The octamer binding factor Oct6: cDNA cloning and expression in early embryonic cells

Dies Meijer*, Anneke Graus, Robert Kraay, An Langeveld, Maarten P. Mulder and Gerard Grosveld

MGC Department of Cell Biology and Genetics, Erasmus University, Postbox 1738, 3000 DR Rotterdam, The Netherlands

Received September 3, 1990; Revised and Accepted November 23, 1990

EMBL accession no. X54628

ABSTRACT

We have cloned a cDNA encoding a novel octamer binding factor Oct6 that is expressed in undifferentiated ES cells. Expression of the Oct6 gene is downregulated upon differentiation of these cells by aggregate formation. Furthermore the gene is transiently up regulated during retinoic acid induced differentiation of P19 EC cells, reaching maximum levels of expression one day after RA addition. Sequence analysis of the cDNA encoding the Oct6 protein indicated that the Oct6 gene is a member of the POU-HOMEODOMAIN gene family. The gene expresses a 3 kb mRNA encoding a 449 amino acid protein with an apparent molecular weight of 45 kD. The sequence of the Oct6 POU domain is identical to that of the rat SCIP (Tst-1) gene. The Oct6 expression pattern suggests a role for this DNA binding protein in neurogenesis as well as early embryogenesis.

INTRODUCTION

Cellular differentiation processes are believed to be the result of differential regulation of expression of the genetic content of the cell. That is, genes are turned on and off in response to intra- or extra cellular cues. Regulation of gene expression can operate at the transcriptional and/or posttranscriptional level. Although some well documented examples of regulation at the posttranscriptional level exist (1–4), the main mode of regulation of gene expression is at the transcriptional level. Regulation is achieved by sequence specific interaction of transcription factors with cis-acting DNA elements in gene promoters and enhancers (5). Some of these cis elements are known to bind multiple related proteins (AP-1, CCAAT and Oct binding proteins; (6). An example is the octamer motif ATTTGCAT, which is a well defined cis element found in a variety of promoters and enhancers. The octamer motif in the immunoglobulin heavy chain (IgH) gene promoter and enhancer confers lymphoid specific expression, through binding of the lymphoid specific binding factors Oct2A and Oct2B (7). In contrast, the same motif is also found in the promoters of widely expressed genes, like U snRNA and histone genes. Probably the ubiquitous octamer binding factor

Oct1 is involved in the cell cycle dependent expression of these genes (8). Recently several additional (Oct3–Oct10) octamer binding proteins were reported to be present in various adult mouse tissues and in embryos at different stages of development (9). To date three of these octamer binding proteins (Oct1, Oct2 and Oct3/Oct4) have been defined by cloning of the corresponding cDNAs (10–16). Sequence analysis showed that they are encoded by different genes belonging to the POU-HOMEODOMAIN (short: POU domain) gene family (17).

Here we report on the expression and cloning of a fourth octamer binding factor, Oct6, that is expressed in undifferentiated mouse embryo derived stem (ES) cells. Furthermore the Oct6 gene shows a biphasic expression pattern during retinoic acid (RA) induced neuronal differentiation of P19 embryonal carcinoma (EC) cells. Sequence analysis of a cDNA encoding the Oct6 protein identified a POU domain within the largest open reading frame. This POU domain is identical to the previously described SCIP(tst-1) POU domain (18, 19). The expression pattern of the Oct6 gene suggests a role for this putative transcription factor in early embryogenesis as well as in neurogenesis.

MATERIALS AND METHODS

Cell culture, transfection and in vitro differentiation

P19 EC cells were grown in DMEM/F10 (1:1) medium supplemented with 5% Foetal Calf Serum (FCS), penicillin and streptomycin. The cells were split every 3 to 4 days. CCE ES (20) cells were grown on a feeder layer of lethally irradiated STO fibroblasts in DMEM, supplemented with 10% FCS, non-essential amino acids (Gibco), 0.1 mM β -mercaptoethanol, penicillin and streptomycin. ES cells were subcultured every 2–3 days on fresh feeder layers.

RA induced differentiation of P19 cells was done as described (21) with some modifications. P19 cells were trypsinized and seeded as a single cell suspension in petri dishes to form cell aggregates. To prevent the cells from sticking to the petri dish, the bottom was first covered with a layer of 1% agar in culture medium. After 4 to 5 days, differentiation of the cell aggregates was induced by plating them on tissue culture dishes in full

* To whom correspondence should be addressed

medium, supplemented with 1 μ molar all trans retinoic acid (RA;Sigma). These modifications of the original protocol prevents massive cell death, occurring during aggregate formation in the presence of RA.

DMSO induced differentiation of P19 cells was done as described (21). In vitro differentiation of ES cells to simple embryoid bodies and cystic embryoid bodies was performed according to Robertson (22).

COS-1 cells were grown in DMEM/F10 medium supplemented with 5% FCS, penicillin and streptomycin. COS-1 cells were transfected using the DEAE/Dextran method. The day before transfection cells were subcultured by plating 0.5×10^6 cells in a 10 cm dish. Two hours before transfection the culture medium was refreshed. Cells were washed once with serumfree medium followed by addition of serumfree medium containing 10 μ g of plasmid DNA and 100 μ g/ml of DEAE/Dextran. After two hours the transfection medium was removed and replaced by serumfree medium containing 0.1 mM chloroquine. After another two hours, the chloroquine containing medium was replaced by 14 ml of medium with 5% FCS. Three days after transfection, cells were harvested for preparation of nuclear extracts and isolation of RNA.

