Development 109, 349-362 (1990)
Printed in Great Britain © The Company of Biologists Limited 1990

Structure and expression of the mouse Oct2a and Oct2b, two differentially spliced products of the same gene

ANTONIS K. HATZOPOULOS, ANASTASIA S. STOYKOVA*, JÖRN R. ERSELIUS, MARTYN GOULDING, TOOMAS NEUMAN and PETER GRUSS

Max-Planck Institute of Biophysical Chemistry, Department of Molecular Cell Biology, 3400 Göttingen, FRG

*on leave of absence from The Bulgarian Academy of Sciences, Regeneration Research Laboratory, Zdrave 2, 1431 Sofia, Bulgaria

Summary

A large family of tissue-specific nuclear proteins interact with the octamer motif ATTTGCAT, a transcriptional regulatory element found in the promoter and enhancer sequences of many genes. As a step towards elucidating the mechanism of this regulation, cDNA clones of the mouse Oct2 protein were isolated. One, called here Oct2b, encodes a larger variant of the previously described Oct2a proteins. The Oct2b cDNA has an insertion of 74 bp close to the 3′ end which creates an open reading frame distinct from Oct2a. As a result, the Oct2b protein has a carboxy end which is similar to that of the ubiquitous octamer-binding protein Oct1. Analysis of the Oct2 gene shows that Oct2a and Oct2b are differentially spliced products of the same gene. The insertion in the Oct2b cDNA results from the inclusion of an additional exon in the mRNA which would otherwise reside in an intron sequence of the Oct2a transcript. RNA analysis demonstrates that both Oct2a and 2b mRNAs are most abundant in B-cells but they are also expressed in a variety of tissues including brain, intestine, testis, kidney, as well as in embryos. Interestingly, the ratio of Oct2a and 2b varies among tissues. *In situ* hybridization studies during mouse embryogenesis show that the Oct2 gene is widely expressed in the developing nervous system. In contrast, expression in the adult brain is confined to very specific areas which include the suprachiasmatic and medial mammillary nuclei, hippocampus, olfactory tract and the olfactory bulb. Oct2 proteins are present in both neuronal and oligodendroglial cells, although they are more abundant in glial cells.

Key words: octamer, Oct2b protein, DNA-binding proteins, Oct2 gene, differential splicing, embryogenesis.

Introduction

The octamer motif, ATTTGCAT, is a transcriptional regulatory element found in the promoter and enhancer sequences of many genes (Falkner and Zachau, 1984; Parslow *et al.* 1984; Bergman *et al.* 1984; Ares *et al.* 1985; Mattaj *et al.* 1985; Mason *et al.* 1985; Mangin *et al.* 1986; Carbon *et al.* 1987; for a review see Hatzopoulos *et al.* 1988). The octamer is necessary and sufficient for B-lymphocyte-specific expression of the immunoglobulin genes (Wirth *et al.* 1987; Lenardo *et al.* 1987; Gerster *et al.* 1987). Moreover, the octamer is also found in the regulatory region of genes that are not lymphoid-specific, including the thymidine kinase gene (Parslow *et al.* 1987), the snRNA genes (Carbon *et al.* 1987) and the histone H2b gene (LaBella *et al.* 1988).

Previously, two proteins have been described that interact with the octamer motif (Singh *et al.* 1986; Landolfi *et al.* 1986; Staudt *et al.* 1986; Rosales *et al.* 1987; Scheidereit *et al.* 1987). One called Oct1 (also NF-A1, NFIII, OBP100 or OTF-1) is ubiquitously expressed and is involved in the cell-cycle regulation of the histone H2b gene (Fletcher *et al.* 1987; Sturm *et al.* 1988*a*), in adenovirus DNA replication (Prujin *et al.* 1986; Rosenfeld *et al.* 1987) and in the expression of the snRNA genes (Ares *et al.* 1985; Mattaj *et al.* 1985; Mangin *et al.* 1986; Carbon *et al.* 1987; Tanaka *et al.* 1988). The second, Oct2 (also NF-A2 or OTF-2), was originally thought to be present only in B-cells (Landolfi *et al.* 1986; Staudt *et al.* 1986) and, as such, responsible for the lymphocyte-specific expression of the immunoglobulin genes. Recently, both Oct1 and Oct2 proteins have been purified and their respective cDNAs have been isolated and characterized (Sturm *et al.* 1987; Fletcher *et al.* 1987; O'Neil and Kelly, 1988; Scheidereit *et al.* 1987; Sturm *et al.* 1988*b*; Staudt *et al.* 1988; Ko *et al.* 1988; Müller *et al.* 1988; Scheidereit *et al.* 1988; Clerc *et al.* 1988). The two sequences share extensive homology in the area responsible for DNA binding called the POU-domain (Herr *et al.* 1988), which is also found in other genes (Bodner *et al.* 1988; Ingraham *et al.* 1988; Finney *et al.* 1988). Part of this

region shows similarity to the homeobox, a conserved sequence originally found in many *Drosophila* genes involved in development (Gehring, 1987). During recent years, a large number of mammalian genes containing homeobox sequences have been isolated (for a review and references see Dressler and Gruss, 1988). Their expression during embryogenesis suggests of a role in pattern formation; however, their exact function is unknown even though they can interact with specific DNA sequences (Odenwald *et al.* 1989). The cloning of Oct1 and Oct2 shows that the homeodomain is a DNA-binding structure employed not only by developmentally regulated genes but also by genes involved in tissue-specific gene expression.

In addition to Oct1 and Oct2, another lymphoid-specific factor known as OTF-2B has been described (Schreiber *et al.* 1988), as well as a testis-specific one (Barberis *et al.* 1987), a factor present in F9 embryonic carcinoma cells termed NF-A3 (Lenardo *et al.* 1989) and a factor present in melanoma cells (Cox *et al.* 1988). Recently, we have shown that a family of proteins, termed Oct1 to Oct10, interact with the octamer motif raising the possibility that a large set of genes is regulated by distinct octamer-binding proteins in a tissue-specific manner (Schöler *et al.* 1989*a*).

An important finding of our previous study was that Oct2-like complexes are also present in nuclear extracts of brain, kidney, mouse day 12 embryo and sperm (Schöler *et al.* 1989*a*). These results raised several interesting questions. Are the Oct2-like complexes due to the presence of the Oct2 protein in the tissues mentioned above, or to a distinct protein that fortuitously has the same mobility shift? In case of the first possibility, where is Oct2 expressed during embryogenesis and what is its role? Are the large number of octamer binding proteins, the products of distinct genes as it appears to be the case for Oct1 and Oct2 or are some of these due to differential splicing of the same gene?

In order to answer these questions, we isolated mouse cDNA sequences from B-cells encoding the murine equivalent of Oct2 (or OTF-2 or NF-A2), here called Oct2a (see below). Using this as a probe, it is shown that Oct2a is expressed in a wide variety of tissues including brain, kidney, intestine, testis as well as in mouse day 12 embryos.

In addition to cDNA clones encoding Oct2a, a second class of cDNAs was isolated, called Oct2b, that showed two differences from Oct2a. The first difference is a deletion of 48 bases (16 amino acids) upstream from the *POU*-domain. This deletion is also found in some human Oct2a clones (Scheidereit *et al.* 1988). The second is an insertion of 74 bases at an area after the homeobox and close to the 3′ end. This insertion creates a distinct longer open reading frame which results in a completely different carboxy terminus. Interestingly, this new carboxy end is very similar to the corresponding area of Oct1 indicating that they might have similar functions.

Initial characterization of the mouse Oct2 gene shows that the 48 base deletion in Oct2b results from the use of a downstream splice acceptor site, while the 74 base insertion is due to the splice of an additional exon found in the intron sequences of the nuclear Oct2a transcript.

The Oct2b cDNA was transcribed *in vitro*, translated in reticulocyte lysates and used in the electrophoresis mobility shift assay. The results suggest that this cDNA encodes the previously identified OTF-2B protein (predicted M_r 75×10^3; Schreiber *et al.* 1988). Oct2b, which was also thought to be lymphoid specific, is expressed in the same tissues as Oct2a, although the ratio of Oct2a to Oct2b varies among tissues.

In situ hybridization in day 12 embryos shows that the Oct2 gene is expressed mainly throughout the developing nervous system, i.e. the neural tube, diencephalon, mesencephalon, metencephalon and myelencephalon. In contrast, expression in the adult brain is confined to a limited number of areas, which include the medial mammillary and suprachiasmatic nuclei, the CA1-CA4 areas of the hippocampus, the olfactory bulb and the olfactory tract. By separating neuronal and oligodendroglial nuclei and preparing nuclear extracts, it is shown that Oct2a protein is found in both cell types but is more abundant in glial cells.

Materials and methods

Isolation of cDNA and chromosomal clones

The mouse B-cell lymphoblastoid cDNA library in λgt11 (Clontech) and the mouse C57 chromosomal library in λEMBL3 (kindly provided by Dr Kenji Imai) were screened with chemically synthesized oligonucleotides of the homeobox sequence of the human Oct2 (Ko *et al.* 1988; Müller *et al.* 1988; Scheidereit *et al.* 1988; Clerc *et al.* 1988). Two single-stranded 90 bases long oligonucleotides were synthesized covering the entire 180 bases of the coding strand of the homeobox sequence. Two 12 nucleotides long primers were also made and used to produce radioactive non-coding strand probes by the Klenow fragment of DNA polymerase. After plating the libraries, denatured phage DNA was transferred to Hybond-N filters (Amersham) and hybridized under low-stringency conditions (6×SSC, 60°C). The filters were washed two times in 2×SSC, 0.1 % SDS at room temperature and two times at 42°C. Positive plaques were isolated, phages were grown and phage DNA was purified according to standard procedures (Maniatis *et al.* 1982). cDNA clones were named mR1 to mR5 and chromosomal clones mCR1 to mCR4. The two different Oct2 transcripts are referred in the text as mOct2a and mOct2b while the protein products and the protein–DNA complexes in the EMSA as Oct2a and Oct2b.

