c-ABL GENE EXPRESSION AND SPERMATOGENESIS: INVESTIGATIONS INTO THE POSSIBLE ROLE OF OCTAMER TRANSCRIPTION FACTORS

c-abl gen expressie en spermatogenese: onderzoek naar de mogelijke rol van octameer transcriptie factoren.

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Erasmus Universiteit Rotterdam
op gezag van de Rector Magnificus
Prof. Dr. C.J. Rijnvos
en volgens besluit van het College van Dekanen.

De openbare verdediging zal plaats vinden op
woensdag 10 juni 1992 om 13.30 uur

door

Dingenus Nicolaas Meijer
geboren te Almkerk

PROMOTIECOMMISSIE

Promotor:	Prof. Dr. D. Bootsma
Overige leden:	Prof. Dr. P.C. van der Vliet Prof. Dr. J.A. Grootegoed Dr. F.G. Grosveld
Co-promotor:	Dr. G.C. Grosveld


Gedrukt door: Drukkerij Haveka B.V., Alblasserdam

Dit proefschrift is bewerkt binnen de vakgroep Celbiologie en Genetica van de Faculteit der Geneeskunde en Gezondheidswetenschappen van de Erasmus Universiteit Rotterdam.

Voor mijn Vader, Zwanie, Wilma
In liefdevollle herinnering aan
mijn Moeder, Hayo, André

Contents

Chapter I

GENERAL INTRODUCTION

The most obvious hallmark of all metazoans is the specialization of particular cell types to carry out different function. This specialization is reflected by the expression of specific genes who's products are required to carry out the specialized tasks of the cell. Thus, red blood cells produce massive amounts of hemoglobin to transport oxygen from the lungs to all parts of the body. Liver cells produce albumin, B-lymphocytes antibodies, eye-lens cells crystallines and so on. Yet the cells that make up the adult organism are all derived from a single cell, the fertilized egg. At the onset of embryogenesis the fertilized egg starts to divide and multiply. The newly generated cells organize themselves into primordial germ layers, from which all the embryonic and extraembryonic tissues are derived, and create body form and structure through a regulated process of further cell division, migration and differentiation. Since each cell within the developing organism contains the full complement of genetic information (except for B and T cells) the central problem in developmental genetics can be phrased as; How is the genetic information regulated during development so that cells acquire a stable differentiated phenotype?

Studies on *Drosophila* and *Caenorhabditis elegans* embryogenesis have suggested the existence of a hierarchy of regulatory genes. The sequential activation of these genes accomplish the transformation of genetic information into body form and structure through a spatial and temporal regulation of effector genes, which in turn determine cell identity. Because of the lack of a large number of developmental mutants, a similar genetic approach to identify developmental control genes in vertebrates is not feasible. Instead a 'reverse' approach has been taken by identifying genes in the genomes of higher organisms that are structuraly related to developmental control genes in *Drosophila* (see [1] for review). Thus, the vertebrate Hox genes were identified on the basis of structural homology with the *antennapaedia* homeo box, while the Pax genes contain a sequence that is homologous to the paired box as present in the *paired* gene in *Drosphila*. Three members (Pax3, 6 and 7) of the Pax gene family contain in addition to the paired box a class specific homeo box; the paired type homeo box. It is assumed that many of these genes, like the Pax and Hox genes play a similar role in higher organisms like *Xenopus*, mouse and human. Consistent with this idea, Hox and Pax genes are expressed during mouse development in a spatially and temporally restricted way. Moreover, the direct involvement in development of at least the Pax genes is becoming evident now that mutations in a number of these genes are associated with genetic developmentally aberrant traits in mouse and human: 1. the Splotch mutation in mouse and van Waardenburg syndrome in human are associated

with a mutation in the mouse and human Pax-3 genes respectively [2, 3, 4]. Because of similarities in phenotype it was suggested previously that the Splotch mutation would provide an adequate animal model for this human trait. 2. Mutations in the Pax-6 gene are associated with the small eye mutation in mice and with aniridia in man, while 3. the Pax-1 gene was found to be mutated in the genome of the mouse mutant *undulated* [5, 6, 7].

An alternative approach to isolate developmentally important regulatory genes would be to start at the bottom of the proposed hierarchy of regulatory events, i.e. with the proteins that interact with cis-acting elements that govern cell type specific expression of the linked genes. Of special interest would be those proteins that are expressed during development and show a cell type restricted expression pattern. Using the well characterized 'octamer' DNA element, this approach has led to the identification of a family of sequence specific DNA-binding factors (Oct factors) that show a cell type specific expression pattern and are expressed at different stages of mouse development ([8-13], for review see [14]). The cloning and characterization of cDNA molecules encoding the Oct1 and Oct2 proteins revealed that they are structurally related to each other and to another mammalian pituitary specific transcription factor, Pit1 (also called GHF-1), and the *C. elegans* regulatory protein *Unc86* [15-21]. The structural homology between the DNA binding domain of these proteins defined a new family of cell type specific transcription factors: The POU domain (gene) family (POU being the acronym for **P**it-1, **O**ct1/2, *Unc86*) which is illustrated in Figure 1 [22, 23, 24]. Using the structural similarity between these genes as a tool, several other members of this family were identified and characterized [10, 25-40].

In the work described in this thesis, we have used the approach described above to identify octamer binding factors that could play a role during the terminal cytodifferentiation of sperm cells (chapter IV and V) in the mouse, and in early embryogenesis, using mouse embryonal carcinoma cells as an *in vitro* model (chapter VI). This work led to the cloning of cDNA molecules encoding octamer binding factors expressed during specific stages of differentiation.

In the following paragraphs, I will briefly review some aspects of the POU domain protein family. Then I will discuss two members of the POU domain family for which genetic data are available, to illustrate the role these proteins play in determining cell identity. In the last paragraph of this general introduction, I will give an outline of the experimental work described in this thesis.

The Octamer sequence and POU proteins.

