,I. (1987) *Cell*, **51**, 71−79.
Dahlberg,J.E. and Lund,E. (1988) In Birnstiel,M.L. (ed.) *Structure and Function of Major and Minor Small Nuclear Ribonucleoprotein Particles*. Springer, Berlin, pp. 38−70.
Das,G., Henning,D., Wright,D. and Reddy,R. (1988) *EMBO J.*, **7**, 503−512.
Engelke,D.R., Ng,S.-Y., Shastry,B.S. and Roeder,R.G. (1980) *Cell*, 19, 717−728.
Geiduschek,E. and Tocchini-Valentini,G. (1988) *Annu. Rev. Biochem.*, **57**, 873.914.
Guthrie,C. and Patterson,D. (1988) *Annu. Rev. Genet.*, **22**, 387−419
Jahn,D., Wingender,E. and Seifart,K.H. (1987) *J. Mol. Biol.*, **192**, 303−313.
Kassavetis,G.A., Braun,B.R., Nguyen,L.H. and Geiduschek,E.P. (1990) *Cell*, **60**, 235−245.
Knuth,M.W., Gunderson,S.I., Thompson,N.E., Strasheim,L.A. and Burgess,R.R. (1990) *J. Biol. Chem.*, **265**, 17911−17920.
Kunkel,G., Maser,R. Calvet J. and Peterson T. (1986) *Proc. Natl. Acad. Sci. USA*, 83, 8575−8579
Kunkel,G. and Pederson,T. (1988) *Genes Dev.*, 2, 196−204
Kunkel,G. and Pederson,T. (1989) *Nucleic Acids Res.*, **17**, 7371−7379.
Kunkel,G. (1991) *Biochim. Biophys. Acta*, 1088, 1−9.
Lobo,S. and Hernandez,N. (1989) *Cell*, 58, 55−67
Margottin,F., Dujardin,G., Gerard,M., Egly,J.M., Huet,J. and Sentenac,A. (1991) *Science*, **251**, 424−426.
Mattaj,I., Dathan,R., Parry,H., Carbon,P. and Krol,A. (1988) *Cell*, **55**, 435−442.

Mimori,T. and Hardin,J.A. (1986) *J. Biol. Chem.*, **261**, 10375–10379.
Moenne,A., Camier,S., Anderson,G., Margottin,F., Beggs,J. and Sentenac,A. (1990) *EMBO J.*, **9**, 271–277
Myers,R.M., Tilly,K. and Maniatis,T. (1986) *Science*, **232**, 613–615.
Nakajima,N., Horikshi,M. and Roeder,R.G. (1988) *Mol. Cell. Biol.*, **8**, 4028–4040.
Oshima,Y., Okada,N., Tani,T., Hoh,Y. and Hoh,M. (1981) *Nucleic Acids Res.*, **9**, 5145–5158.
Reddy,R. (1988) *J. Biol. Chem.*, **263**, 15980–15984.
Reeves,W.H. and Sthoeger,Z.M. (1989) *J. Biol. Chem.*, **264**, 5047–5052.
Schneider,H.R., Waldschmidt,R., Jahn,D. and Seifart,K.H. (1989) *Nucleic Acids Res.*, **17**, 5003–5016.
Schneider,H.R., Waldschmidt,R. and Seifart,K.H. (1990) *Nucleic Acids Res.*, **18**, 4743–4750.
Schwarz,L.B., Sklar,V.E.F, Jaehning,J.A., Weinmann,R. and Roeder,R.G. (1974) *J. Biol. Chem.*, **249**, 5889–5898
Seifart,K.H. and Benecke,B.J. (1975) *Eur. J. Biochem.*, **53**, 293–300
Seifart,K.H., Wang,L.R., Waldschmidt,R., Jahn,D. and Wingender,E. (1989) *J. Biol. Chem.*, **264**, 1702–1708.
Simmen,K.A., Bernués,J., Parry,H.D., Stunnenberg,H.G., Berkenstamm,A., Cavallini,B., Egly,J.M. and Mattaj,I.W. (1991) *EMBO J.*, **10**, 1853–1862.
Steinberg,T.H., Mathews,D.E., Durbin,R.D. and Burgess,R.R. (1990) *J. Biol. Chem.*, **265**, 499–505.
Waldschmidt,R., Jahn,D. and Seifart,K.H. (1988) *J. Biol. Chem.*, **263**, 13350–13356.
Weingart,S., Sommer,U., Gerhold,H. and Seifart,K.H. (1989) *Eur. J. Biochem.*, **183**, 145–153.
Yaneva,M., Ochs,R., McRorie,D.K., Zweig,S. and Busch,H. (1985) *Biochim. Biophys. Acta*, **841**, 22–29.
Yaneva,M., Wen,J., Ayala,A. and Cook,R. (1989) *J. Biol. Chem.*, **264**, 13407–13411.

*Received on April 16, 1991; revised on May 23, 1991*