Nuclear extract preparation and bandshift assay

Cells were harvested by trypsinization. The single cell suspension was washed once with full medium and once with ice cold PBS. Preparation of nuclear extract (XT) and cytoplasmic RNA was done according to Schreiber et al (23) and Cough (24). Protein concentration of the nuclear XTs was determined, using a direct spectrophotometric method of (25). Protein concentrations of nuclear XTs were typically in the order of 5 mg/ml. Bandshift assays were performed according to Barberis et al. (26), using a 32 P end labelled double stranded synthetic oligonucleotide. The nucleotide sequence of this oligonucleotide used throughout this study was derived from the mouse type I c-abl promoter sequence (27) gagaggaATTTGCATttccaccgaccttc. Typically, 2 fmoles (10,000 cpm) of ds oligonucleotide was incubated with 5 μ g of nuclear protein in 10 mM Hepes (pH 7.9), 60 mM KCl, 1 mM DTT, 1 mM EDTA, 4% Ficoll and 2 μ grams of poly dI-dC at room temperature, for twenty minutes. DNA-protein complexes were separated on a 4% polyacrylamide gel in $0.25 \times$ TBE buffer. Electrophoresis was carried out at 150 Volts for 2 hours, using a protean II slab gel apparatus (Biorad). After electrophoresis, the gel was fixed in 10% methanol, 10% acetic acid for 20 minutes. The dried gel was exposed to autoradiographic film (Fuji RX) without intensifying screen. The proteolytic clipping bandshift assay (PCBA) was done as described (28). Nuclear XT preparation of whole tissue (Brain) was done according to Lichtsteiner et al. (29).

Sequence of the double stranded oligonucleotides are.

abl I	;	TAGGAATTTGCATTTCCGATC
U2	;	TGGTTGTGGCCGTCACAAAGAGGCGGGGCT
		ATGCAAAATAGGGTGTGCCGGGCGAGTCGGG
ad2	;	AGGCCAATATGATAATGAGGGGGT
ad4	;	CACGCCTTATTTGCATATTACT
ICP4	;	AGGGCGGTAATGAGATGCCATGT
abl I(mut)	;	TAGGAATGTTCAATTCCGATC
hep ⁺ oct ⁺	;	CGAGTGCTCATGAATATGCAAAATCAATTGG
hep ⁺ oct ⁻	;	CGAGTGCTCATGAATATCAGTCGCCATTGG

The 300 bp EcoRI fragment used in the binding competition experiment (Fig.6) were derived from the plasmids p6W_sTKCAT (FD) and p6W_sTKCAT(o-) (Fd) (9, 30).

Northern blot analysis

Adult BCBA mice were used as a source of tissues for RNA extraction. Total RNA was extracted, using the method of Auffrey and Rougeon (31). Cytoplasmic RNA from tissue culture cells was prepared, according to the method of Cough (24). Ten μ g of denatured total RNA was separated on a 1% agarose gel containing 0.66 M Formaldehyde (32). RNA was transferred to nitrocellulose (S&S) or Biotrans (Dupont) blotting membranes, by capillary action. Hybridization was done overnight at 42°C in hybridization buffer, containing 50% formamide. Hybridization probes were labeled with [α - 32 P] dATP and [α - 32 P] dCTP using the random hexamer primed labeling method (33). Blots were washed to a final stringency of $0.1 \times$ SSC, 0.1% SDS at 65°C and exposed to Kodak X-AR5 films using an intensifying screen at -80°C.

Cloning and sequencing

Construction of the Brain cDNA library in β gt10 was essentially done as described (34). All screening and cloning manipulations were carried out following standard protocols (35). The library was screened with a mouse Oct2 POU domain probe. Positive clones were selected and subcloned into the EcoRI site of pTZ18 (Pharmacia). Suitable restriction endonuclease sites in the selected clones were used to construct subclones in the M13 cloning vectors mp18 and mp19. Single stranded recombinant phage DNA was used as template for sequencing according to the method of Sanger (36) using [α - 35 S] dATP and the Sequenase (USB) enzyme. Sequence data were compiled and analysed, using the Microgenie software package (Beckman).

RESULTS

A family of octamer binding factors is differentially expressed during EC cell differentiation

Embryonal carcinoma cells provide an excellent in vitro system to study early embryonic events. Depending on the culture conditions, these cells can differentiate into a wide spectrum of different cell types. We employed the P19 EC cell system to investigate expression of octamer binding factors during retinoic acid (RA) and dimethylsulfoxide (DMSO) induced differentiation of these cells. P19 cells can be induced to differentiate into neuroectodermal cell types, mainly astroglia cells and neurons, by high levels of RA. Treatment of P19 cells with DMSO results in mesodermal derivatives, including skeletal and cardiac muscle cell types (37, 38). P19 cells were induced to differentiate following the RA and DMSO protocol and cells were harvested at the indicated day (Fig. 1A,B). Nuclear extracts and cytoplasmic RNA were prepared from the same batch of cells. Using a double stranded oligonucleotide containing an octamer motif, bandshift experiments were performed. Results of these experiments are presented in Fig. 1A,B. Nuclear extracts of undifferentiated P19 cells gave rise to two complexes. The lower mobility complex is found in all cell lines tested and has been labelled Oct1 (also called NFA1, OTF1, NFIII, OBP100 (17)). The higher mobility complex was labelled Oct4 as it comigrated with the previously described ES/EC cell specific complex Oct4 (9) in a coelectrophoresis experiment (data not shown). When the cells are aggregated, two changes are apparent (Fig. 1A, d0). The most dramatic change consists of induction of a second Oct4-like complex with a slightly higher mobility, concomitant with a decreased intensity of the Oct4 complex (small arrows in Fig. 1A). At the same time, a weak complex appears that we refer

Figure 1. Octamer binding factors in nuclear extracts of differentiating P19 and ES cells. Radiolabelled probe was incubated with nuclear extracts of P19 or ES cells at different stages of differentiation. 3–5 μ g of extract was used per lane. The position of the Oct1, Oct4 and Oct6 complexes are indicated. Small arrows locate the Oct4 and Oct4-like complex. Panel A; RA induced differentiation of P19 cells. Nuclear extracts of undifferentiated P19 cells (lane –), aggregates (d0), at day one of RA induction (d1), day two (d2), day four (d4), and day eight (d8). As a control a bandshift of mouse brain nuclear extract is shown. Panel B; DMSO induced differentiation of P19 cells. Indications as in panel A. Panel C; Differentiation of ES cells. Nuclear extracts of undifferentiated ES cells (ES), simple embryoid bodies at day 4 of differentiation (EB4) and cystic embryoid bodies at day 8 of differentiation (EB8). Free probe is indicated by F.