Subcloning and sequencing

The cDNA inserts of λgt11 clones mR2 and mR4 were cloned into the plasmid vector Bluescript (Stratagene) and mapped with various restriction enzymes (plasmids p1.2 and p2.2, respectively). In addition, the 2.2 kb insert of clone mR4 (sequence in Fig. 1) was divided in two fragments by using the *Hin*dIII single restriction site at position 1342. The resulting 1.3 and 0.9 kb fragments were cloned in Bluescript (plasmids pK and pL, respectively). Similarly, the 1.2 kb insert of clone mR2 was divided by *Hin*dIII into two fragments of 0.62 kb (5′ part) and 0.58 kb (3′ part) long and cloned in Bluescript (plasmids pF and pE, respectively). The 385 bp fragment of mR4 from the *Sac*I site at position 1590 to the *Nae*I site at

position 1975 in Fig. 1 was also subcloned in Bluescript (plasmid pN). Fragments K, L, E and F were cloned into M13 in both orientations and sequenced with the dideoxy method (Sanger *et al.* 1977) using both deazaGTP and dITP nucleotides and chemically synthesized internal primers (Sequenase Kit from US Biochemical). Various chromosomal fragments covering the area that hybridized with the cDNA inserts were also subcloned in Bluescript and sequenced directly using the dideoxy method (Sanger *et al.* 1977).

RNA isolation and RNase protection analysis

Female NMRI outbred mice were mated overnight and the day of the vaginal plug was noted as day 0. Embryos were dissected free of extraembryonic tissue and frozen in liquid nitrogen. RNA was isolated by homogenizing mouse day 12 embryos and tissues of adult mice in 5 M guanidinium thiocyanate, 50 mM Tris–HCl pH 7.5, 10 mM Na_2EDTA pH 8.0 and 5% v/v β-mercaptoethanol (Chirgwin *et al.* 1979) and centrifuged for 10 min at 10000 revs min^{-1}. 1/10 vol of 20% Sarkosyl was added to the supernatant and the samples were heated at 65°C for 3 min. 0.1 g ml^{-1} CsCl was added and samples were centrifuged through a 5.7 M CsCl, 100 mM Na_2EDTA pH 8.0 cushion in a Beckman SW28 rotor spun at 25000 revs min^{-1} at 18°C overnight. The RNA pellets were redissolved in 3 ml 5 mM Na_2EDTA pH 8.0, 0.5% Sarkosyl and 5% β-mercaptoethanol. After phenol–chloroform–isoamyl alcohol (25:24:1) extractions, the RNA was precipitated by addition of 1/10 vol 3 M sodium acetate pH 5.2 and 2.5 vol ethanol. For the RNase protection experiments, plasmids pF and pN were linearized using the *Bam*HI site within the Bluescript polylinker and the radioactively labelled antisense strand was synthesized by T3 polymerase. After phenol–chloroform–isoamyl alcohol (25:24:1) extraction and two ethanol precipitations, the probes were dissolved in 80% formamide, 20 mM Pipes pH 6.4, 400 mM NaCl, 1 mM Na_2EDTA pH 8.0. RNA pellets (4–40 mg) were also dissolved in 30 μl of the same buffer and 1 μl of probe (100000–500000 cts min^{-1}) was added. The samples were heated for 10 min at 85°C and subsequently hybridization was carried overnight at 65°C. Next day, the samples were treated with 0.7 μg RNase A and 70 units RNase T1 for one hour at 37°C in 350 μl 10 mM Tris–HCl pH 7.5, 5 mM EDTA, 300 mM NaCl. Subsequently, 50 μg proteinase K and 20 μl 10% SDS were added and reactions carried for 15 min. (All three enzymes were from Boehringer Mannheim). After phenol–chloroform–isoamyl alcohol (25:24:1) extraction and ethanol precipitation, the products were separated in 6% denaturing polyacrylamide gels. The gels were dried and autoradiographed.

Northern analysis

Total or poly $(A)^+$ was denatured at 68°C, separated on a 1.2% agarose–formaldehyde gel and blotted onto Hybond N (Amersham) using 10×SSC. Prehybridization was carried out at 42°C in 50% formamide, 6×SSC, 5×Denhardt's, 0.1% SDS and 100 μg ml^{-1} denatured salmon sperm DNA for 4 h. Hybridization was carried out under the same conditions for 12 h in the presence of the ^{32}P-labelled probe (1×10^6 cts min^{-1}). The filters were washed twice in 2×SSC, 0.1% SDS at room temperature for 20 min, followed by two washes of 10 min at 60°C in 0.2×SSC and 0.1% SDS and exposed to Kodak XAR film for 3 days.

In situ *hybridization*

Plasmid pF was linearized and single-stranded RNA probes were transcribed *in vitro* using 100 mCi ^{35}S-UTP and T3 or T7 polymerases (Promega Biotech.). After DNase digestion, probes were precipitated with 10% trichloroacetic acid and collected on nitrocellulose filters (Millipore). Probes were eluted from the filters in 50 mM EDTA pH 8.0, 0.1% SDS at 65°C. Following ethanol precipitation, probes were partially degraded with 0.2 N NaOH on ice for 30–60 min and neutralized with 1 M acetic acid. After ethanol precipitation, the probes were resuspended in 50% formamide, 10 mM DTT. Sections (8 μm) were prepared and hybridized as described (Hogan *et al.* 1986; Dony and Gruss, 1987). Sections were cut in a cryostat and transferred onto subbed slides. The sections were dried at 55°C, fixed in 4% paraformaldehyde (PFA) and dehydrated in graded ethanol. Slides were kept at −20°C until the day of hybridization. Slides were dipped in distilled water and incubated at 70°C in 2×SSC (standard saline citrate). After a second rinse with distilled water, slides were digested with 0.125 mg ml^{-1} pronase for 10 min at room temperature and the digestion was stopped in 0.2% glycine for 30 s. Slides were rinsed in PBS and refixed in 4% PFA for 20 min and rinsed again in PBS. Slides were acetylated in 0.1 M triethanolamine with 1/400 volume acetic anhydride, made fresh. Slides were rinsed again in PBS and dehydraded in graded ethanol. After prolonged air drying, the hybridization mixture was added. The probe was diluted to approximately 5×10^4 cts ml^{-1} in a buffer containing 50% formamide, 2×SSC, 10 mM Tris, 10 mM $NaPO_4$ pH 6.8, 5 mM EDTA, 10% dextran sulphate, 10 mM DTT, and 150 μg ml^{-1} tRNA, 150 μg ml^{-1} denatured salmon sperm DNA, 0.1 mM UTP, 10 μM S-ATP, 1 mM ADPβS and 10 mM β-mercaptoethanol. The hybridization mix was boiled for 2 min, applied directly onto sections and covered with a siliconized cover slip. Hybridization was done overnight in a humid chamber at 50°C. The slides were washed for 3–4 h in 50% formamide, 2×SSC and 10 mM β-mercaptoethanol at 37°C followed by RNase digestion. A second wash in 50% formamide, 2×SSC, 10 mM β-mercaptoethanol was done overnight and the slides were then dehydrated in graded ethanol. Slides were dipped in Kodak NTB-2 emulsion diluted 1:1 with water and allowed to dry in a dark chamber for 2–3 h.

Slides were placed in a dark plastic box, wrapped in foil and allowed to expose for 7–20 days at 4°C. Development was done at room temperature for 3 min in Kodak D-19, followed by 30 s in 1% acetic acid and 3 min in 30% sodium thiosulphate. After repeated washes in distilled water, the mouse embryo day 12 slides were stained with Giemsa and allowed to dry. The adult mouse brain sections were stained with 0.5% toluidine blue in 10 mM sodium acetate pH 4.6 for 10 min, quickly rinsed in distilled water and then dehydrated in acetone, acetone:xylene (1:1) and xylene for 5 min each. Photomicrographs were taken with a Leitz Labovert bright-field/dark-field microscope.

Nuclear extract preparation and electrophoretic mobility shift assay (EMSA)

Nuclear extracts were prepared as described before (Dignam *et al.* 1983; Schöler *et al.* 1989*a*). The separation of neuronal and glial nuclei was carried essentially as previously described (Stoykova *et al.* 1979). Binding conditions and electrophoresis were as described in Schöler *et al.* 1989*a*. The DNA fragment harbouring the octamer motif (Gerster *et al.* 1987, termed 1W in Schöler *et al.* 1989*a*) contains the region 518–564 of the immunoglobulin heavy chain enhancer except at position 562 (numbering according to Ephrussi *et al.* 1985).

In vitro transcription and translation

Plasmid p2.2 was linearized using the *Bam*HI and *Sal*I restrictions sites in the Bluescript polylinker and the sense and

antisense RNA strands were synthesized with the T3 and T7 RNA (Promega) polymerases, respectively. The reaction conditions were 40 mM Tris–HCl pH 8.0, 8 mM $MgCl_2$, 25 mM NaCl, 2 mM spermidine, 5 mM DTT, 2 mM NTPs, 2 mM $m^7G(5')ppp5'G$, 2 μg template and 2 μl enzyme. 1/50 of the RNA template was translated in a rabbit reticulocyte lysate following the manufacturer's specifications (BRL). 1/6 of the translation mix was used in the EMSA as described above.

Results

Isolation of mouse Oct2a and Oct2b clones

Two single-stranded 90 nucleotides long fragments spanning the entire homeobox area of the human Oct2 sequence (Ko *et al.* 1988; Müller *et al.* 1988; Scheidereit *et al.* 1988; Clerc *et al.* 1988) were synthesized, radioactively labelled and used as probes to screen a mouse B-cell λgt11 cDNA library. Five clones were isolated called mR1 to mR5. The insert size ranged from 1 to 3.5 kb. Two clones, mR2 (1.2 kb) and mR4 (2.2 kb) were selected for further characterization. Sequence analysis showed that the mR2 insert coded for an incomplete sequence at the 5′ end of the mouse Oct2 gene, here called mOct2a (see below). The sequence of the mR4 insert showed that it is also a mouse Oct2 clone, but when it was compared to mOct2a and the published human sequences (Ko *et al.* 1988; Müller *et al.* 1988; Scheidereit *et al.* 1988; Clerc *et al.* 1988) two differences emerged. Therefore, this sequence was tentatively named mOct2b. The first difference is a deletion of 48 bp upstream of the POU-domain at the end of the glutamine-rich region which removes 16 amino acids from the protein product (Fig. 1). This deletion was also present in one of the human Oct2a clones (Scheidereit *et al.* 1988). The second difference is an insertion of 74 bp close to the 3′end of the human and mouse Oct2a, at 37 bp upstream from the stop codon (Fig. 1). Since the 74 bp does not cause an in frame insertion, a new open reading frame is created. As a result, the last 12 amino acids of Oct2a are not present and instead 132 new residues are added before a new stop codon is encountered.