The octamer sequence ATGCAAAT was first identified as a conserved element in the promoters (and enhancer) of both the light and heavy chain immunoglobulin genes [41, 42, 43]. The octamer element plays an important

role in the B-cell specific activity of these promoters [42, 43, 44]. This correlates with the observation that two B-cell specific nuclear factors (Oct2A and Oct2B) and a ubiquitous factor (Oct1) interact with this element in *in vitro* binding assays [8, 12, 45]. Unexpectedly, this same conserved element was also found in a number of ubiquitously expressed genes such as the histone H2B and the small nuclear RNA genes [46, 47]. Moreover it is also present in the origin of replication of adenoviruses (see [48]). The octamer element in the H2B promoter confers cell-cycle dependent regulation on this gene through interaction with Oct1, while the same combination is also involved in replication of the adenovirus genome [48-51]. Apart from being a strong B-cell specific enhancer and promoter element, the octamer sequence also stimulates transcription from a linked gene promoter in embryonic stem cells and embryonal carcinoma cells [52]. These cells express Oct1 and two additional factors called Oct3/4 and Oct6 [11, 53]. In fact, using the octamer element in gel retardation assays a number of additional octamer binding factors were identified in a variety of organs and tissues and at different developmental stages of the mouse [11] (reviewed in [14]). Thus, through the interaction with a whole family of cell type specific nuclear factors, the octamer element is involved in a number of biological regulatory mechanisms. All the octamer binding proteins characterized so far belong to the POU domain family of proteins. It is

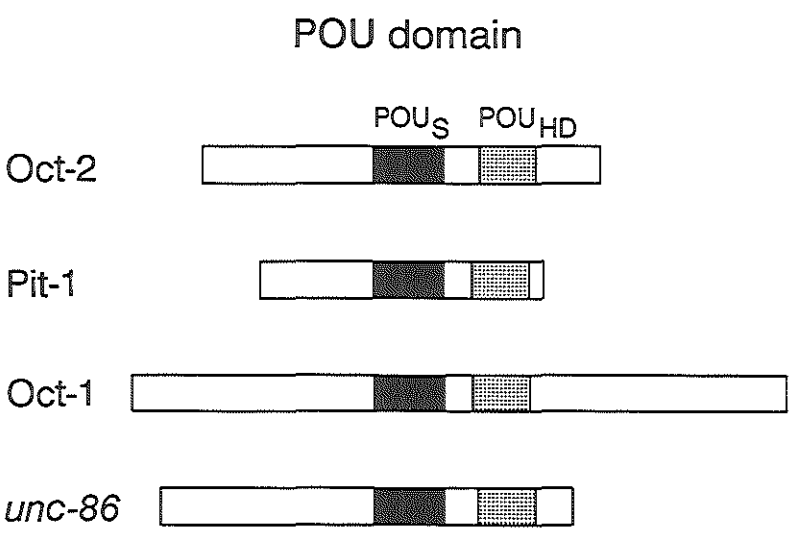


Figure 1. Structural organization and alignment of the four founding fathers of the POU domain protein family. The proteins are aligned with respect to the POU domain as indicated. There is no further homology among these proteins outside the POU domain. The Oct2 protein is also called OTF2 or NF-A2; Pit-1 is also called GHF-1, and Oct1 is also called OBP100, OTF-1, NF-A1 or NFAIII. POU_S is the POU-specific domain and POU_{HD} is the POU-homeo domain. This figure was adapted from [23].

expected that all other octamer binding proteins that have not been characterized yet will belong to this family also.

Structure of the POU domain.

As mentioned before, the POU domain was first recognized as a region that is highly conserved among the three mammalian transcription factors Pit-1 and Oct1/Oct2 and the *C. elegans* gene *Unc86* (see Figure 1). The characterization of additional members of the POU domain gene family and pair wise comparison of the primary structure of the POU domain enabled a further classification of these proteins into five groups [29, 36] (Figure 2). The POU domain is 150-160 amino acid residues in length and can be subdivided in two major regions of very high homology; the POU-specific (POU_S) and the POU-homeo (POU_{HD}) domain. The POU specific region can be further subdivided into two regions referred to as the POU_S-A and POU_S-B region (see Figure 2). A short 14-25 amino acid long linker sequence separates the POU_S from the POU_{HD} domain. This linker is poorly conserved between members of different classes but highly conserved among class III members. The 60 amino acid POU_{HD} domain is distantly related to the classical *Drosophila* homeo box [54]. The homeo domain of the POU proteins shares 22 invariant amino acid residues, 13 of which are also found in the *Drosophila antennapedia* homeo box at similar positions (Figure 2). Thus the homeo domain of the POU proteins are more related to family members than to the homeo domain of *Drosophila* homeobox proteins and the vertebrate hox genes. Therefore they define a new class of homeo proteins. Using the Chou-Fasman algorithm it is predicted that this region will contain three α-helical structures similar to those in the *Drosophila* homeo domain proteins (see below). The POU_{HD} domain contains clusters of basic amino acids at both ends while the POU_S domain has a similar cluster at its amino terminus. The highest homology within the POU_{HD} domain is found in the carboxyl terminal part, as all members of the POU family contain the sequence RVWFCN. This region shows also the highest homology with the classical homeobox proteins. The cystein residue being characteristic for the POU homeo proteins, as most *Drosophila* homeo box proteins have a glutamine (Q) residue, or a serine (S) residue in case of the Pax type homeo box, at this position. It has been suggested that different functional domains of proteins will be encoded by separate exons. This does not seem to be true for the POU gene family, as the different homologous domains of the POU domain are not encoded by separate exons ([55-58], and chapter VIII of this thesis). In fact, the number of exons encoding the different POU proteins various from 1 (Oct6) to over 14 (Oct2) and intron/exon borders in multi exon POU genes are not found in homologous positions. This, together with the fact that POU genes have been cloned from species as diverse as worms, insects, amphibians, birds and mammals, is indicative of their long evolutionary history.