to as the Oct6 complex as it comigrates exactly with the EC/ES cell complex Oct6 (9). Culturing aggregates in tissue culture dishes, in the presence of RA for one day, leads to a further increase in the amount of Oct6 complex and to further reduction of the Oct4 and Oct4-like complexes (d1). At day two (d2), the Oct4 complex is no longer detectable and also the Oct6 complex decreases in intensity. At day four (d4) a new complex appears with a mobility that is slightly slower than the Oct4 complex. Furthermore a complex is observed with a mobility similar to Oct6. When electrophoresis was prolonged, it was evident that this complex has a lower mobility than the Oct6 complex observed in lanes d0, d1 and d2 (data not shown). At day 8, an additional complex is observed that migrates much slower than the Oct6 complex. From day five onwards, many neuronal cells could be identified in the differentiating cultures by virtue of their long processes. Throughout differentiation of the cells, the intensity of the Oct1 complex remains roughly constant. Some of the octamer complexes observed during P19 cell differentiation, are also seen when whole brain nuclear extracts were used. It is clear from these results, that differentiation of P19 cells along the neur ectodermal pathway correlates with a highly complex temporally controlled expression pattern of a family of octamer binding factors.

To check whether the observed differential expression of octamer factors was restricted to RA induced differentiation of P19 cells, we performed a similar experiment, now using DMSO as a differentiation inducing agent. Results of this experiment are shown in Fig. 1B. This experiment differs from the previous one in that DMSO was already present during aggregate formation. Over the timespan studied, the only change observed in the expression of octamer binding factors is the gradual appearance of the higher mobility Oct4-like complex (small arrows Fig. 1B). Again Oct1 expression seems to be relatively constant. Therefore the temporal expression of octamer binding factors appears to be specific for RA induced differentiation of P19 cells.

Figure 2. Expression of mb1 mRNA in differentiating P19 and ES cells. Cytoplasmic RNA and nuclear extracts were prepared from the same batch of cells. Nuclear extracts were used in the bandshift experiments presented in Fig. 1. Twenty μ g of cytoplasmic RNA was denatured, separated on a 1% agarose gel in 0.66M formaldehyde and blotted onto Biotrans (NEN) membranes. The blots were probed with a 300 bp PvuII mb1 probe in hybridization buffer containing 50% formamide at 42°C for 18 hours. Blots were washed to a final stringency of 0.1 $\times$ SSC, 65°C. Exposure was for three days. Indications of the lanes as in Fig. 1.

EC cells are often regarded as closely resembling embryonic stem cells. We were therefore interested to see whether the observed octamer factors were also expressed during ES cell differentiation. ES cells can be induced to differentiate by growth in suspension where they will form aggregates. After a few days, aggregates will form a layer of endoderm cells on their outer surface. The structures thus formed are termed embryoid bodies. Continued culturing of these embryoid bodies in suspension leads to formation of cystic bodies and further differentiation of the cells into ectodermal and mesodermal cell types. When nuclear extracts of undifferentiated ES cells were assayed for the presence of octamer binding factors, three complexes were observed; Oct1,

Figure 3. Protease clipping bandshift assay of the Oct6 protein. Radiolabelled probe was incubated at roomtemperature with nuclear extracts of Oct6 transfected COS-1 cells (left), MES68cl2 (middle) and Epi7 cells (right). After ten minutes, different amounts of the endoprotease ArgC was added (amounts as indicated at the bottom of the lanes). Incubation at roomtemperature was continued for another ten minutes, before loading the samples on the gel. The position of the Oct6 degradation products are indicated with a triangle. Oct1 degradation products are indicated with a dot. The position of undegraded proteins are indicated with an arrow. Free probe is indicated by F.

Oct4 and Oct6 (Fig. 1C). Previously, these complexes have been described to be present in D3 ES cells and F9 EC cells (9). Differentiation of ES cells correlates with a drastic decrease in intensity of the Oct4 and Oct6 complexes (Fig. 1C; EB4, EB8). The residual intensity of the Oct4 and Oct6 complexes at day 4 (EB4) is probably due to remaining undifferentiated ES cells in the core of the aggregates. Again, appearance of the higher mobility Oct4-like complex is observed during ES cell differentiation (Fig. 1C). No additional complexes were observed. Clearly P19 EC cells differ from true ES cells in that they do not express the Oct6 protein(s) in the undifferentiated state. The regulated expression of the Oct6 complex during P19 cell differentiation and expression in undifferentiated ES cells suggests a role of this protein in both systems. As a first step in defining the role of the Oct6 protein in these systems, we set out to clone a cDNA encoding the Oct6 protein.