When the human and mouse sequences are compared, they show a high degree of homology. Most base changes are occuring in the third position of the codon. Consequently, only a few amino acid substitutions occur between human and mouse sequences (Fig. 1). None of the amino acid substitutions are located within the POU-domain, which is required for DNA binding (Sturm and Herr, 1988).

The predicted protein sequence is also shown in Fig. 1. Upstream of the methionine assigned as position 1, multiple stop codons are found in all three reading frames. Methionines are found at positions 6 and 13. The first two, 1 and 6, match well the optimal AUG context (Kozak, 1986), but at this point it is unclear which one is actually used. Moreover, at the 3′ end no typical poly (A) tail is found indicating that the clones were produced by internal priming in an A-rich area of the message.

The carboxy ends of Oct2b and Oct1 are similar

The 132 new amino acids in the carboxy end of Oct2b are mainly serines–threonines (28 %), alanines (15 %), leucines–isoleucines (15 %), glycines (13.6 %) and prolines (9.8 %) with very few charged amino acids. This carboxy terminus of Oct2b shows extensive similarity to the carboxy end of Oct1. Computer homology plots show that, besides the extensive homology found in the area of the POU-domain, this is a second area that the two sequences are very homologous. In Fig. 2 the last 133 amino acids of Oct2b are aligned with the last 128 amino acids of Oct1. The homology to Oct1 starts within the insertion at amino acid 457 and extends into the Oct2b open reading frame. This similarity suggests that the carboxy terminal domains of the two proteins may have similar functions.

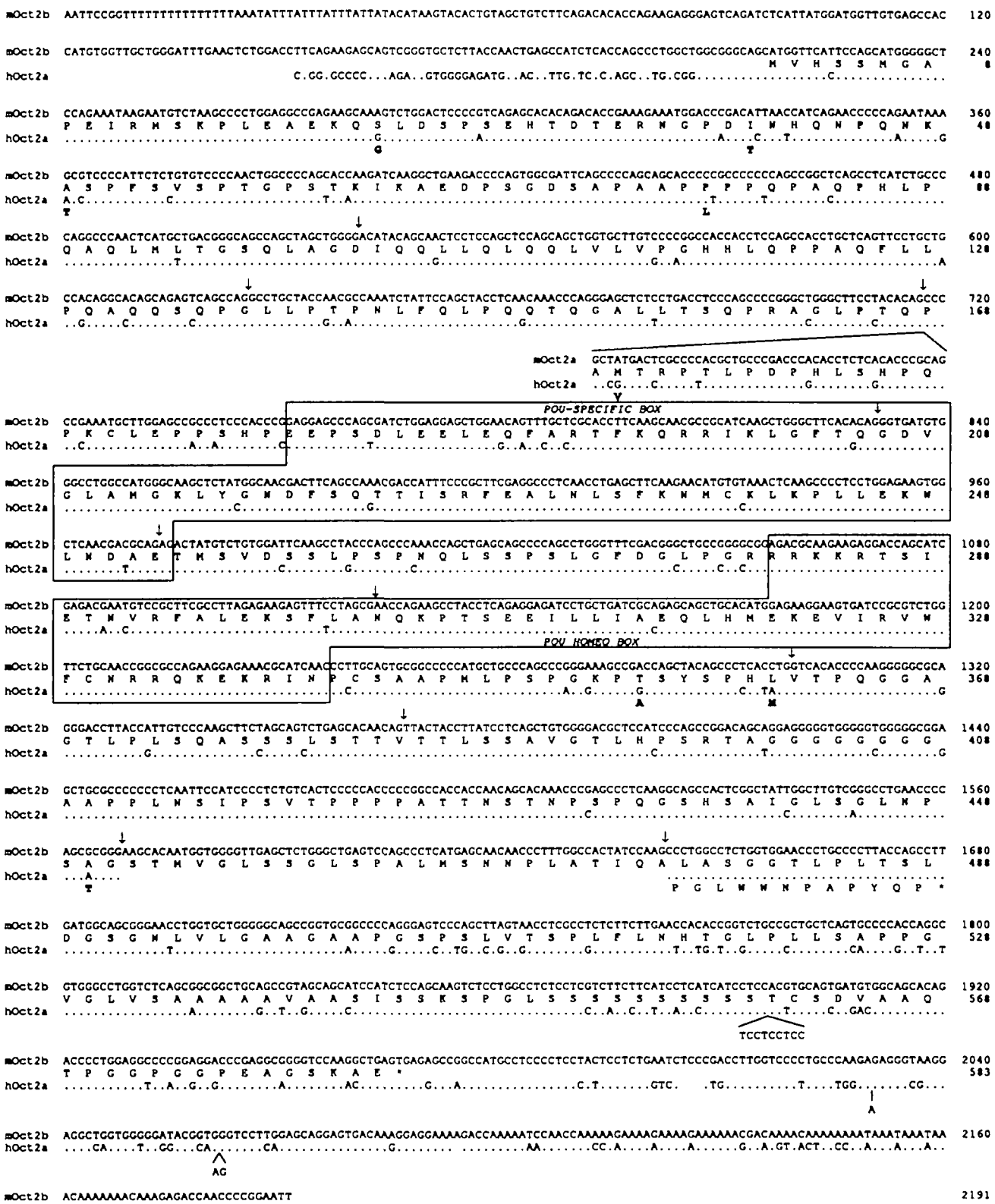
Clone mR4 encodes a distinct protein

The open reading frame of clone mR4 encodes for a protein that is 103 amino acids longer than Oct2a or 119 amino acids longer than the Oct2a variant with the 16 residues deletion (Scheidereit *et al.* 1988). Since Oct2a runs with an apparent M_r of around 60×10^3 (Scheidereit *et al.* 1987; Müller *et al.* 1988), the new clone would appear as a $70–75\times10^3$ protein. Recently a new octamer-binding factor called OTF-2B present in certain B-cell lines has been identified (Schreiber *et al.* 1988). The predicted M_r of this protein is 75×10^3 suggesting that mOct2b might encode OTF-2B. To confirm this, the mR4 insert was excised and cloned in the plasmid vector Bluescript. Using the T3 and T7 polymerases the 'sense' (coding) and 'antisense' (non-

Fig. 1. Sequence of mouse Oct2b. The 2191 nucleotides of the entire insert of λgt11 clone mR4 is shown as mOct2b. The predicted amino acid sequence is shown below the nucleotide sequence in one letter code. The first in frame methionine is assigned number 1. The published sequence of the human Oct2a clone (hOct2a) is shown for comparison. Identical nucleotides are depicted as dots. When a base substitution creates a codon for a different amino acid, the amino acid found in the human protein is shown below the hOct2a sequence with bold letters. The 48 bp deletion in mOct2b is shown as an insertion after nucleotide position 717. In this case, the mouse mOct2a sequence is shown for comparison. In mOct2b, there is a 74 bp insertion after nucleotide position 1568, which creates a new open reading frame. As a result the carboxy ends encoded by mOct2b and hOct2a are different. The carboxy ends encoded by mOct2a and hOct2a will be identical since there are no nucleotide differences between mouse and human sequences in this area. At the 3′ end of the hOct2a, small gaps and insertions were allowed in order to maximize alignment. These differences do not represent differences between Oct2a and 2b but rather interspecies variations between mouse and human. The POU-specific and POU homeo sequences are boxed. The arrows above the mOct2b sequence indicate the location of exon-intron boundaries mapped so far as determined by analysis of chromosomal clones (see also Fig. 4). Note that the arrow after nucleotide 520 does not represent the first exon-intron junction since the complete structure of the Oct2 gene is not known yet. Asterisks indicate the two different stop codons in mOct2a and mOct2b.

coding) RNAs, respectively, were synthesized. The RNAs produced were then translated *in vitro* using rabbit reticulocyte lysates. After translation, the reticulocyte extracts were used in the electrophoretic mobility shift assay (EMSA) employing a radioactive fragment from the immunoglobulin heavy chain enhancer which contains the octamer motif (Fig. 3). The results confirmed that the protein product encoded by mOct2b has an identical mobility in the EMSA as the complex Oct2b. As shown in Fig. 3, the reticulocyte extract also contains endogenous Oct1 protein. A second complex that runs faster than Oct2a is observed, but this is present in both the 'sense' and 'antisense' lanes.

The *in vitro* translation was repeated using radioac-



```
                616                                                                  678
Oct1        SLNPGTLSGALSPALMSNSTVATIQALASGGSLPITSLDATGNLVF   ANAGGAPNIVTAPLFLN
                 ||  ||||||||  ||||||||||||||||||| |||||      | | | ||| ||||||
Oct2b       GSTMVGLSSGLSPALMSNNPLATIQALASGGTLPLTSLDGSGNLVLGAAGAAPGSPSLVTSPLFLN
            451                                                                  516

                679                                                                  743
Oct1        PQNLSLLTSNP VSLVSAAAASAGNSAPVASLHATSTSAESIQNSLFTVASASGAASTTTTA SKAE
               | |||  | | |||||||    |    |  ||||  |   |   ||  |        | ||||
Oct2b       HTGLPLLSAPPGVGLVSAAAAAVAASISSKSPGLSSSSSSSSSSTCSDVAAQTPGGPGGPEAGSKAE
            517                                                                  583
```

Fig. 2. Amino acid comparison of the carboxy ends of Oct1 and Oct2b. The last 133 amino acids of Oct2b and the last 128 amino acids of Oct1 are aligned. Small gaps were introduced in the Oct1 sequence for maximum alignment. For the comparison two conservative changes, threonine to serine and leucine to isoleucine were also taken into account.

tive methionine to label the newly synthesized proteins. After resolving the products in protein–SDS gels, gels were dried and autoradiographed. The mR4 encoded protein runs with an apparent M_r of approximately 70×10^3 as predicted (data not shown). Taken together these results indicate that the protein encoded by the mOct2b open reading frame is the previously described OTF-2B factor (Schreiber *et al.* 1988).