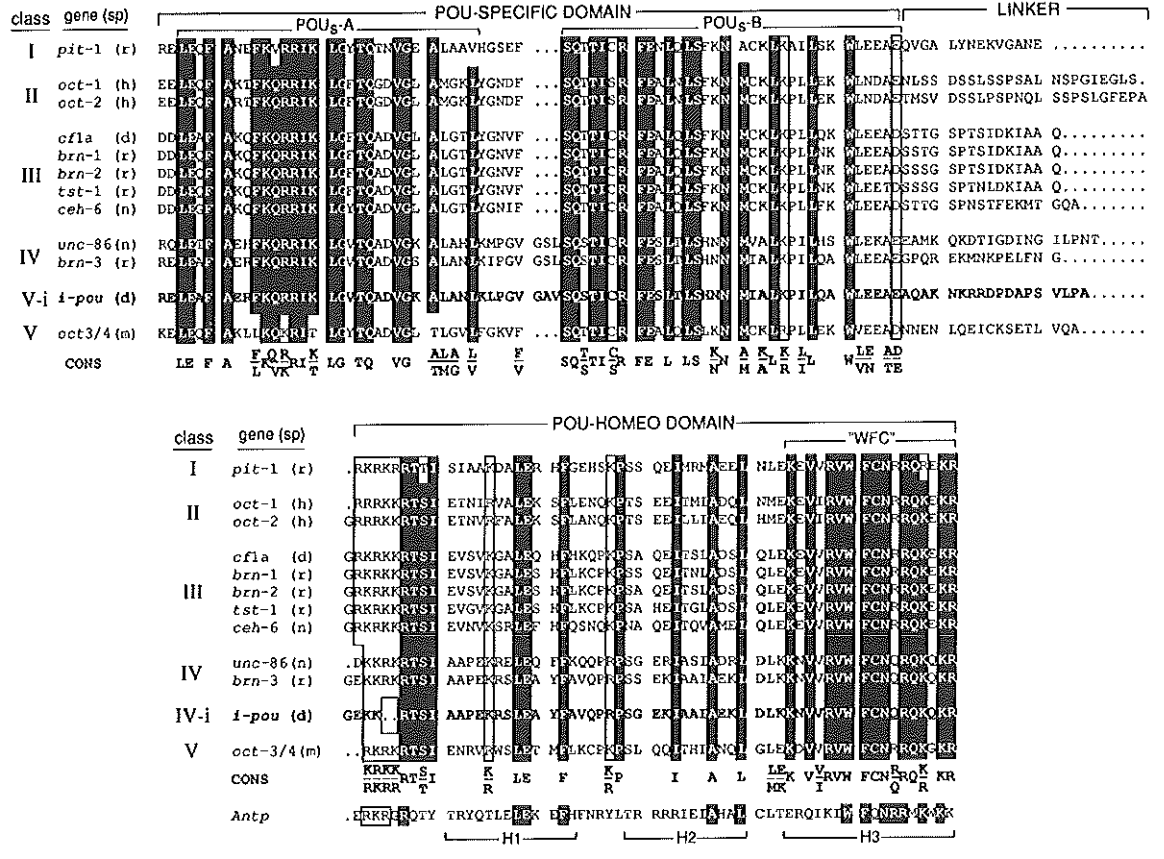


Figure 2. Primary sequence comparison of a family of POU proteins defines five classes of related proteins. Homologies are boxed and identities are highlighted on a black background. The consensus (cons) amino acid sequence is shown under the POU_S and the POU_{HD} domains. The antennapaedia homeo domain is shown with the positions of three helical domains indicated as H1, H2 and H3. The species from which factors were cloned are rat (r), mouse (m), *C. elegans* (n), *Drosophila* (d) and human (h). The factor Tst-1 is also called SCIP and Oct6. Alternative names for the factors Oct1, Oct2 and Pit-1 are mentioned in the legend to figure 1. This figure is taken from [96].

The POU domain and DNA binding.

The demonstration that the POU domain constitutes the DNA binding domain of the POU proteins provided the first conclusive evidence that the homeo motif is a DNA binding domain [24]. Recently the detailed structure of the engrailed homeodomain bound to DNA has been described using X-ray crystallography [59]. The homeo domain is folded in three α-helices (see Figure 3). Helices 2 and 3 form a helix-turn-helix motif related to the structure of prokaryotic regulatory proteins, such as the lambda cro protein and the phage 434 repressor (see [60]). α-Helix 3, the "recognition" helix, binds to the major groove of the DNA helix via specific hydrogen bonding between amino acid side chains and bases. Helices 2 and 1 lay in an anti-parallel orientation perpendicular to the major groove and make hydrophobic contacts with helix 3. Furthermore, two amino acid residues N-terminal of helix 1 contact bases via the minor groove (Figure 3B). Although no X-Ray or NMR data are available yet, it is anticipated that the overall structure of the POU homeodomain bound to DNA will be very similar to that of the engrailed homeodomain. Supportive evidence for this assumption comes from the characterization of the I-POU protein in *Drosophila* [38]. This protein fails to bind DNA but instead strongly forms heterodimers with another POU protein, Cf1a [31]. It was shown that the inability of the I-POU

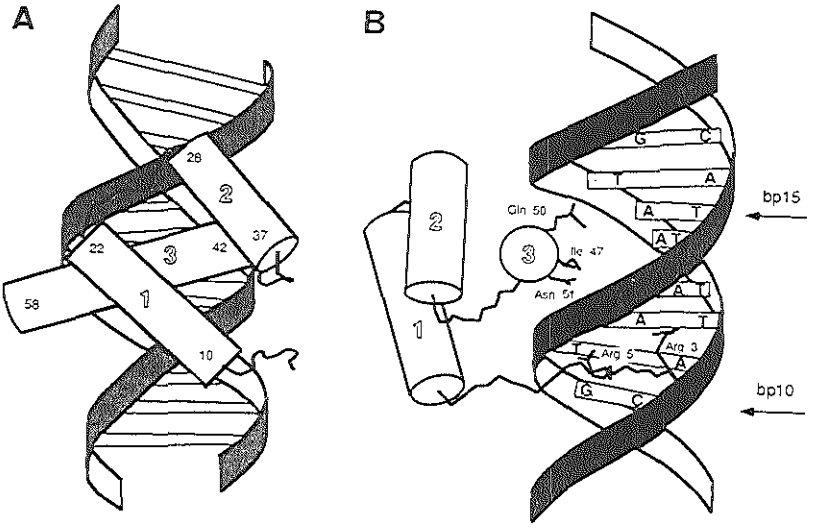


Figure 3. A. Schematic drawing showing the relationship of the engrailed homeo domain α helices with respect to the double helical DNA. Cylinders represent the α helices (numbered 1 to 3), while ribbons are used to show the sugar-phosphate backbone of the DNA. The B figure is as A, but rotated 90,° and gives a view along the axis of helix 3 which lies in the major groove. Critical contacts between amino acid side chains and bases are indicated. These figures are taken from reference [59].

protein to bind DNA is due to the absence of two basic amino acid residues (RK), N-terminal of helix 1. These two amino acid residues are in a similar position as the two basic residues that make minor groove contacts in the engrailed protein. Reintroduction of these residues into I-POU turns the polypeptide into a DNA binding protein that is no longer capable of binding to Cf1a. Furthermore a mutation in the carboxyl terminal basic cluster within the Pit-1 homeodomain (substituting the arginine for a glycine residue directly flanking the WFCN motif) completely abolished DNA binding [61]. Introduction of a proline residue in the third helix had the same detrimental effect on DNA binding, underscoring the importance of the recognition helix in DNA binding. However the mutation of the cysteine residue within the WFCN motif into a glutamine (which is representative of the *antennapaedia* class of homeo proteins) did not effect DNA binding, while in the classic homeo proteins this residue is important for specificity [61, 62, 63].