Cloning of an Oct6 cDNA

The three octamer binding factors that have been identified to date, by cloning of the corresponding cDNAs, are encoded by different members of the POU domain gene family. As the POU domain constitutes the DNA binding domain of these proteins it is tempting to assume that the octamer factors identified here are also encoded by genes belonging to this family. Recently it was demonstrated that a number of additional POU domain genes are expressed in different regions of the rat brain, kidney and testis (19). The Oct6 protein in ES cells and differentiating P19 cells may well be encoded by a member of this gene family. Given the high homology within the POU domain among the different members of this gene family, we tried to clone a cDNA encoding the Oct6 protein based on this homology. Using a mouse Oct2 POU domain probe, derived from a testis specific Oct2 cDNA (Meijer et al. unpublished results), a mouse brain cDNA library was screened. Six clones (called mb1 to mb6) were isolated, that hybridized with varying intensity to the Oct2 probe. DNA fragments derived from these clones were used as probes to screen Northern blots containing RNA from differentiating P19 and ES cells. It was anticipated, that the expression pattern

Figure 4. Expression of Oct6 mRNA in different mouse tissues and cell lines. Twenty μ g of denatured total RNA was separated on a 1% agarose gel in 0.66M formaldehyde and blotted onto Biotrans blotting membrane. Probe and hybridization conditions as in Fig. 2. Lane 1–4; total RNA from whole embryos at day 10 (e10), day 13 (e13), day 16 (e16) and day 19 (e19) of gestation. Lane 5–12; Total RNA from adult mouse tissues. Lane 13 and 14; Total RNA from MES68cl2 and undifferentiated CCE ES cells.

of the Oct6 mRNA would correlate with the expression pattern of the Oct6 protein in our bandshift experiments. A probe derived from clone mb1 fulfilled this criterion and detected a transcript of approximately 3 kb (Fig. 2A,B). Interestingly, at day 8 of RA induction of P19 cells, expression of mb1 mRNA reappears after its high transient expression around day 1. This correlates exactly with the reappearance of the Oct6 bandshift complex at day 8. However at this stage of differentiation, the Oct6 complex is obscured by the presence of a protein complex with a slightly lower mobility, that first appears at day 4.

In vitro transcription/translation of the mb1 cDNA generated a protein that did bind to the octamer probe, but gave rise to a complex with a slightly higher mobility than the Oct6 complex in a bandshift assay (data not shown). Sequence analysis indicated, that this clone represented a 5' end truncated cDNA. Instead of trying to clone a longer cDNA, the mb1 cDNA clone was extended at its 5' side with 320 bp of genomic sequences. These were derived from a mouse β EMBL3 clone, that contains the genomic counterpart of the 5' end of the mb1 cDNA (not

Figure 5. Nucleotide and predicted amino acid sequence of the extended mb1 cDNA encoding Oct6. Total length of the presented sequence is 1517nt. Nucleotide sequence of cDNA clone mb1 starts at position 155. The first 154 nt are derived from a genomic 510 bp *SacI*-*KpnI* fragment. Position 1 of the presented sequence represents the CAP site of the Oct6 mRNA as determined by primer extension (Fig. 4). The POU specific and POU-HOMEO domains are indicated by the boxed amino acid sequences. Homopolymeric, or quasi-homopolymeric amino acid sequences are underlined. The *KpnI* site that was used to link the genome Oct6 sequences onto the cDNA is indicated. The position of the (–) strand oligonucleotide that was used in the primer extension experiment is indicated with an arrow (pos. 38 to 54).

shown). As indicated in Fig. 6, the 5' end of the cDNA maps at position 155. Downstream of this position, the sequence of the genomic DNA and the mb1 cDNA are exactly colinear to the *KpnI* site at position 344. For convenience the 5' cDNA sequences were replaced by genomic sequences from this *KpnI* site on. Upstream of position 155 the reading frame remains open. No consensus splice acceptor or donor sites are present (Fig. 5). However, at positions –25 and –80 with respect to the presented sequence a TATA box sequence and CCAAT box are present respectively (not shown), suggesting that this genomic fragment also contains part of the Oct6 promoter. To prove that the extended cDNA contained the entire Oct6 open reading frame, plus part of the Oct6 promoter the construct was cloned into the SV40 based expression vector pcDX (39) in the antisense orientation with respect to the SV40 early promoter and transfected into COS-1 cells. In this way an Oct6 protein can only be produced when transcription starts from the putative Oct6 promoter.

As can be concluded from the results presented in Fig. 3 (see also Fig. 6) the extended mb1 clone described above produced

a protein that gives rise to an octamer complex in a bandshift assay with the same mobility as the Oct6 complex observed when using MES68cl2 nuclear extracts (an adenovirus E1a transformed MES1 cell line (44); M.P.M. and A.L. unpublished results).

Furthermore primer extension experiments indicated that the CAP site used on the extended cDNA template in transfected COS cells is the same as that used in MES68cl2 cells (data not shown).

To further corroborate the identity of the Oct6 protein encoded by the cloned DNA with the endogenous Oct6 protein present in the MES68cl2 cell line a protease clipping bandshift experiment (PCBA; 28) was performed. The MES68cl2 cell line was used as a reference instead of ES cells or RA induced P19 cells, because it has a high endogenous level of Oct6 (see Fig. 4) and gave a clear degradation pattern in the PCBA. The presence of Oct4 in ES cells and P19 cells would have obscured the clipping pattern of Oct6. Limited proteolytic degradation of the protein encoded by the cloned DNA gives rise to three complexes (Fig. 3, left panel); indicated with triangles. The same degradation products are observed when Oct6 containing MES68cl2 nuclear

extracts were used (Fig. 3, middle panel). The two additional bands observed in the MES68cl2 experiment (indicated with a dot) are derived from degradation of Oct1, as can be concluded from the third panel that shows the degradation pattern of Oct1 in a cell line that only expresses Oct1 (Epi7).

Taken together these results indicate that the construct indeed contains the entire Oct6 open reading frame and that transcription is initiated from the Oct6 promoter using the authentic CAP site.

Since the distance between the CAP site and the 5' end of the mb1 cDNA measures 154 bp, the ORF of the construct is extended by 33 amino acids to the methionine indicated as the start codon in Fig. 6.

Northern blot analysis

The Oct6 complex was observed in differentiating P19 cells, adult brain and ES cells. In order to get a more complete picture of the expression of the Oct6 gene, various tissues of adult animals and whole embryos at different stages of gestation were analyzed using northern blots. The result of this limited survey is shown in Fig. 4. No adult tissue other than brain has an appreciable level of Oct6 expression. Furthermore, no expression of Oct6 was observed in whole embryos at day 10, 13, 16 and 19 pc. RNA samples of MES68cl2 and ES cells were included as extra controls.