Oct2a and Oct2b are differentially spliced products of the same gene

Preliminary Southern analysis of human and mouse chromosomal DNA with the cloned cDNAs as probes showed that Oct2 is probably a large single copy gene (our unpublished data; Müller *et al.* 1988; Scheidereit *et al.* 1988). In addition, apart from the 48 bp and 74 bp deletions/insertions, the isolated mouse Oct2 cDNA clones are identical and colinear with the human sequences. These data strongly suggest that Oct2a and Oct2b are differentially spliced products of the same gene.

In order to confirm this and to see how the different exons are organized, the structure of the mouse Oct2 gene was analysed in chromosomal clones isolated by using the synthetic oligonucleotides spanning the human Oct2 homeobox as a probe. Four independent overlapping clones were isolated (termed mCR1 to 4) spanning an area of approximately 18 kb. The restriction enzyme map of this area is shown in Fig. 4. Subcloning, Southern analysis of these clones using different parts of the cDNAs as probes and sequencing of the various chromosomal fragments identified 10 exons of the mouse Oct2 gene. It should be noted that this is not the complete structure of the gene, but simply the organization of the intron–exon junctions of the mOct2a and 2b cDNAs from position 520 in the sequence shown in Fig. 1 to the end. Nevertheless, the structure of this part of the gene was sufficient at this point to enable us to explain how the differences between mOct2a and 2b arise. It is noteworthy that about 1500 bases of transcribed mRNA sequence are spread in an area of around 9 kb. The DNA structure of this part of the gene is characterized by the large number of exons ranging in size from 67 to more than 550 bp and corresponding introns ranging in size from 75 bp to about 2 kb. The intron–exon organization is

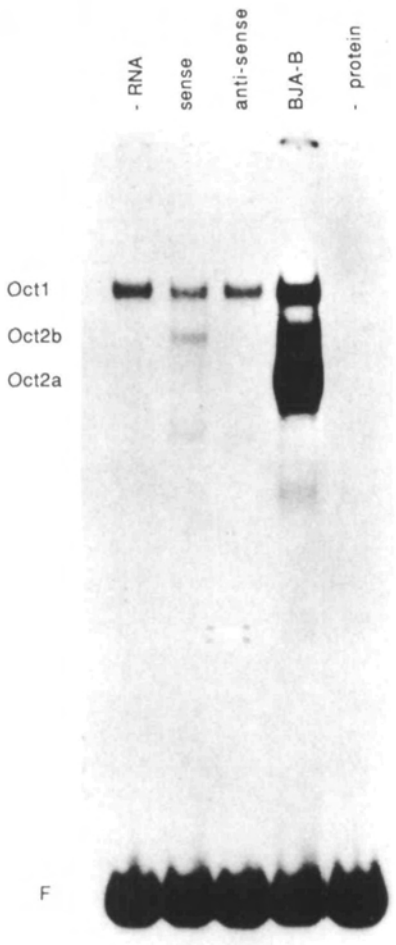


Fig. 3. The insert of clone mR4 encodes the Oct2b octamer binding protein. RNA from the coding (sense) and non-coding (antisense) strand of clone mR4 was synthesized *in vitro* and then translated in rabbit reticulocyte lysates. The translation products were then used in the electrophoretic mobility shift assay (EMSA) with a radioactively labelled oligonucleotide containing the octamer motif (see Materials and methods). In the lane, −RNA, the reticulocyte extract was incubated without addition of exogenous RNA. Sense and antisense lanes are reticulocyte extracts where the coding and non-coding RNA strands of the mR4 insert were translated, respectively. BJA-B is a nuclear extract from a human mature B-cell line. In −protein lane, the oligonucleotide was incubated without any extract. The expected pattern of the BJA-B nuclear extract is observed and the protein complexes due to Oct1, Oct2a and Oct2b are indicated. F marks the position of the unbound free oligonucleotide.

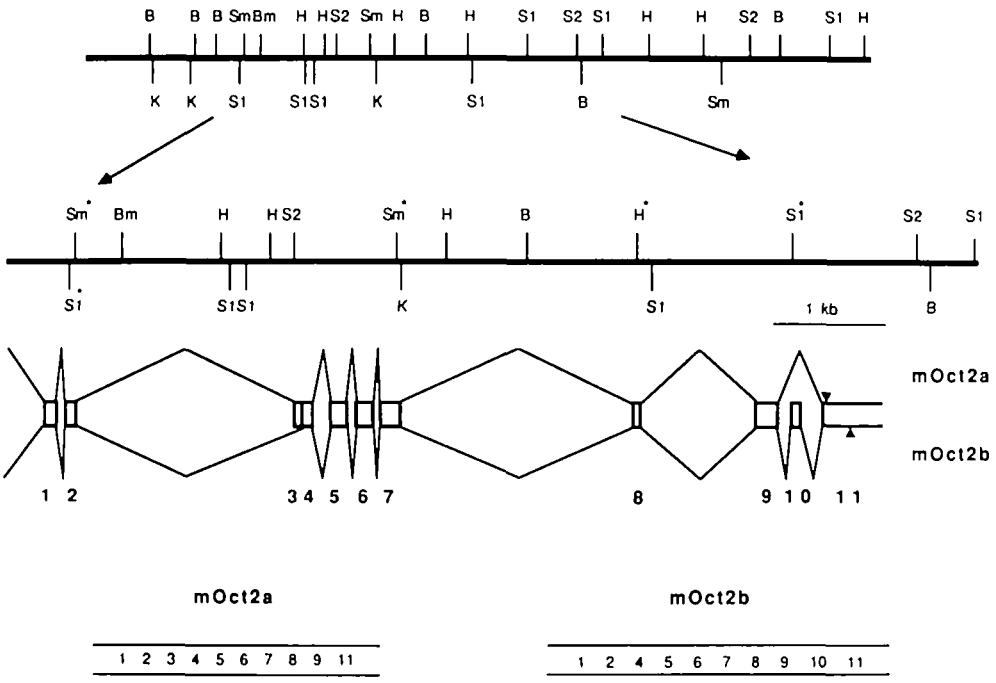


Fig. 4. Partial structure of the mouse Oct2 gene. The restriction map of the 18 kb of the mouse genome that contain part of the Oct2 gene is shown on top. Below is an expanded area that contains the Oct2 gene exons mapped so far. Arrows with letters mark the location of restriction sites. B *Bgl*II; Bm, *Bam*HI; H, *Hin*dIII; K, *Kpn*I; S1, *Sac*I; S2, *Sac*II; Sm, *Sma*I. Restriction sites marked with asterisks are also present in the cDNAs. The exons are denoted by open boxes. The splicing patterns of mOct2a and mOct2b are shown above and below the exons, respectively. The numbering of 1 to 11 is used for better visualization of the structure of the mOct2a and 2b transcripts after the differential splicing (shown at the bottom of the figure). Open rectangles indicate that the structure is incomplete. Black arrowheads above and below exon 11 mark the location of stop codons in Oct2a and Oct2b transcripts, respectively.

drawn in scale in Fig. 4. The location of the intron–exon junctions in the cDNAs are shown with arrows in Fig. 1.

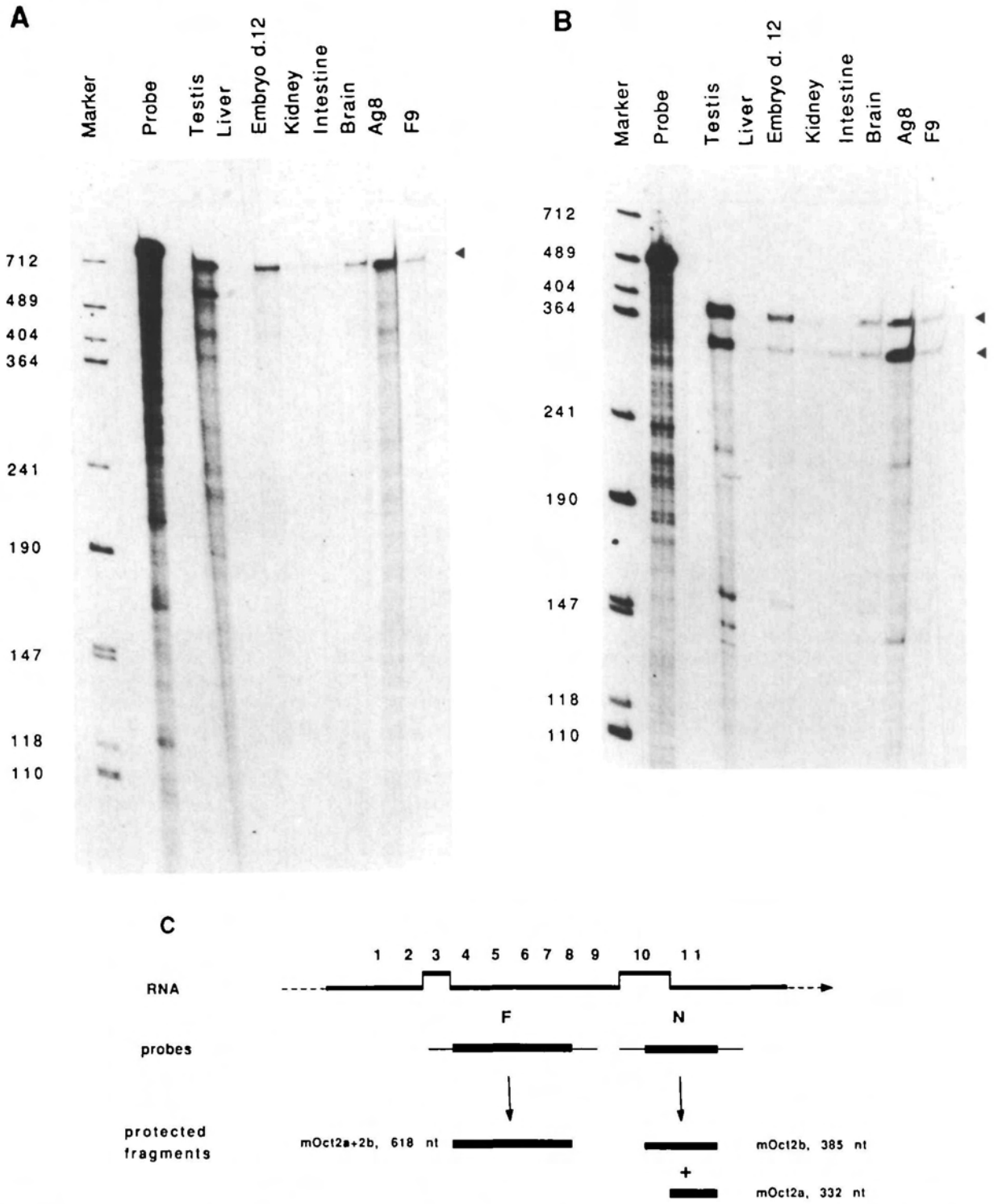
Sequencing of the chromosomal clones shows that the two differences between mOct2a and mOct2b are derived as follows (Fig. 4). In the first case, the 48 bp deletion in mOct2b is generated by the use of a second splice acceptor site 48 bp downstream of the one used in the case of mOct2a (exon 4 and 3 4, respectively, in Fig. 4). In both cases, the splice site fits the consensus mouse sequence (Shapiro and Senapathy, 1987). The second difference is derived by splicing in mOct2b of an extra exon (number 10 in Fig. 4) which in the case of mOct2a is located in the intron area between exons 9 and 11. Thus, although exon 11 is the same in both transcripts, a different protein is produced since exon 10 shifts the open reading frame of mOct2b. The two transcripts now use two different stop codons (indicated by arrowheads in Fig. 4).