Deletion analysis of various other POU domain proteins shows that, whereas an intact homeo domain is required for DNA binding, the contribution of the POU specific domain is variable and depends on the exact sequence of the DNA binding site [61, 64]. For the Pit-1 protein, the homeodomain suffices for low affinity binding with relaxed sequence specificity, while the presence of the POU_S domain increases site specificity and affinity (see further Ingraham *et al.* [61] for a detailed mutational analysis of the POU_S domain). A detailed analysis of the contributions of the Oct1 POU_S domain in binding of the POU domain to the ad2 octamer sequence was undertaken by Verrijzer *et al.* [64]. Their results showed that the POU_S domain contacts the 5' half of the ATGATAAT sequence, while the homeodomain contacts the 3' half of this sequence. At high protein concentrations, the isolated POU_S domain was shown to bind to the the 5'part of several octamer and related binding sites. Hydroxy radical footprinting showed that the majority of the backbone contacts are made by the POU_{HD}. The specificity of binding is determined by base contacts, while the binding energy stems almost completely from electrostatic interactions between the protein and the phosphate backbone [60]. Curiously enough, the presence of the POU_S domain greatly increases the affinity of the POU domain for the ad2 octamer element. This could indicate that part of the binding energy of the POU domain stems from protein-protein interactions, possibly between the POU_S and the POU_{HD} domains. The contributions of the POU_S domain to DNA binding affinity depends on the specific recognition site. The Oct1 POU domain was also shown to induce conformational changes (bending) in the bound DNA and is sufficient to stimulate adenovirus DNA replication in an *in vitro* replication assay [51, 65]. Both functions are lost upon deletion of the POU_S domain.

Recently it was shown that the DNA binding affinity of the Pit-1 protein can be modulated by phosphorylation. Phosphorylation of a serine residue, next to the amino terminal basic cluster in the POU_{HD}, modifies the

conformation of Pit-1 on DNA recognition sites and results in either decreased or increased affinity depending on the flanking nucleotide sequence [66]. Furthermore, phosphorylation of the Oct1 protein was shown to be modulated as a function of the cell cycle correlating with the proposed involvement of this protein in the cell cycle regulated expression of the H2B gene [49, 50]. Thus phosphorylation is an important mechanism through which the (DNA binding) activity of these POU proteins can be regulated.

The POU domain and protein-protein interactions.

Several studies have indicated that the POU domain is not only a DNA binding motif but is also involved in protein-protein interactions. For example, together with at least one other cellular factor, the Oct1 protein is recruted by the Herpes Simplex Virus (HSV) protein VP16 (also called α-TIF and Vmw65) into a multiprotein complex on the TAATGARAT response elements, to stimulate transcription of the immediate early gene promoters of HSV [67-71]. The Oct1 protein alone binds with low affinity to this response site, but DNA binding is greatly increased upon interaction with VP16 and other proteins. VP16 itself does not bind to DNA with high affinity, but instead contributes a strong transactivation domain to the complex. The interaction between VP16 and Oct1 is specific for Oct1, as the closely related Oct2 protein fails to form a complex with VP16. A detailed mutational analysis by Stern *et al.* revealed that seven amino acid differences between helix 1 and helix 2 of the Oct1 and Oct2 POU_{HD} domains dictates the specificity of interaction with VP16 [72]. As discussed above, the helices 1 and 2 are not directly involved in DNA binding but instead function in highly specific protein-protein interactions (see below). The VP16/Oct1 interaction clearly illustrates how non DNA binding proteins can specifically influence differential transcriptional regulation of two proteins that have identical DNA binding specificities. It also presents an extreme example of modularity in the sense that the DNA binding interface (Oct1) and the transcriptional activation domain (VP16) are located on separate proteins. The case of VP16 has stimulated a lot of speculation about the possible existence of similar cellular factors. Such factors have not been described to date.

As already mentioned in the previous paragraph, the *Drosophila* I-POU protein does not bind DNA, but forms heterodimers with the Cf1a and not with any of the other known POU factors [38, 73]. The region required for heterodimer formation is entirely limited to the POU_{HD} domain and encompasses only the amino terminal basic region and helices 1 and 2. The I-POU protein does not dimerize with the Brn-2 protein, another octamer binding factor, although this protein has a POU_{HD} domain that differs only in three positions in helix 1 with Cf1-a. Only when all three amino acid residues in the Brn-2 protein are mutated to the corresponding residues in Cf1-a the protein is able to dimerize with I-POU. It appears that protein-protein

interaction in both cases (Oct1/VP16 and I-POU/Cf1-a) relies on critical determinants in the POU_{HD} helices 1 and 2. These regions in the POU_{HD} show the highest variability in primary sequence between the different classes (see Figure 2). This variability would provide a structural basis for selective protein-protein interaction between POU proteins and other factors. Interestingly, the amino acid sequence separating helix 1 and 2 is identical for the Oct6 (called tst-1 in Figure 2) and Oct3/4 proteins [36]. These proteins belong to different classes and are both expressed in undifferentiated ES and EC cells. It has been suggested by Schöler *et al.* that the conserved linker sequence between helices 1 and 2 in these proteins could provide an interaction interface for an ES/EC cell specific factor [36].

Most of the POU proteins bind as monomers to their recognition sequence, except the Pit-1 protein which binds as a dimer as a consequence of synergistic, DNA binding dependent, protein-protein interactions. Dimer formation of Pit-1 on DNA requires the POU_S domain [61]. Furthermore, cooperative binding of the Oct1 protein to flanking octamer and heptamer sequences in the Ig promoters was shown to depend on the POU_S domain as well [74]. Whether the POU_S domain of other POU proteins plays a role in protein-protein interactions remains to be established.

Transactivation domains of POU proteins.