Sequence of Oct6

The full length Oct6 clone was sequenced on both strands. The complete nucleotide sequence is presented in Fig. 5. Translation of the sequence reveals a long open reading frame (ORF), encoding a protein of 449 amino acids with a calculated molecular weight of 45 kD. As expected, this reading frame contains a POU domain (residue 253–397). Comparison of the Oct6 POU domain with published POU domain sequences revealed a 100% identity with the rat SCIP/Tst-1 POU domain (18, 19), indicating that Oct6 is encoded by the mouse counterpart of the rat SCIP/Tst-1 gene. Apart from the POU domain, the protein does not show a high degree of homology with other members of the POU domain gene family. According to the terminology of He et al. (19) the Oct6 protein is a Class III POU domain protein. The *Drosophila* Cf1a gene is the only other member of this subclass of POU domain genes, for which a sequence is published (apart from the POU domain itself) (40). Comparison of the Oct6 amino acid sequence with Cf1a did reveal a short stretch of homology in the aminoterminal part of the proteins. This homology is due to the presence of a relatively long (9) stretch of alanine residues. Interestingly the Oct6 protein contains several homopolymeric amino acid stretches. Apart from a stretch of alanine residues there are several stretches rich in glycine, histidine and proline. Glycine rich amino acid domains are also present in the Oct4/Oct3 protein (16, 15). This protein is also very proline rich. The amino acid sequence of the Oct6 protein carboxyterminal of the POU domain is particularly rich in proline residue (25%). The functional importance of the domains outlined above remains to be determined.

The Oct6 protein binds to the octamer motif

To determine whether the Oct6 protein binds to the octamer motif contained within the probe DNA, DNA binding competition experiments were performed. As competitor sequences we used two DNA fragments of which the first consists of a six times repeated synthetic oligonucleotide, derived from the IG enhancer sequence, containing the octamer binding site, while the second consists of the same repeat with a mutated octamer binding site

(FD and Fd respectively (30, 41)). In Fig. 6 it is shown that a thousand fold molar excess of FD competitor inhibited the binding of Oct6 protein to the radiolabelled probe, while the mutated Fd competitor failed to show this effect. Therefore, binding competition depended on the presence of an intact octamer motif which showed that interaction between the Oct6 protein and the probe is mediated by these sequences.

The Oct6 protein binds to other sequence motifs as well

It has been shown that the Oct1 and Oct2 proteins bind to sequences differing considerably from the consensus octa motif (49, 50) albeit with lower affinity. The putative binding domain (POU domain) of Oct6 differs from that of Oct1/Oct2. It is therefore possible that the spectrum of DNA binding sites for Oct6 differs from that of Oct1/Oct2. To test whether the Oct6 protein shows differential affinities for octamer and octamer related binding sites as compared with Oct1 we performed bandshift assays using MES68cl2 nuclear extracts that contain Oct1 and Oct6 protein and different probes known to be binding sites for Oct1. The probes used are listed in figure 7A and are aligned with respect to the consensus octamer motif. The ratio between probe shifted by Oct1 versus Oct6 is taken as an indication of their relative affinities for that site. The three probes that contain a perfect match to the consensus octa motif but differ in their flanking sequences (abl1, U2 and ad4) all bind Oct1 and Oct6 albeit with different affinities (Fig. 7B). This indicates that flanking sequences contribute to binding affinities and that these contributions are different for Oct1 and Oct6. The ad2 octa motif differs at two positions from the consensus octamer sequences. However the ratio between Oct1 and Oct6 is now shifted in favour of the Oct6 complex indicating that this mutation more strongly affects Oct1 binding to the ad2 octa motif than Oct6 binding. The strongest differential affinity between Oct1 and Oct6 is seen with the Herpes simplex virus (HSV) ICP4 gene promoter TAATGARAT motif. This site poorly binds to Oct1 whereas

Figure 6. The Oct6 protein binds to the octamer motif. Radiolabelled probe was incubated with nuclear extracts of Oct6 transfected COS cells (lane 1 to 3) or MES68cl2 nuclear extract (lane 4 to 6) either in the absence (lane 1 and 4) or in the presence of a 1000 fold molar excess of cold competitor DNA (lane 2, 3 and 5, 6). The FD competitor is a 300 bp EcoRI restriction fragment containing six tandem copies of the sequence; ctgagcaaaacaccctgggtaATTGCAATTctaaaa-taagtga (30). The Fd competitor is the same as FD except that it contains an octamer motif mutated at two positions; ATgTtCAT (41).

Figure 7. The octamer binding factors Oct1 and Oct6 bind with different affinities to sequences related to the consensus octamer motif. Panel A shows the sequence of the different OCTA motifs present in the ds oligonucleotides used. Nucleotides fitting the octamer consensus sequence are indicated in black. Only the octamer motif plus flanking nucleotides are shown (see Materials and Methods). abl I: mouse type I c-abl promoter (27). U2: Xenopus U2 small nuclear RNA distal sequences element (46) ad4: adenovirus serotype 4 ITR (48), ad2: adenovirus serotype 2 ITR (48) ICP4: HSV immediate early ICP4 gene promoter (47) abl I (mut): mutated version of abl I hep⁺oct⁺: heptamer (boxed) and octamer elements of the murine 104–2 IgH promoter. hep⁺oct⁻: intact heptamer motif, mutated octa motif (50) (B) Bandshift experiment with MES68cl2 nuclear extracts showing relative affinities of the Oct1 and Oct6 proteins for the different probes (see A). F indicates free fragment. (C) The Oct6 protein binds to the heptamer sequence. Bandshift experiment with ds oligonucleotides hep⁺oct⁺ and hep⁺oct⁻ and nuclear extracts of Oct6 transfected COS cells. DNA protein complexes are indicated as C1 and C2. F indicates free fragment (D) Bandshift competition experiment to assay the relative affinities of Oct1 and Oct6 for the TAATGARAT motif. MES68cl2 nuclear proteins were incubated with radiolabeled abl I probe in the presence of increasing amount of cold ICP4 probe as competitor (right to left). The competitor was added in 2-fold serial dilutions starting with a 1000 fold excess of TAATGARAT binding sites to octa binding sites.

it is a good binding site for Oct6. The abl I (mut) oligo contains three mutations in the octa motif. This mutant was shown to abolish binding by Oct1 and Oct2 (30) and has a similar dramatic effect on binding by Oct6.