Expression of Oct2a and Oct2b

One of the main objectives of this work was to determine whether the Oct2 gene is expressed in tissues other than B lymphocytes and in particular during embryogenesis as was suggested by the EMSA studies (Schöler *et al.* 1989*a*). With the identification of the novel Oct2b transcript, it also was important to test if there are differences in the pattern of expression of Oct2a and Oct2b.

Total RNA was isolated from various tissues, mouse embryos and cell lines, and the presence of Oct2 transcripts was analysed by using the RNase protection assay. For this purpose, two probes were used. The first ontains a 618 bp fragment isolated from the λgt11 clone mR2. This fragment covers the entire POU-specific and homeoboxes common in both mOct2a and mOct2b transcripts (fragment F in Fig. 5C, see also Materials and methods). A prominent band of the expected size is protected with the mouse B-cell line Ag8 RNA (Fig. 5A). The same band is protected with RNA from testis, kidney, intestine, brain, mouse day 12 embryo and F9 cells. Liver seems negative although a faint band appears after prolonged film exposures. In testis, an equally strong band of around 520 bp is shown, indicating that an additional transcript exists in testis that contains a truncated POU-domain. Of all the tissues tested, B-cells appear to contain the highest amounts of mOct2a and 2b transcripts, since in this experiment 10 times less B-cell RNA was used compared to the other tissues.

The probe used in the experiment described in Fig. 5A cannot distinguish between the mOct2a and mOct2b transcripts. For this purpose, a second DNA fragment was used containing the region from the *Sac*I site in position 1590 to the *Nae*I site at position 1975 of mOct2b (numbering as in Fig. 1, fragment N in Fig. 5C). The *Sac*I site is located within the 74 bp insertion and is not found in mOct2a. As a result, 385 nt



and 332 nt fragments should be protected by the mOct2b and mOct2a transcripts, respectively. As shown in Fig. 5B, two bands of the expected size are protected, the upper one representing the mOct2b transcript and the lower one the mOct2a.

Both transcripts are very abundant in B-cells (again 10 times less RNA was used for the Ag8 lane than in the other lanes). Interestingly, the ratio of mOct2a to 2b varies among tissues. mOct2a is more abundant than mOct2b in B-cells and intestine. In F9 cells and in kidney, equal amounts of the two transcripts exist. In testis, day 12 embryo and to some extent in brain, it appears that mOct2b transcripts are more abundant. At present, the significance of this observation is unclear. Several less-intense bands are also protected indicating that transcripts probably highly homologous but not identical to Oct2 exist. The fact that the ratio of mOct2a to 2b varies between tissues and the transcripts are detected in the F9 cell line rules out the possibility of B lymphocyte contamination of the RNA preparations.

Fig. 5. Expression of mOct2a and mOct2b in various tissues. RNase protection experiments were performed using total RNA and radioactive antisense RNA probes. The RNA source is indicated above the lanes. Ag8 is a mouse B-cell line; F9 is a mouse embryonic teratocarcinoma cell line. In each case, 40 μg of total RNA were used except in lane Ag8 where only 4 μg were used. In panel A, the probe is a 720 nt fragment in which 618 nt covering the POU-domain common in both Oct2a and Oct2b transcripts are protected (probe F). In panel B the probe is a 490 nt fragment which contains a 385 nt sequence from mOct2b (position 1590 to 1975 in Fig. 1; probe N). A 385 nt fragment is protected by the mOct2b transcript and only a 332 nt fragment by the mOct2a transcript (Upper and lower arrowheads in panel B). The radioactive marker shown in the left lane in both A and B is Bluescript plasmid DNA restricted with *Hpa*II. Panel C is a schematic representation of the RNase protection experiments shown in panels A and B. The top line represents the transcript derived from the Oct2 gene. Broken lines at the two ends indicate that the complete structure of the 7.5 kb transcript is not yet known. The arrow denotes the 5′ to 3′ orientation. The 48 nt deletion and 74 nt insertion in the mOct2b clone mR4 are elevated. Numbers above show the location of the exons as marked in Fig. 4. The solid boxes in probes F and N represent the Oct2 sequences while the thin lines represent the presence of Bluescript vector sequences. The protected fragments are shown as solid boxes at the bottom, 618 nt for probe F, 385 and 332 nt for probe N (see also Materials and methods).

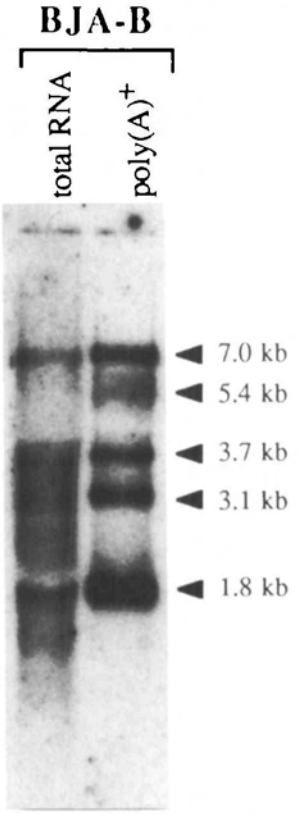


Fig. 6. Northern blot analysis of RNA from BJA-B cells hybridized with fragment F (cDNA probe). 25 μg of total RNA (lane 1) and 3.5 μg of poly $(A)^+$ RNA (lane 2) were denatured, seperated on a 1.2 % agarose-formaldehyde gel and blotted with 10×SSC. Hybridization was carried out as described in Materials and methods. The approximate sizes of the transcripts are given.

Expression of the Oct2 gene during embryogenesis and in adult brain

The fragment F used in the RNase protection experiment (Fig. 5A) was also used as a probe for *in situ* hybridization of mouse day 12 embryo sections.

A Northern blot of BJA-B mRNA probed with fragment F revealed five transcripts ranging from 1.8 to 7.0 kb (Fig. 6). This is consistent with other analyses of BJA-B RNA with Oct-2 cDNA probes, hybridizing also to multiple transcripts in a similiar pattern (Müller *et al.* 1988; Staudt *et al.* 1988; Scheidereit *et al.* 1988). The same pattern was also observed when fragment N was used as a probe (data not shown). This demonstrates that fragment F does not cross-hybridize with other POU- and homeobox-genes. Accordingly, fragment F is an appropiate probe to study the expression of the Oct-2 gene in mouse embryo as a first step towards identifying its role in mouse development.

The Oct2 gene is widely expressed in the developing nervous system (Fig. 7). In particular, expression is found in the neural tube and in the developing brain, mainly in diencephalon, mesencephalon, metencephalon and myelencephalon. Interestingly, the expression level of Oct2 is not homogeneous but stronger in particular areas. For example, in the diencephalon, strong hybridization is observed in the areas where the optic chiasma and the mammillary bodies will develop, which corresponds to the observed hybridization in suprachiasmatic and medial mammillary nuclei in the adult brain (see below). In addition, reproducible expression was found in an area running parallel and ventrally to the vertebrae in the region of the developing dorsal aorta, but, at present, this structure has not been conclusively identified (area marked by arrowheads in Fig. 7, panel A).

Although the expression of Oct2 is widespread in the developing brain, in the adult brain expression is confined to very small and distinct areas like the medial mammillary nuclei, the olfactory bulb and tract and the CA1–CA4 areas of the hippocampus. Very strong expression was also observed in the suprachiasmatic nuclei (data not shown). Although piriform cortex appears positive, hybridization was also observed in the negative control at a lower level. In addition, expression is confined only to the CA1–CA4 areas of the hippocampus. The dentate gyrus appears positive due to intensive staining by toluidine blue.