Most transcription factors contain a transactivation domain that can be physically separated from the DNA binding domain (see for example [75]). The POU proteins do not seem to form an exception to this general rule, although it has been suggested that the amino terminal part of the Oct2 POU_S domain could function in transactivation [76]. The transactivation domains of the Oct1 [77], Oct2 [76, 77, 78], Oct3/4 [34, 79], Oct6 ([80] this thesis) and Pit-1/GHF-1 [61, 81] proteins have been determined. The different transactivation domains do not share an overall similar structure, but instead are characterized by an abundance of one or a few particular amino acid residues [82]. Thus the Oct1 protein has a carboxyl terminal (relative to the POU domain) serine/threonine- and an amino terminal glutamine-rich activation domain, while the Oct2 protein has an amino terminal glutamine/leucine/proline- and a carboxyl terminal serine/threonine-rich domain. The transactivation domain of Oct3/4 is rather proline-rich and is located at the extreme amino terminus of the protein. Oct6 has a glycine/alanine-rich transactivation domain located in the amino terminal third of the protein. The Pit-1/GHF-1 transactivation domain is also located in the amino terminus of the protein and is rich in serine/threonine residues. This great diversity in transactivation domains might indicate that the different POU proteins interact with different components of the basic transcription machinery or with different auxiliary factors [83]. All these proteins have been analyzed in the context of transcriptional activation, using either artificial or natural promoters. However, the Oct6 (tst-1/SCIP) protein has

also been shown to function as a transcritional repressor of the P_0 gene promoter [84, 85]. The P_0 gene is a member of the immunoglobulin superfamily and is specifically expressed in glia cells of the peripheral nervous system [86]. Furthermore, the relative importance of the different activation domains in Oct2 seem to depend, to a certain extent, on the reporter construct used [76, 77]. Thus the validity of the results of this type of experiments should be treated cautiously.

The *Unc86* and Pit-1 POU proteins and development.
In contrast to the classical *Drosophila* and mammalian homeobox genes, most POU homeo genes show a highly cell type restricted expression pattern. From these expression data alone, it is not clear whether the POU proteins play an active role in cellular differentiation or that they are only involved in maintaining a fully differentiated phenotype. The genetic data available for the *C.elegans Unc86* and murine Pit-1 (GHF-1) gene suggest that these POU proteins are indeed actively involved in differentiation, in that they link a particular cell identity to cell lineage.

***Unc86* and cell lineage.**
The effects of *Unc86* mutations are seen in three post-embryonic neuroblast lineages. Figure 4A shows the lineage of the V5.paa neuroblast. In wild type animals, Unc86 protein appears within a few minutes after neuroblast division in one daughter cell but not in the other [87]. The daughter cell that does not express Unc86 will differentiate into a dopaminergic neuron, while the Unc86 expressing daughter will undergo another cell division, giving rise to a different type of neuron and a programmed cell death. In the absence of functional Unc86 protein the daughter repeats the division pattern of the parental cell generating a chain of dopaminergic neurons (Figure 4A). Thus the Unc86 protein is required to distinguish one daugther from its mother cell by reprogramming the expression of an unknown number of genes. The appearance of the Unc86 protein is sensitive to an asymmetric feature of the cell division, probably in a way that does not involve cell-cell interactions.

Besides its role in cell lineage, the Unc86 protein also participates in the determination of certain neural identities. The 57 neurons (approximately one-fifth of the total 302 neurons in the adult organism) that express the Unc86 protein comprise 27 different types of neurons, which share no common feature. This is taken to suggest that the Unc86 protein controls different sets of genes in different cell types. One candidate for control by *Unc86* is the homeo domain gene mec-3 [88]. Although expression of this gene depends on *Unc86*, this is clearly not sufficient as only 10 cells of the 57 *Unc86* expressing cells do actually express the mec-3 gene. This is suggestive for the Unc86 protein being part of a combinatorial code that establishes cellular identity.

Pit-1 and pituitary gland development.

In mammals, the anterior pituitary gland originates from an invagination of the oral ectoderm at the roof of the primitive mouth. This structure, known as Rathke's pouch, appears at around day 8.5 of embryonic development of the mouse. By embryonic day 12, Rathke's pouch detaches from the oral ectoderm and becomes an independent structure. The anterior pituitary originates from the ventral part of rathke's pouch while the dorsal part gives rise to the intermediate lobe of the pituitary. Five phenotypically distinct cell

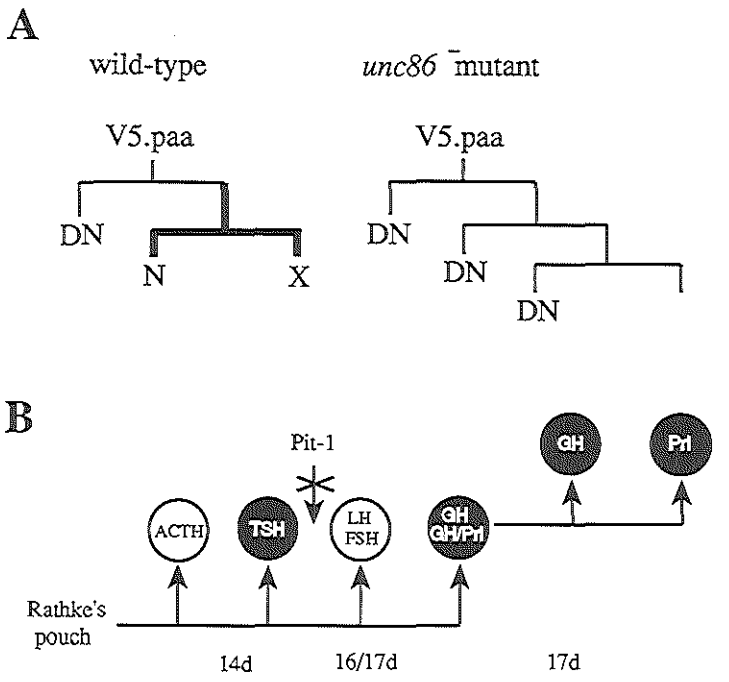


Figure 4. A. Schematic representation of the *C. elegans* V5.paa neuroblast lineage. The other two neuroblast lineages that are affected by *unc86* mutations are the Q and the T.pp neuroblasts. The thick line represent cells that express the unc86 protein in wild type animals. DN is a dopaminergic neuron and N is a sensory neuron that expresses the homeo box gene mec3. X is a programmed cell death. In *unc86*⁻ mutant animals, the division pattern of the parental cell is repeated resulting in the generation of a chain of dopaminergic neurons. The figure is taken from reference [87].