A second conserved element within the promoters of IgH genes, the heptamer motif located 2–22 bp upstream of the octa element, was shown to be a binding site for Oct1 and Oct2 (50). Two ds oligonucleotides were used to assay binding of Oct6 to the heptamer motif. The result of this experiment is shown in fig. 7C. Clearly Oct6 binds to the heptamer element in the absence of an intact octamer motif (complex C2). In the presence of an intact octamer element a lower mobility complex (C1) is observed indicating that both binding sites are occupied by Oct6.

To show that the ratio of Oct1 and Oct6 complexes seen with the ICP4 probe indeed reflects the different affinities of the two proteins for this binding site we performed a binding competition assay. Oct1 and Oct6 complex formation on the abl I octa motif was challenged by increasing amounts of cold ICP4 competitor. Clearly the ICP4 oligo more efficiently competes out the formation of the Oct6 complex than the formation of the Oct1 complex, reflecting a higher affinity of the Oct6 protein for the TAATGARAT motif. From this limited survey it is clear that Oct6 binds the same spectrum of binding sites as Oct1 but exhibits different affinities.

DISCUSSION

During early embryogenesis, the totipotent cells of the inner cell mass (ICM) of the blastocyst embryo become committed to specific differentiation pathways. Murine embryonic stem cells and embryonal carcinoma cells provide a culture system in which it is possible to study these early embryonic events. Here we used P19 EC cells and CCE ES cells to study octamer binding factors that are differentially regulated during differentiation of these cells. Several interesting points emerged from this survey. The first change that is observed, when EC cells and ES cells form aggregates, is the induction of an Oct4-like complex, that runs ahead of the Oct4 complex in a bandshift assay. Whether the protein present in this faster migrating complex is encoded by the same Oct4 gene remains to be determined.

Second, a family of octamer binding factors is observed during neuronal differentiation of P19 cells. These additional complexes appear in a temporally ordered fashion during the differentiation process, suggesting that each of them plays a role in successive steps of differentiation. The P19 EC cells provide an excellent system to study the role of these octamer factors in the establishment of glial and neuronal cell lineage and subsequent differentiation of these lineages since the different complexes only appear in the RA-induced differentiation of P19 EC cells. Of particular interest is the appearance of the Oct6 complex as it is one of the first complexes to be induced in P19 cells upon RA addition to the culture. This induction is biphasic; the early expression is transient but reappears later during differentiation (d8), suggesting that the protein may play a role in the establishment of the neuronal differentiation direction (around day 1 of RA induction), as well as in the establishment or induction of a more differentiated phenotype (day 8). Furthermore, the protein is expressed in undifferentiated ES cells and is down regulated upon differentiation of these cells. As a first step in defining the diverse roles of the Oct6 protein in these different differentiation processes we cloned a cDNA encoding the Oct6 protein. Three lines of evidence suggest that we have cloned the gene encoding the Oct6 protein. First, the expression

pattern of the gene at the RNA level correlates with the expression pattern of the Oct6 complex during P19 EC cell and ES cell differentiation. This correlation is also found in the cell line MES68cl2, which has the highest amount of Oct6 protein and mRNA of all cell lines analyzed. Second, in bandshift assays the protein expressed from the extended cDNA construct in COS-1 cells gave the same mobility shift as the endogenous Oct6 protein, present in nuclear extracts of differentiating P19 cells, MES68cl2 or mouse brain. Third, limited proteolytic degradation experiments on the protein expressed from the cloned Oct6 gene in COS-1 cells revealed a pattern of degradation, that is identical to the degradation pattern of the endogenous Oct6 protein in MES68cl2 cells.

As anticipated, sequence analysis of the cloned Oct6 gene revealed the presence of a POU domain within the longest open reading frame. The original cDNA clone mb1 appeared to miss part of the N-terminus of the ORF, since both *in vitro* transcription/translation of this clone, as well as expression in COS-1 cells produced a protein with a higher mobility in a bandshift assay than the endogenous Oct6 protein in ES cells. Since the mb1 sequence started at nucleic acid position 155, the *in frame* methionine at amino acid position 51 probably functioned as the start codon in these experiments. Extension of mb1 with the homologous genomic sequences, supplies two more *in frame* methionine residues (pos. 1 and 23). Because the presented sequence encodes a protein of the same size as the endogenous Oct6 protein, we favour the first methionine to be the initiation codon, since it has a better resemblance to the Kozak sequence (42). The genomic extension of the cDNA clone supplied the authentic CAP site to the construct, which was used in the COS-1 cells transfected with this DNA inserted in an SV-40 based expression vector. Therefore, the genomic fragment must at least contain the minimal promoter of the Oct6 gene. Since the Oct6 mRNA measures 3 kb of which the extended cDNA represents half the size, in which the CAP site is contained, the 3' untranslated region is estimated to be 1.5 kb. From the sequence it is obvious that the cDNA lacks the poly (A) addition sequence; probably cDNA synthesis was primed from an A-rich sequence, present in the 3' UTR.

The DNA binding competition experiments, with an intact and a mutated octamer sequence clearly showed that interaction between the Oct6 protein and the probe DNA was mediated by the octamer sequence.