Octamer-binding proteins in neuronal and glial cells

While *in situ* hybridization gives a first indication of which areas of the brain the Oct2 gene is expressed, it is not accurate enough to distinguish the cell types that express the gene. As an initial approach to determine in which cells Oct2 proteins are present, mixed brain nuclei were isolated and further fractionated into neuronal and glial nuclei. As shown earlier, the nuclear

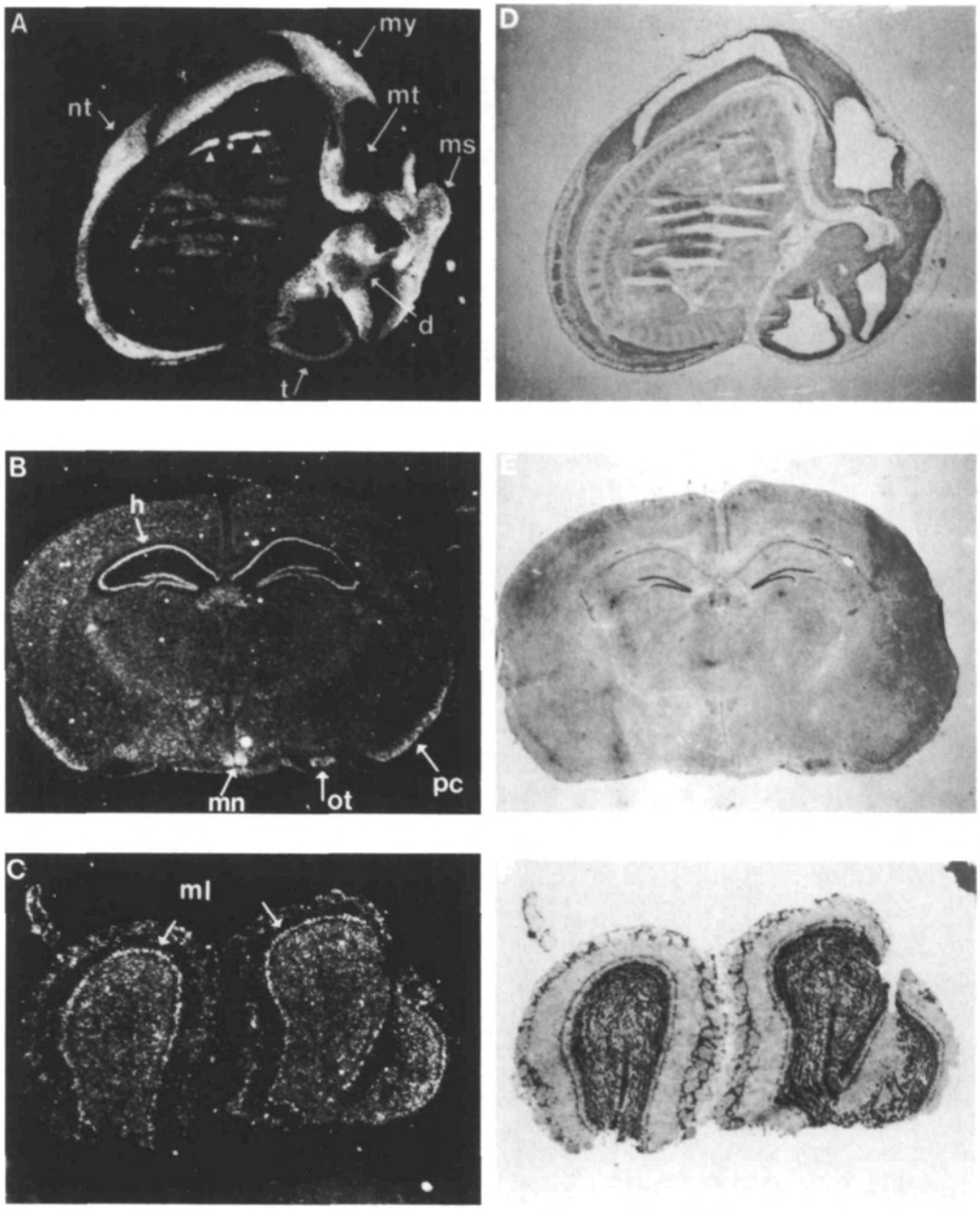


Fig. 7. Expression of the Oct2 gene during mouse embryogenesis and in adult brain. Mouse day 12 embryo and adult mouse brain sections were hybridized with probe F (see Fig. 5C). Left panels A, B, C are dark-field pictures while right panels D, E, F are the corresponding bright field. (A, D) Sagittal section of mouse day 12 embryo. Arrows point to areas of the developing nervous system; t, telencephalon; d, diencephalon; ms, mesencephalon; mt, metencephalon; my, myelencephalon; nt, neural tube. Arrowheads mark the positive area around the dorsal aorta which has not been conclusively identified. (B, E) Section through adult brain. mn, medial mammillary nuclei; ot, olfactory tract; cp, piriform cortex; h, hippocampus. Section through the olfactory bulb. Arrows indicate the mitral and granular cell layer (ml).

fractions obtained by this procedure are at least 90 % pure and are representative of the nuclei of medium-sized neurons and dark- and medium-type oligodendrocytes of rat cerebrum (Stoykova *et al.* 1985). The presence of the different octamer-binding proteins was analysed in the electrophoresis mobility shift assay using the fragment containing the octamer motif (Fig. 8). The results show that both cell types contain the same array of proteins as the whole brain nuclear extract (Schöler *et al.* 1989*a*). Oct2a is found in both cell types but is more abundant in glial cells than in neurons. Oct1 is present in only very small amounts in neurons. Oct3 shows the opposite pattern being more abundant in neuronal than glial cells. Amounts of Oct 8, 9 and 10 are reduced in both cell types.

Discussion

A large number of proteins in tissues, cell lines and

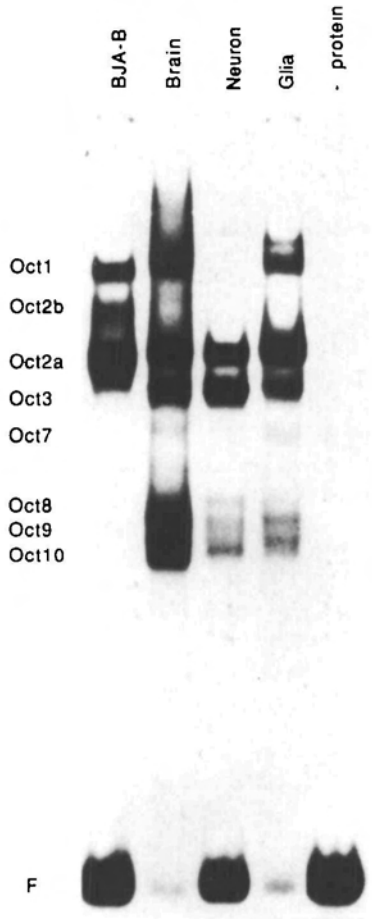


Fig. 8. Distribution of octamer-binding proteins in neuronal and oligodendroglial cells. A radioactively labelled oligonucleotide containing the octamer motif was incubated with 6 μg nuclear extracts from brain, neurons, glial cells, BJA-B (a human B-cell line) or without extract (−protein lane). The observed complexes are named on the left side according to the nomenclature of Schöler *et al.* (1989*a*). F, free oligonucleotide.

embryos bind to the octamer motif ATTTGCAT (Singh *et al.* 1986; Landolfi *et al.* 1986; Staudt *et al.* 1986; Rosales *et al.* 1987; Scheidereit *et al.* 1987; Barberis *et al.* 1987; Cox *et al.* 1988; Lenardo *et al.* 1989; Schöler *et al.* 1989*a*; Schöler *et al.* 1989*b*). These proteins interact specifically with this transcriptional regulatory sequence since single base exchanges prevent binding (Staudt *et al.* 1986; Schöler *et al.* 1989*a*). There are several possibilities about how this large family of transcriptional factors can be produced. For example, every one of these proteins might be the product of a distinct gene. Alternatively, post-transcriptional and post-translational modifications may give rise to some of the diversity observed in octamer-binding factors. While the cloning of Oct1 and Oct2a showed that these two are the products of distinct genes (Sturm *et al.* 1988*b*; Ko *et al.* 1988; Müller *et al.* 1988; Scheidereit *et al.* 1988; Clerc *et al.* 1988), the results presented here demonstrate that two of the octamer-binding proteins, namely Oct2a and Oct2b, are generated by differential splicing of the same gene. The cloning of the remaining factors, Oct3 to Oct10, will reveal their relation to Oct1 and Oct2 and whether the third proposed mechanism, i.e. post-translational modifications, also contributes to the generation of the family of octamer-binding proteins. Recently, it was shown that a large family of POU-domain-containing genes exists which could be placed into four groups (He *et al.* 1989). These results demonstrate that, besides the large family of octamer-binding proteins, other multimember families exist that bind sequences similar but not identical to the octamer motif.

The two factors, Oct2a (463 aa) and Oct2b (583 aa), share common features but also display certain differences. They have in common the POU-domain responsible for DNA binding and the first 167 amino acids of the amino terminus containing a glutamine-rich domain. Oct2b lacks 16 amino acids in the end of this domain and upstream from the beginning of the POU-domain (which starts at amino acid 178, Fig. 1). This deletion occurs because in the mOct2b transcript another splice acceptor site is used, 48 bp downstream from the one used in mOct2a. This deletion is not specific for Oct2b since human Oct2a clones have been isolated that also lack these 16 amino acids (Scheidereit *et al.* 1988).

Oct2a and Oct2b also share the 111 amino acids downstream from the homeo-specific part of the POU-domain, residues 341 to 451 (Fig. 1). This area contains the four properly spaced leucines that might be involved in protein–protein homo- and/or heterodimerizations (Landschultz *et al.* 1988). At this point, there is a 74 bp insertion in mOct2b which results in a new open reading frame generating a different carboxy end. This difference between Oct2a and 2b arises by the splice in mOct2b of a new exon, which otherwise resides in an intron of the mOct2a transcript. The carboxy end of Oct2b is very similar to the one of Oct1 and they might have similar functions. Clerc *et al.* have described another Oct2 transcript which stops after the glycine residue at position 451 in Fig. 1 (clone 3–1 in Clerc *et al.* 1988). The sequence that follows in their case is homologous to the sequence in the mouse intron following exon 9 in Fig. 4. It seems that this Oct2 gene transcript includes this intron sequence and a downstream splice acceptor site is used.

It is possible that more, presently unidentified, differentially spliced products are derived from the Oct2 gene which might not be resolved by the two probes used in the RNase protection experiment. More transcripts, and consequently more protein products, might be hidden in the protected bands. In light of this, the RNA results as well as the *in situ* hybridization might not represent specifically Oct2a or Oct2b but the expression of the Oct2 gene in general. This issue will only be resolved when the complete array of the Oct2 gene transcripts is described.