B. Schematic representation of the five different cell types that develop in the anterior pituitary gland. The time points during embryonic development of the rat at which these cell types become identifiable, on the basis of the trophic factor they express, is indicated. The Pit-1 protein is expressed in thyrotrophs, somatotrophs and lactotrophs (indicated in black). These cell types are depleted in dwarf mutant mice. Thyrotrophs appear before the Pit-1 protein is detectable. ACTH is adrenocorticotropin hormone, TSH is thyroid stimulating hormone, LH is luteinizing hormone, FSH is follicle stimulating hormone, GH is growth hormone (somatotropin), Prl is prolactin. This figure is taken from [96].

types will differentiate from the anterior pituitary anlagen in a stereotypical order (see Figure 4B). These cell types are defined by the factor they produce (corticotrophs, thyrotrophs, gonadotrophs, somatotrophs and lactotrophs). Pit-1 gene expression is first detected in all five distinct cell types in the anterior pituitary around day 15 of embryonic development of the rat [89](see also [90]). The Pit-1 protein is detected at low levels around this time, but is restricted to thyrotrophs, lactotrophs and somatotrophs, and precedes the expression of prolactin and growth hormone genes on embryonic day 16-17. This is consistent with the proposed role for Pit-1 in activation of these two transcription units [91].(It should be noted here that there is some disagreement in the literature over the role of Pit-1 in prolactin gene activation and in the cell types in which the Pit-1 gene is expressed; see [92])

A role for Pit-1 in pituitary gland development was firmly established with the molecular characterization of a form of dwarfism (dw locus) in mouse [55]. These genetic dwarf mice produce no detectable levels of growth hormone (GH), prolactin (Prl) and thyroid-stimulating hormone (TSH), and mature somatotrophic, lactotrophic and thyrotrophic cell types are depleted. Two alleles of the dw locus were shown to result from mutations in the Pit-1 transcription unit. In the dwarf Jackson mouse there is a large (>4 kb) disruption (an inversion or insertion) in the Pit-1 gene, while the Snell dwarf mouse is characterized by a point mutation in the homeobox of the gene. This point mutation converts the tryptophan residue in the RVWFCN motif in the recognition helix to a cysteine residue (RVCFCN), resulting in a protein that is no longer able to bind to its recognition elements. The low levels of Pit-1 mRNA in Snell dwarf pituitaries, and the presence of Pit-1 binding sites in the Pit-1 gene promoter is consistent with a mechanism in which Pit-1 boosts its own transcription [93, 94]. Thus the loss of three pituitary cell types in the dw mutants indicates that Pit-1 is involved in at least a part of the pituitary developmental program. It also suggests that the Pit-1 protein is directly or indirectly involved in the proliferation and survival of these cell types. Indeed, inhibition of Pit-1 synthesis by complementary oligonucleotides led to a decrease in proliferation of somatotrophic cell lines and a decrease in GH and Prl expression. The fact that the GH gene and Prl gene are expressed only in a subset of the pituitary cells that express the Pit-1 protein further suggests the existence of additional mechanisms that restrict the actions of this protein [89, 95]. Mutations at a second dwarf locus (df) result in a phenotype that is similar to the two allelic dw mutants including the depletion of somatotrophic, lactotrophic and thyrotrophic cell types in homozygous Ames dwarf animals. No Pit-1 expression could be detected in the hypoplastic pituitaries of Ames dwarf mice [55]. This could suggest that the df gene is involved in the regulation of the Pit-1 gene, or, together with Pit-1, is involved in the specification and/or maintenance of the three pituitary cell types affected by the mutation.

Outline of this thesis.

The work described in this thesis aims at the elucidation of mechanisms that govern cellular differentiation events in male germ cell development (spermatogenesis), especially during the post-meiotic phase (spermiogenesis), and in embryonal carcinoma cells. Chapter II describes the molecular characterization of a variant c-abl mRNA (TSabl) that is specifically expressed at high levels during spermiogenesis and was suggested to play an important role in this proces. The TSabl mRNA is transcribed from the proximal of the two c-abl promoters and is alternatively processed resulting in a removal of most of the 3'UTR, without an effect on the coding capicity of the mRNAs. The high levels of this shortened mRNA in post-meiotic male germ cells could be due to two not mutually exclusive mechanisms, i.e. a higher mRNA stability or/and continued transcription of the gene during the later phases of spermiogenesis. Chapter III describes experiments that tried to address the question whether the TSabl mRNA has a longer half life as a consequence of the removal of most of the 3'UTR. In chapter IV a preliminary analysis of the c-abl promoter is presented, using DNAseI footprinting and gel retardation assays. The results of these experiments hinted at the possibility that there exists a testis specific octamer binding factor that could be involved in the haploid specific regulation of gene expression. This stimulated us to undertake the experiments described in chapter V that aimed at the cloning of testis specific cDNAs encoding octamer binding factors. We show that the POU domain gene Oct2 is highly expressed in spermatogenic cells, generating two transcripts through a mechanism of alternative processing and/or promoter usage. This chapter closes with a discussion of testis specific gene expression.

The temporally regulated expression of a family of octamer binding factors during 'neuronal 'differentiation of P19 EC cells is described in chapter VI. One factor, Oct6, is expressed in a bi-phasic pattern, suggesting that it might play a role at different stages of development. This factor is further characterized by cloning of the cognate cDNA and was found to be the mouse homologue of the rat Tst-1 POU gene [29]. This gene is highly expressed in rat testis but not in mouse testis (this thesis). Chapter VII describes the functional mapping of the protein domains involved in transcriptional activation and DNA binding. In chapter VIII the genomic organization of the Oct6 gene is described. Furthermore, we present an initial characterization of the Oct6 promoter, to begin to address the important question of how this transcriptional regulator is regulated itself. In the last chapter some aspects of the Oct6 gene are discussed in relation to its possible function in differentiation, drawing on examples from other members of the POU domain gene family.