Bandshift assays using different binding sites revealed that Oct6 does not only bind to the consensus octamer sequence but also to such degenerate sites as the TAATGARAT and the IgH heptamer motif. Clearly Oct6 and Oct1 differ in their affinity for these sites probably reflecting differences in their POU domain. The observation that Oct6 has a higher affinity for the ICP4 TAATGARAT motif than Oct1 (7B and D) might have important implications for the transcriptional regulation of the HSV IE genes. Activation or inactivation of IE gene expression is the important step leading to the lytic cycle or a latent state of the virus (53). Activation of IE genes requires the assembly of a multiprotein complex on the TAATGARAT motif containing Oct1, the viral protein Vmw65 and a third cellular factor X (51). Critical determinants of Vmw65/Oct1 interaction in the Oct1 protein have been mapped to the homeo domain (52). It will be very interesting to see whether the Oct6 protein can interact with Vmw65 or whether it can efficiently compete with binding of the Oct1/Vmw65/X complex to the TAATGARAT motif. Such antagonist action may well be involved in the regulation of HSV IE genes.

Comparison of the Oct6 POU domain sequence with other known POU domains revealed a 100% identity with the rat SCIP/Tst-1 POU domain (18, 19). The Oct6 protein is probably encoded by the mouse counterpart of this rat gene. This is further corroborated by the following observations: First, the length of the Oct6 transcript is similar to that of the reported SCIP/Tst-1 mRNA (18). Second, Southern blot analysis indicated, that the Oct6 gene is a single copy gene (data not shown), excluding the possibility that the Oct6 protein is encoded by a gene closely related to the SCIP gene. And third, the expression patterns of Oct6 and SCIP overlap. However, an interesting difference exists between Oct6 expression in mouse and SCIP expression in rat. SCIP/Tst-1 was reported to be expressed in rat testis (19). No expression of Oct6 was observed in mouse testis. Such interspecies differences in testicular gene expression have been observed for other genes as well. It is of interest to note that we have detected expression of Oct2 in mouse testis, while He et al (19) showed that Oct2 is not expressed in rat testis (Meijer et al. unpublished results). The reason for these differences is enigmatic.

Outside the POU domain the protein contains a large number of (quasi) homopolymeric stretches of a limited number of amino acids; alanine, glycine, histidine and proline. Proline rich protein domains have been identified in a number of transcription factors (6). It was shown that the carboxyterminal proline rich part of CTF/NF-1 functions as a transactivation domain (45). By extension it is possible that the Oct6 protein acts as a transactivator and that this function is mediated by its proline rich domain. Indeed preliminary experiments indicate that the Oct6 protein is capable of activating a minimal promoter linked to a synthetic enhancer containing a multimerized octamer motif. Whether this function is indeed mediated by the proline rich domain of Oct6 remains to be determined.

It was shown, that SCIP/Tst-1 is probably involved in the differentiation of PNS and CNS glial cells into myelinating Schwann cells and oligodendrocytes (18). Furthermore, *in situ* hybridization data indicated, that SCIP/Tst-1 is also expressed by a restricted set of neurons in the adult brain, as well as in different regions of the developing nervous system during embryogenesis (19). Since we used nuclear extracts from whole populations of differentiating P19 cells, we cannot assign any of the observed octamer factors to individual cell types. However in differentiating P19 cells the main type of glial cells that are formed are astrocytes. No myelinating cells were found to be present (38). This suggests that expression of Oct6 in differentiated P19 cells represents expression in neuronal cells. Recently, we have cloned the complete Oct6 gene from a mouse genomic library. This will enable us to study the regulatory sequences of the gene, that direct expression in the different cell lineages. Construction of an Oct6 promoter driven marker gene will further enable us to study the precise temporal expression in individual cell types, in relation to different differentiation markers. Further studies will aim at the identification of target genes of the Oct6 protein and on the regulation of the Oct6 gene itself.

ACKNOWLEDGEMENTS

We thank Prof. Dirk Bootsma for continuous support. We gratefully acknowledge Ernie de Boer, Uli Schibler and Meinrad Busslinger for technical suggestions and Hans Schöler for providing plasmids and communicating sequence data prior to publication and Peter Verrijzer for oligo's and stimulating

discussion. We thank Rita Boucke and Jeannette Lokker for typing the manuscript, Tom de Vries Lentsch and Mirko Kuit for photographic work, and Sjoef van Baal for computer assistance.