It is interesting that the differential splicing scheme of the Oct2 gene resembles the patterns seen in the sex-specific splicing of many *Drosophila* genes, namely *Sxl*, *tra* and *tra-2*, which are involved in a hierarchical way in

sex determination (for a review and references see Baker, 1989). In this case, downstream acceptor sites or splicing of additional exons create different transcripts in males and females. The products of one differential splice regulate the splicing pattern of the genes below in the hierarchy. Moreover, the splice in male flies is the default splicing while the splice pattern in females requires other proteins which probably interact with the splicing machinery. It will be interesting to check if either the mOct2a or the mOct2b splicing pattern is the default one or if both of them are regulated specifically by other proteins. Since the ratio of Oct2a to Oct2b is different in the various tissues, their splicing might be regulated in a tissue-specific manner.

Differential splicing is also responsible for the generation of a family of peptides interacting with the transcriptional regulatory element known as the CCAAT box (Mermod *et al.* 1989). Alternate splicing of the same gene might turn out to be a widespread mechanism for generating families of related multifunctional regulatory proteins.

It is noteworthy that there is no direct correlation between the levels of the Oct2 RNAs as detected in the RNase protection and the levels of Oct2a and Oct2b binding activities as detected in the EMSA. For example, while mOct2a and 2b transcripts are found in F9 cells, no Oct2a and 2b complexes were seen in the EMSA (Schöler *et al.* 1989*a*,*b*). In addition, while the amount of Oct2a RNA is lower in kidney than in brain and embryo, the levels of Oct2a complexes are comparable in the EMSA (see Fig. 5B and Schöler *et al.* 1989*a*). Furthermore, while in brain the amounts of mOct2a and 2b RNAs are about equal, and in embryos RNA levels of Oct2b are higher than 2a, no Oct2b binding is observed *in vitro* (compare Fig. 5B and Fig. 8). It is possible that Oct2b does not bind to the octamer motif as strongly as Oct2a. In addition, apparently other mechanisms operating post-transcriptionally are responsible for this disrepancy. Alternatively, proteins like Oct3 to Oct10 are differentially spliced products of the Oct2 gene as well and are responsible for the protected bands in the RNase protection experiments. Cloning of the remaining octamer-binding proteins will only resolve this issue.

While this manuscript was in preparation, He *et al.* also showed *in situ* hybridization results with an Oct2 probe in embryos and adult brain (He *et al.* 1989). Although basically they made the same conclusions, there are some differences concerning the expression in adult brain. For example, they did not show expression in the olfactory bulb and hippocampus, although they also observed high expression in suprachiasmatic and medial mammillary nuclei. In addition, they did not see expression in testis in their RNA analysis. At this point, we do not have an explanation for these discrepancies except that different probes and experimental conditions were used.

The EMSA assay shows that Oct2a is present in both neuronal and glial cells but it is more abundant in glial cells. Staudt *et al.* have also observed high levels of expression of Oct2 in a glioma cell line (Staudt *et al.* 1988). Eventually, only a specific antibody will provide a more detailed analysis about the cell types where Oct2 proteins are found. In particular, antibodies that specifically recognize the different polypeptides will be essential in describing where the Oct2 proteins are present. These antibodies could be used to analyse in more detail the expression pattern of these factors during mouse embryogenesis and in adult tissues. The analysis at the protein level will also help to resolve the discrepancy between the binding activity and the RNA levels. For example, if protein levels follow the RNA levels, then binding is regulated post-translationally. If, on the other hand, the protein levels reflect the binding pattern, then post-transcriptional and translational controls may be acting.

Oct1 is involved in cell-cycle regulation of gene expression as well as in DNA replication. This finding correlates with high amounts of Oct1 protein in fast growing cells (Sturm *et al.* 1988*b*). Consequently, it is interesting that Oct1 is not present in considerable amounts in neurons, while its level in glial cells is comparable to other cell types. Neurons in brain have lost their ability to divide and do not regenerate after damage. In contrast, glial cells are able to regenerate, although they do not divide under normal conditions.

The results presented here indicate that Oct2 proteins play an important role in regulating transcription in the developing nervous system, adult brain, testis, kidney and intestine in addition to regulating immunoglobulin gene expression. The expression of Oct2 gene in the developing nervous system in embryos correlates with our previous observation that brain and embryo nuclear extracts have identical patterns of octamer complexes in the EMSA (Schöler *et al.* 1989*a*). One prediction of this finding is that the other brain octamer-binding proteins, namely Oct3 and Oct7 to Oct10, may be expressed also in the developing embryonic nervous system.

Future research will need to focus on the function of the different octamer proteins. It will be interesting to see if Oct2a and Oct2b are redundant and have the same function or if they play different roles. The fact that the octamer motif is found both in promoters close to the TATA box and in enhancer sequences thousands of bases away, raises the possibility that one protein, Oct2a or Oct2b, binds to promoter and the other to enhancer sequences. Further insights into the regulation of gene expression by this family of transcriptional factors will come from analysis of the mechanisms responsible for differential splicing of the Oct2 gene.

We are greatly indebted to H. R. Schöler, U. Deutsch and G. Dressler for help with the experimental work. We thank B. Föhring for help with the computer analysis, R. Altschäffel for photographic work and H. P. Geithe for the synthesis of oligonucleotides. We also to thank A. Püschel and H. R. Schöler for helpful comments on the manuscript. A.H., M. G. and T.N. were supported by the Max-Planck Society. A.H. and A.S were recipients of post-doctoral fellowships by the A. von Humboldt Organization. This work was supported by the

Max-Planck Society and by the Bundesministerium für Forschung und Technologie (BMFT).

References

ARES JR, M., MANGIN, M. AND WEINER, A. M. (1985). Orientation dependent transcriptional activation of a human U2 snRNA gene. *Molec. cell. Biol.* **5**, 1560–1570.

BAKER, B. S. (1989). Sex in flies: the splice of life. *Nature* **340**, 521–524.

BARBERIS, A., SUPERTI-FURGA, G. AND BUSSLINGER, M. (1987). Mutually exclusive interaction of the CCAAT-binding factor and of a displacement protein with overlapping sequences of a histone gene promoter. *Cell* **50**, 347–359.

BERGMAN, Y., RICE, D., GROSSCHELD, R. AND BALTIMORE, D. (1984). Two regulatory elements for immunoglobulin κ light chain gene expression. *Proc. natn. Acad. Sci. U.S.A.* **81**, 7041–7045.

BODNER, M., CASTRILLO, J.-L., THEILL, L. E., DEERINCK, T., ELLISMAN, M. AND KARIN, M. (1988). The pituitary-specific transcription factor GHF-1 is a homeobox-containing protein. *Cell* **55**, 505–518.

CARBON, P., MURGO, S., EBEL, J.-P., KROL, A., TEBB, G. AND MATTAJ, I. W. (1987). A common octamer motif binding protein is involved in the transcription of U6 snRNA by RNA polymerase III and U2 snRNA by RNA polymerase II. *Cell* **51**, 71–79.

CHIRGWIN, J. M., PRZYBYLA, A. E., MACDONALD, R. J. AND RUTTER, W. J. (1979). Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. *Biochemistry* **18**, 5294–5299.

CLERC, R. G., CORCORAN, L. M., LEBOWITZ, J. H., BALTIMORE, D. AND SHARP, P. A. (1988). The B-cell-specific Oct-2 protein contains POU box- and homeo box-type domains. *Genes Dev.* **2**, 1570–1581.

COX, P. M., TEMPERLEY, S. M., KUMAR, H. AND GODING, C. R. (1988). A distinct octamer-binding protein present in malignant melanoma cells. *Nucleic Acids Res.* **23**, 11 047–11 056.

DIGNAM, J. D., LEBOWITZ, R. M. AND ROEDER, R. G. (1983). Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Res.* **11**, 1475–1489.

DONY, C. AND GRUSS, P. (1987). Specific expression of the *Hox 1.3* homeo box gene in murine embryonic structures originating or induced by the mesoderm. *EMBO J.* **6**, 2965–2975.

DRESSLER, G. R. AND GRUSS, P. (1988). Do multigene families regulate vertebrate development? *Trends Genet.* **4**, 214–219.

EPHRUSSI, A., CHURCH, G. M., TONEGAWA, S. AND GILBERT, W. (1985). B-lineage-specific interactions of an immunoglobulin enhancer with cellular factors in vivo. *Science* **227**, 134–140.

FALKNER, F. G. AND ZACHAU, H. G. (1984). Correct transcription of an immunoglobulin κ gene requires an upstream fragment containing conserved sequence elements. *Nature* **310**, 71–74.

FINNEY, M., RUVKUM, G. AND HORVITZ, H. R. (1988). The *C. elegans* cell lineage and differentiation gene *unc-86* encodes a protein containing a homeo domain and extended sequence similarity to mammalian transcription factors. *Cell* **55**, 757–769.

FLETCHER, C., HEINTZ, N. AND ROEDER, R. G. (1987). Purification and characterization of OTF-1, a transcription factor regulating cell cycle expression of a human H2b gene. *Cell* **51**, 773–781.

GEHRING, W. J. (1987). Homeo boxes in the study of development. *Science* **236**, 1245–1252.

GERSTER, T., MATTHIAS, P., THALI, M., JIRICNY, J. AND SCHAFFNER, W. (1987). Cell type-specificity elements of the immunoglobulin heavy chain gene enhancer. *EMBO J.* **6**, 1323–1330.

HATZOPOULOS, A. K., SCHLOKAT, U. AND GRUSS, P. (1988). Enhancers and other *cis*-acting regulatory sequences. In *Transcription and Splicing* (eds B. D. Hames and D. M. Glover), pp. 43–96. IRL Press, Oxford, Washington DC.

HE, X., TREACY, M. N., SIMMONS, D. M., INGRAHAM, H. A., SWANSON, L. W. AND ROSENFELD, M. G. (1989). Expression of a large family of POU-domain regulatory genes in mammalian brain development. *Nature* **340**, 35–42.