REFERENCES

1. **Kessel, M. and P. Gruss.** 1990. Murine developmental control genes. *Science* **249:** 374-9
2. **Baldwin, C.T., C.F. Hoth, J.A. Amos, E.O. Dasilva and A. Milunsky.** 1992. An Exonic Mutation in the HuP2 Paired Domain Gene Causes Waardenburg's Syndrome. *Nature* **355:** 637-638
3. **Epstein, D.J., M. Vekemans and P. Gros.** 1991. splotch (Sp2H), a mutation affecting development of the mouse neural tube, shows a deletion within the paired homeodomain of Pax-3. *Cell* **67:** 767-74
4. **Tassabehji, M., A.P. Read, V.E. Newton, R. Harris, R. Balling, P. Gruss and T. Strachan.** 1992. Waardenburg's Syndrome Patients Have Mutations in the Human Homologue of the Pax-3 Paired Box Gene. *Nature* **355:** 635-636
5. **Chalepakis, G., R. Fritsch, H. Fickenscher, U. Deutsch, M. Goulding and P. Gruss.** 1991. The molecular basis of the undulated/Pax-1 mutation. *Cell* **66:** 873-84
6. **Hill, R.E., J. Favor, B. Hogan, C. Ton, G.F. Saunders, I.M. Hanson, J. Prosser, T. Jordan, N.D. Hastie and V. Vanheyningen.** 1991. Mouse Small Eye Results from Mutations in a Paired-Like Homeobox-Containing Gene. *Nature* **354:** 522-525
7. **Ton, C., H. Hirvonen, H. Miwa, M.M. Weil, P. Monaghan, T. Jordan, V. Vanheyningen, N.D. Hastie, H. Meijersheijboer, M. Drechsler, B. Royerpokora, F. Collins, A. Swaroop, L.C. Strong and G.F. Saunders.** 1991. Positional Cloning and Characterization of a Paired Box- Containing and Homeobox-Containing Gene from the Aniridia Region. *Cell* **67:** 1059-1074
8. **Staudt, L.M., H. Singh, R. Sen, T. Wirth, P.A. Sharp and D. Baltimore.** 1986. A lymphoid-specific protein binding to the octamer motif of immunoglobulin genes. *Nature* **323:** 640-3
9. **Cox, P.M., S.M. Temperley, H. Kumar and C.R. Goding.** 1988. A distinct octamer-binding protein present in malignant melanoma cells. *Nucleic Acids Res* **16:** 11047-56
10. **Meijer, D., A. Graus, R. Kraay, A. Langeveld, M.P. Mulder and G. Grosveld.** 1990. The octamer binding factor Oct6: cDNA cloning and expression in early embryonic cells. *Nucleic Acids Res* **18:** 7357-65
11. **Schöler, H.R., A.K. Hatzopoulos, R. Balling, N. Suzuki and P. Gruss.** 1989. A family of octamer-specific proteins present during mouse embryogenesis: evidence for germline-specific expression of an Oct factor. *Embo J* **8:** 2543-50
12. **Schreiber, E., P. Matthias, M.M. Müller and W. Schaffner.** 1988. Identification of a novel lymphoid specific octamer binding protein (OTF-2B) by proteolytic clipping bandshift assay (PCBA). *Embo J* **7:** 4221-9
13. **Schreiber, E., K. Harshman, I. Kemler, U. Malipiero, W. Schaffner and A. Fontana.** 1990. Astrocytes and glioblastoma cells express novel octamer-DNA binding proteins distinct from the ubiquitous Oct-1 and B cell type Oct-2 proteins. *Nucleic Acids Res* **18:** 5495-503
14. **Schöler, H.R.** 1991. Octamania: The POU factors in murine development. *Trends Genet* **7:** 323-9
15. **Bodner, M., J.L. Castrillo, L.E. Theill, T. Deerinck, M. Ellisman and M. Karin.** 1988. The pituitary-specific transcription factor GHF-1 is a homeobox-containing protein. *Cell* **55:** 505-18
16. **Clerc, R.G., L.M. Corcoran, J.H. LeBowitz, D. Baltimore and P.A. Sharp.** 1988. The B-cell-specific Oct-2 protein contains POU box- and homeo box-type domains. *Genes Dev* **2:** 1570-81
17. **Ko, H.S., P. Fast, W. McBride and L.M. Staudt.** 1988. A human protein specific for the immunoglobulin octamer DNA motif contains a functional homeobox domain. *Cell* **55:** 135-44
18. **Ingraham, H.A., R.P. Chen, H.J. Mangalam, H.P. Elsholtz, S.E. Flynn, C.R. Lin, D.M. Simmons, L. Swanson and M.G. Rosenfeld.** 1988. A tissue-specific transcription factor containing a homeodomain specifies a pituitary phenotype. *Cell* **55:** 519-29

19. **Müller, M.M., S. Ruppert, W. Schaffner and P. Matthias.** 1988. A cloned octamer transcription factor stimulates transcription from lymphoid-specific promoters in non-B cells. *Nature* 336: 544-51
20. **Scheidereit, C., J.A. Cromlish, T. Gerster, K. Kawakami, C.-G. Balmaceda, R.A. Currie and R.G. Roeder.** 1988. A human lymphoid-specific transcription factor that ctivates immunoglobulin genes is a homeobox protein. *Nature* 336: 551-557
21. **Sturm, R.A., G. Das and W. Herr.** 1988. The ubiquitous octamer-binding protein Oct-1 contains a POU domain with a homeo box subdomain. *Genes Dev* 2: 1582-99
22. **Herr, W., R.A. Sturm, R.G. Clerc, L.M. Corcoran, D. Baltimore, P.A. Sharp, H.A. Ingraham, M.G. Rosenfeld, M. Finney, G. Ruvkun and H.R. Horvitz.** 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: 1513-6
23. **Robertson, M.** 1988. Homoeo boxes, POU proteins and the limits to promiscuity *Nature* 336: 522-4
24. **Sturm, R.A. and W. Herr.** 1988. The POU domain is a bipartite DNA-binding structure. *Nature* 336: 601-4
25. **Agarwal, V.R. and S.M. Sato.** 1991. XLPOU 1 and XLPOU 2, two novel POU domain genes expressed in the dorsoanterior region of Xenopus embryos. *Dev Biol* 147: 363-73
26. **Billin, A.N., K.A. Cockerill and S.J. Poole.** 1991. Isolation of a family of Drosophila POU domain genes expressed in early development. *Mech Dev* 34: 75-84
27. **Dick, T., X.H. Yang, S.L. Yeo and W. Chia.** 1991. Two closely linked Drosophila POU domain genes are expressed in neuroblasts and sensory elements. *Proc Natl Acad Sci USA* 88: 7645-9
28. **Goldsborough, A., A. Ashworth and K. Willison.** 1990. Cloning and sequencing of POU-boxes expressed in mouse testis. *Nucleic Acids Res* 18: 1634
29. **He, X., M.N. Treacy, D.M. Simmons, H.A. Ingraham, L.W. Swanson and M.G. Rosenfeld.** 1989. Expression of a large family of POU-domain regulatory genes in mammalian brain development. *Nature* 340: 35-41
30. **Hinkley, C.S., J.F. Martin, D. Leibham and M. Perry.** 1992. Sequential Expression of Multiple POU Proteins During Amphibian Early Development. *Mol Cell Biol* 12: 638-649
31. **Johnson, W.A. and J. Hirsh.** 1990. Binding of a Drosophila POU-domain protein to a sequence element regulating gene expression in specific dopaminergic neurons. *Nature* 343: 467-70
32. **Lloyd, A. and S. Sakonju.** 1991. Characterization of Two Drosophila POU Domain Genes, Related to oct-1 and oct-2, and the Regulation of Their Expression Patterns. *Mech Dev* 36: 87-102
33. **Monuki, E.S., G. Weinmaster, R. Kuhn and G. Lemke.** 1989. SCIP: a glial POU domain gene regulated by cyclic AMP. *Neuron* 3: 783-93
34. **Okamoto, K., H. Okazawa, A. Okuda, M. Sakai, M. Muramatsu and H. Hamada.** 1990. A novel octamer binding transcription factor is differentially expressed in mouse embryonic cells. *Cell* 60: 461-72
35. **Petryniak, B., L.M. Staudt, C.E. Postema, W.T. McCormack and C.B. Thompson.** 1990. Characterization of chicken octamer-binding proteins demonstrates that POU domain-containing homeobox transcription factors have been highly conserved during vertebrate evolution. *Proc Natl Acad Sci USA* 87: 1099-103
36. **Schöler, H.R., S. Ruppert, N. Suzuki, K. Chowdhury and P. Gruss.** 1990. New type of POU domain in germ line-specific protein Oct-4. *Nature* 344: 435-9
37. **Suzuki, N., H. Rohdewohld, T. Neuman, P. Gruss and H.R. Schöler.** 1990. Oct-6: a POU transcription factor expressed in embryonal stem cells and in the developing brain. *Embo J* 9: 3723-32
38. **Treacy, M.N., X. He and M.G. Rosenfeld.** 1991. I-POU: a POU-domain protein that inhibits neuron-specific gene activation. *Nature* 350: 577-584
39. **Burglin, T., M. Finney, A. Coulson and G. Ruvkun.** 1989. *Caenorhabditis elegans* has scores of homoeobox-containing genes. *Nature* 341: 239-243