REFERENCES

- Owen, D. and Kühn, L.C. (1987). *EMBO J.* **6**, 1287–1293.
- Rothenberger, S., Müllner, E.W. and Kühn, L.C. (1990) *Nucl. Acid Res.* **18**, 1175–1179.
- Baker, B.S. (1989) *Nature* **340**, 521–524.
- Bingham, P.M., Chou, T.B., Mims, I. and Zachar, Z. (1988) *Trends in Gen.* **4**, 134–138.
- Hatzopoulos, A.K., Schlokot, U. and Gruss, P. (1988) in: *Transcription and Splicing* eds. B.D. Hames and D.M. Glover (IRL Press, Oxford).
- Mitchell, P.J. and Tjian, R. (1989) *Science* **245**, 371–378.
- Kemler, I. and Schaffner, W. (1990) *FASEB J.* **4**, 1444–1449.
- Labella, F., Sive, H.L., Roeder, R.G. and Heintz, N. (1988) *Genes & Dev.* **2**, 32–39.
- Schöler, H.R., Hatzopoulos, A.K., Balling, R., Suzuki, N. and Gruss, P. (1989) *EMBO J.* **8**, 2543–2550.
- Ko, H.S., Fast, P., McBride, W. and Staudt, L.M. (1988) *Cell* **55**, 135–144.
- Sturm, R.A., Das, G. and Herr, W. (1988) *Genes & Dev.* **2**, 1582–1599.
- Müller, M.M., Ruppert, S., Schaffner and Matthias, P. (1988) *Nature* **336**, 544–551.
- Scheidereit, C., Cromlish, J.A., Gerster, T., Kauvakamic, K., Balmaceda, C.G., Currie, R.A. and Roeder, R.G. (1988) *Nature* **336**, 551–557.
- Clerc, R.G., Corcoran, L.M., LeBowitz, D. and Sharp P.A. (1988) *Genes & Dev.* **2**, 1570–1581.
- Schöler, H.R., Ruppert, S., Suzuki, N., Chowdhary, K. and Gruss, P. (1990) *Nature* **344**, 435–439.
- Okamoto, K., Okazawa, H., Okuda, A., Sakai, M., Muramatsu, M. and Hamada, H. (1990) *Cell* **60**, 461–472.
- Herr, W., Sturm, R.A., Clerc, R.G., Corcoran, L.M., Baltimore, D., Sharp, P.A., Ingraham, H.A., Rosenfeld, M.G., Finney, M., Ruvkun, G. and Howitz, H.R. (1988) *Genes & Dev.* **2**, 1513–1516.
- Monuki, E.S., Weinmaster, G., Kuhn, R. and Lemke, G. (1989) *Neuron* **3**, 783–793.
- He, X., Treacy, M.N., Simmons, D.M., Ingraham, H.A., Swanson, L.W. and Rosenfeld, M.G. (1989) *Nature* **340**, 35–42.
- Robertson, E., Bradley, A., Kuehn, M. and Evans, M. (1986) *Nature* **33**, 445–448.
- Rudnicki, M.A. and McBurney, M.W. (1987) in: *Teratocarcinomas and Embryonic Stem Cells: A Practical Approach*, E.J. Robertson, ed. (IRL Press, Oxford).
- Robertson, E.J. (1987) in: *Teratocarcinomas and Embryonic Stem Cells: A Practical Approach*. E.J. Robertson, ed. (IRL Press, Oxford).
- Schreiber, E., Matthias, P., Müller, M.M. and Schaffner, W. (1989) *Nucl. Acid Res.* **17**, 6419.
- Cough, N.M. (1988) *Anal. Biochem.* **173**, 93–95.
- Kalb, V.R. and Bernlohr, R.M. (1977) *Anal. Biochem.* **82**, 362–371.
- Barberis, A., Superti-Furga, G. and Busslinger, M. (1987) *Cell* **50**, 347–359.
- Bernards, A., Paskind, M. and Baltimore, D. (1988) *Oncogene* **2**, 297–304.
- Schreiber, E., Matthias, P., Müller, M.M. and Schaffner, W. (1988) *EMBO J.* **7**, 4221–4229.
- Lichtsteiner, S., Wuarin, J. and Schibler, U. (1987) *Cell* **51**, 963–973.
- Gerster, T., Matthias, P., Thali, M., Jiricny, J. and Schaffner, W. (1987) *EMBO J.* **6**, 1323–1330.
- Auffray, A. and Rougeon, F. (1980) *Eur. J. Biochem.* **107**, 303–314.
- Fourney, R.M., Miyakoshi, J., Day III, R.S. and Paterson, M.C. (1988) *Focus* **10**, 5–7.
- Feinberg, A.P. and Vogelstein, B. (1983) *Anal. Biochem.* **132**, 6–13.
- Meijer, D., Hermans, A., von Lindern, M., van Agthoven, T., de Klein, A., Mackenbach, P., Grootegoed, A., Talarico, D., Della Valle, G. and Grosveld, G. (1987) *EMBO J.* **6**, 4041–4048.
- Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular cloning: A laboratory manual* (Cold Spring Harbor, New York; Cold Spring Harbor Laboratory).
- Sanger, F.G., Nicklen, S. and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
- Jones-Villeneuve, E.M.V., Rudnicki, M.A., Harris, J.F. and McBurney, M.W. (1983) *Mol. Cell. Biol.* **3**, 2271–2279.
- McBurney, M.W., Reuhl, K.R., Ally, A.I., Nasipuri, S., Bell, J.C. and Craig, J. (1988) *J. Neurosci.* **8**, 1063–1073.
- Okayama, H. and Berg, P. (1983) *Mol. Cell Biol.* **3**, 280–289.
- Johnson, W.A. and Hirsh, J. (1990) *Nature* **343**, 467–470.
- Schöler, H.R., Balling, R., Hatzopoulos, A.K., Suzuki, N. and Gruss, P. (1989) *EMBO J.* **8**, 2551–2557.
- Kozak, M. (1989) *J. Cell Biol.* **108**, 229–241.
- Elsholtz, H.P., Albert, V.R., Treacy, M.N. and Rosenfeld, M.G. (1990) *Genes & Dev.* **4**, 43–51.
- Mummery, C.L., Feyen, A., Molenaar, W.H., van den Brink, C.E. and de Laat, S.W. (1986) *Exp. Cell Res.* **15**, 229–242.
- Mermod, N., O'Neill, E.A., Kelly, T.J. and Tjian, R. (1989) *Cell* **58**, 741–753.
- Bohmann, D., Keller, W., Dale, T., Schöler, H.R., Tebb, G. and Mattaj, I.W. (1987) *Nature* **325**, 268–272.
- O'Hare, P. and Goding, C.R. (1988) *Cell* **52**, 435–445.
- Verrijzer, C.P., Kal, A.J. and van der Vliet, P.C. (1990) *Genes & Dev.* in the press.
- Baumruker, T., Sturm, R. and Herr, W. (1988) *Genes & Dev.* **2**, 1400–1413.
- Kemler, I., Schreiber, E., Müller, M.M., Matthias, P. and Schaffner, W. (1989) *EMBO J.* **8**, 2001–2008.
- Gerster, T. and Roeder, R.G. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 6347–6351.
- Stern, S., Tanaka, M. and Herr, W. (1989) *Nature* **341**, 624–630.
- Roizman, B. and Sears, A.E. (1987) *Ann. Rev. Microbiol.* **41**, 543–571.