HERR, W., STURM, R. A., CLERC, R. G., CORCORAN, L. M., BALTIMORE, D., SHARP, P. A., INGRAHAM, H. A., ROSENFELD, M. G., FINNEY, M., RUVKUM, G. AND HORVITZ, H. R. (1988). The POU domain: a large conserved region in the mammalian *pit*-1, *oct*-1, *oct*-2, and *Caenorhabditis elegans unc*-86 gene products. *Genes Dev.* **2**, 1513–1516.

HOGAN, B., CONSTANTINI, F. AND LACY, E. (1986). In *Manipulating the Mouse Embryo: A Laboratory Manual*, pp. 228–242. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

INGRAHAM, H. A., CHEN, R., MANGALAM, H. J., ELSHOLTZ, H. P., FLYNN, S. E., LIN, C. R., SIMMONS, D. M., SWANSON, L. AND ROSENFELD, M. G. (1988). A tissue-specific transcription factor containing a homeodomain specifies a pituitary phenotype. *Cell* **55**, 519–529.

KO, H.-S., FAST, P., MCBRIDE, W. AND STAUDT, L. M. (1988). A human protein specific for the immunoglobulin octamer DNA motif contains a functional homeobox domain. *Cell* **55**, 135–144.

KOZAK, M. (1986). Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell* **44**, 283–292.

LABELLA, F., SIVE, H. L., ROEDER, R. G. AND HEINTZ, N. (1988). Cell-cycle regulation of a human histone H2b gene is mediated by the H2b subtype-specific consensus element. *Genes Dev.* **2**, 32–39.

LANDOLFI, N. F., CAPRA, J. D. AND TUCKER, P. W. (1986). Interaction of cell-type-specific nuclear proteins with immunoglobulin V_H promoter region sequences. *Nature* **323**, 548–551.

LANDSCHULTZ, W. H., JOHNSON, P. F. AND MCKNIGHT, S. L. (1988). The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. *Science* **240**, 1759–1764.

LENARDO, M., PIERCE, J. W. AND BALTIMORE, D. (1987). Protein-binding sites in Ig gene enhancer determine transcriptional activity and inducibility. *Science* **236**, 1573–1577.

LENARDO, M. J., STAUDT, L., ROBBINS, P., KUANG, A., MULLIGAN, R. C. AND BALTIMORE, D. (1989). Repression of the IgH enhancer in teratocarcinoma cells associated with a novel octamer factor. *Science* **243**, 544–546.

MANGIN, M., ARES JR, M. AND WEINER, A. M. (1986). Human U2 small nuclear RNA genes contain an upstream enhancer. *EMBO J.* **5**, 987–995.

MANIATIS, T., FRITSCH, E. F. AND SAMBROOK, J. (1982). *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

MASON, J. O., WILLIAMS, G. T. AND NEUBERGER, M. S. (1985). Transcription cell type specificity is conferred by an immunoglobulin V_H promoter that includes a functional consensus sequence. *Cell* **41**, 479–487.

MATTAJ, I. W., LIENHARD, S., JIRICNY, J. AND DEROBERTIS, E. M. (1985). An enhancer like sequence within the *Xenopus* U2 gene promoter facilitates the formation of stable transcription complexes. *Nature* **316**, 163–167.

MERMOD, N., O'NEIL, E. A., KELLY, T. J. AND TJIAN, R. (1989). The proline-rich transcriptional activator of CTF/NF-1 is distinct from the replication and DNA binding domain. *Cell* **58**, 741–753.

MÜLLER, M. M., RUPPERT, S., SCHAFFNER, W. AND MATTHIAS, P. (1988). A cloned octamer transcription factor stimulates transcription from lymphoid-specific promoters in non-B cells. *Nature* **336**, 544–551.

ODENWALD, W. F., GARBEN, J., ARNHEITER, H., TOURNIER-LASSERVE, E. AND LAZZARINI, R. A. (1989). The *Hox-1.3* homeo box protein is a sequence-specific DNA-binding phosphoprotein. *Genes Dev.* **3**, 158–172.

O'NEIL, E. A. AND KELLY, T. J. (1988). Purification and characterization of nuclear factor III (origin recognition protein C) a sequence-specific DNA binding protein required for efficient initiation of adenovirus DNA replication *J. biol. Chem.* **263**, 931–937.

PARSLOW, T. G., BLAIR, D. L., MURPHY, W. J. AND GRANNER, D. K. (1984). Structure of the 5′ ends of immunoglobulin genes: A novel conserved sequence. *Proc. natn. Acad. Sci. U.S.A.* **81**, 2650–2654.

PARSLOW, T. G., JONES, S. D., BOND, B. AND YAMAMOTO, K. R.

(1987). The immunoglobulin octanucleotide: independent activity and selective interaction with enhancers. *Science* **235**, 1498–1501.

PRUJIN, G. J. M., VAN DRIEL, W. AND VAN DER VLIET, P. C. (1986). Nuclear factor III, a novel sequence-specific DNA-binding protein from HeLa cells stimulating adenovirus DNA replication. *Nature* **322**, 656–659.

ROSALES, R., VIGNERON, M., MACCHI, M., DAVIDSON, I., XIAO, J. H. AND CHAMBON, P. (1987). *In vitro* binding of cell-specific and ubiquitous nuclear proteins to the octamer motif of the SV40 enhancer and related motifs present in other promoters and enhancers. *EMBO J.* **6**, 3015–3025.

ROSENFELD, P. J., O'NEIL, E. A., WIDES, R. J. AND KELLY, T. J. (1987). Sequence-specific interactions between cellular DNA-binding proteins and the adenovirus origin of replication. *Molec. cell. Biol.* **7**, 875–886.

SANGER, M. P., NICKLEN, S. AND COULSEN, A. R. (1977). DNA sequencing with chain termination inhibitors. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463–5467.

SCHEIDEREIT, C., CROMLISH, J. A., GERSTER, T., KAWAKAMI, K., BALMACEDA, C.-G., CURRIE, R. A. AND ROEDER, R. G. (1988). A human lymphoid-specific transcription factor that activates immunoglobulin genes is a homeobox protein. *Nature* **336**, 551–557.

SCHEIDEREIT, C., HEGUY, A. AND ROEDER, R. G. (1987). Identification and purification of a human lymphoid-specific octamer binding protein (OTF-2) that activates transcription of an immunoglobulin promoter *in vitro*. *Cell* **51**, 783–793.

SCHÖLER, H. R., BALLING, R., HATZOPOULOS, A. K., SUZUKI, N. AND GRUSS, P. (1989*b*). Octamer binding proteins confer transcriptional activity in early mouse embryogenesis. *EMBO J.* **8**, 2551–2557.

SCHÖLER, H. R., HATZOPOULOS, A. K., BALLING, R., SUZUKI, N. AND GRUSS, P. (1989*a*). A family of octamer-specific proteins present during mouse embryogenesis: evidence for germline-specific expression of an Oct factor. *EMBO J.* **8**, 2543–2550.

SCHREIBER, E., MATTHIAS, P., MÜLLER, M. M. AND SCHAFFNER, W. (1988). Identification of a novel lymphoid specific octamer binding protein (OTF-2B) by proteolytic clipping bandshift assay (PCBA). *EMBO J.* **7**, 4221–4229.

SHAPIRO, M. B. AND SENAPATHY, P. (1987). RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucleic Acids Res* **15**, 7155–7174.

SINGH, H., SEN, R., BALTIMORE, D. AND SHARP, P. A. (1986). A nuclear factor that binds to a conserved sequence motif in transcriptional control elements of immunoglobulin genes. *Nature* **319**, 154–158.

STAUDT, L. M., CLERC, R. G., SINGH, H., LEBOWITZ, J. H., SHARP, P. A. AND BALTIMORE, D. (1988). Cloning of a lymphoid-specific cDNA encoding a protein binding the regulatory octamer DNA motif. *Science* **241**, 577–580.

STAUDT, L. M., SINGH, H., SEN, R., WIRTH, T., SHARP, P. A. AND BALTIMORE, D. (1986). A lymphoid-specific protein binding to the octamer motif of immunoglobulin genes. *Nature* **323**, 640–643.

STOYKOVA, A. S., DABEVA, M. D., DIMOVA, R. N. AND HADJIOLOV, A. A. (1979). Ribosomal RNA precursors in neuronal and glial rat brain nuclei. *J. Neurochem.* **33**, 931–937.

STOYKOVA, A. S., DABEVA, M. D., DIMOVA, R. N. AND HADJIOLOV, A. A. (1985). Ribosome biogenesis and nucleolar ultrastructure in neuronal and oligodendroglial rat brain cells. *J. Neurochem.* **45**, 1667–1676.

STURM, R., BAUMRUKER, T., FRANZA JR, B. R. AND HERR, W. (1987). A 100-kD HeLa cell octamer binding protein (OBP100) interacts differently with two separate octamer-related sequences within the SV40 enhancer. *Genes Dev.* **2**, 1147–1160.

STURM, R. A., DALTON, S. AND WELLS, J. R. E. (1988*a*). Conservation of histone H2A/H2B intergene regions: a role for the H2B specific element in divergent transcription. *Nucleic Acids Res.* **16**, 8571–8586.

STURM, R. A., DAS, G. AND HERR, W. (1988*b*). The ubiquitous octamer-binding protein Oct-1 contains a POU domain with a homeo box subdomain. *Genes Dev.* **2**, 1582–1599.

STURM, R. A. AND HERR, W. (1988). The POU domain is a bipartite DNA-binding structure. *Nature* **336**, 601–604.

TANAKA, M., GROSSNIKLAUS, U., HERR, W. AND HERNANDEZ, N. (1988). Activation of the U2 snRNA promoter by the octamer motif defines a new class of RNA polymerase II enhancer elements. *Genes Dev.* **2**, 1764–1778.

WIRTH, T., STAUDT, L. AND BALTIMORE, D. (1987). An octamer oligonucleotide upstream of a TATA motif is sufficient for lymphoid-specific promoter activity. *Nature* **329**, 174–178.

(*Accepted 2 March 1990*)