40. **Rosner, M.H., M.A. Vigano, K. Ozato, P.M. Timmons, F. Poirier, P.W. Rigby and L.M. Staudt.** 1990. A POU-domain transcription factor in early stem cells and germ cells of the mammalian embryo. *Nature* **345:** 686-92
41. **Parslow, T.G., D.C. Blair, W.J. Murphy and D.K. Granner.** 1984. Structure of the 5′ ends of immunoglobulin genes: a novel conserved sequence. *Proc. Natl. Acad. Sci. USA* **81:** 2650-2654
42. **Mason, J.O., G.T. Williams and M.S. Neuberger.** 1985. Transcription cell type specificity is conferred by an immunoglobulin VH gene promoter that includes a functional consensus sequence. *Cell* **41:** 479-487
43. **Falkner, F.G. and H.Z. Zachau.** 1984. Correct transcription of an immunoglobulin κ gene requires an upstream fragment containing conserved sequence elements. *Nature* **310:** 71-74
44. **Wirth, T., L. Staudt and D. Baltimore.** 1987. An octamer oligonucleotide upstream of a TATA motif is sufficient for lymphoid-specific promoter activity. *Nature* **329:** 174-8
45. **Singh, H., R. Sen, D. Baltimore and P.A. Sharp.** 1986. A nuclear factor that binds to a conserved sequence motif in transcriptional control elements of immunoglobulin genes. *Nature* **319:** 154-8
46. **Mattaj, I.W., S. Lienhard, J. Jiricny and E.M. De Robertis.** 1985. An enhancer-like sequence within the Xenopus U2 gene promoter facilitates the formation of stable transcription complexes. *Nature* **316:** 163-7
47. **Sive, H.L., N. Heintz and R.G. Roeder.** 1986. Multiple sequence elements are required for maximal in vitro transcription of a human histone H2B gene. *Mol Cell Biol* **6:** 3329-40
48. **Pruijn, G.J., R.T. van Miltenburg, J.A. Claessens and P.C. van der Vliet.** 1988. Interaction between the octamer-binding protein nuclear factor III and the adenovirus origin of DNA replication. *J Virol* **62:** 3092-102
49. **LaBella, F., H.L. Sive, R.G. Roeder and N. Heintz.** 1988. Cell-cycle regulation of a human histone H2b gene is mediated by the H2b subtype-specific consensus element. *Genes Dev* **2:** 32-9
50. **Segil, N., S.B. Roberts and N. Heintz.** 1991. Mitotic Phosphorylation of the Oct-1 Homeodomain and Regulation of Oct-1 DNA Binding Activity. *Science* **254:** 1814-1816
51. **Verrijzer, C.P., A.J. Kal and P.C. van der Vliet.** 1990. The DNA binding domain (POU domain) of transcription factor oct-1 suffices for stimulation of DNA replication. *Embo J* **9:** 1883-8
52. **Schöler, H.R., R. Balling, A.K. Hatzopoulos, N. Suzuki and P. Gruss.** 1989. Octamer binding proteins confer transcriptional activity in early mouse embryogenesis. *Embo J* **8:** 2551-7
53. **Lenardo, M.J., L. Staudt, P. Robbins, A. Kuang, R.C. Mulligan and D. Baltimore.** 1989. Repression of the IgH enhancer in teratocarcinoma cells associated with a novel octamer factor. *Science* **243:** 544-6
54. **Scott, M.P., J.W. Tamkun and G.3. Hartzell.** 1989. The structure and function of the homeodomain. *Biochim Biophys Acta* **989:** 25-48
55. **Li, S., E.B. Crenshaw III, E.J. Rawson, D.M. Simmons, L.W. Swanson and M.G. Rosenfeld.** 1990. Dwarf locus mutants lacking three pituitary cell types result from mutations in the POU-domain gene pit-1. *Nature* **347:** 528-33
56. **Kuhn, R., E.S. Monuki and G. Lemke.** 1991. The gene encoding the transcription factor SCIP has features of an expressed retroposon. *Mol Cell Biol* **11:** 4642-50
57. **Yeom, Y.I., H.S. Ha, R. Balling, H.R. Schöler and K. Artzt.** 1991. Structure, Expression and Chromosomal Location of the Oct-4 Gene. *Mech Dev* **35:** 171-179
58. **Okazawa, H., K. Okamoto, F. Ishino, K.T. Ishino, S. Takeda, Y. Toyoda, M. Muramatsu and H. Hamada.** 1991. The oct3 gene, a gene for an embryonic transcription factor, is controlled by a retinoic acid repressible enhancer. *Embo J* **10:** 2997-3005
59. **Kissinger, C.R., B.S. Liu, E. Martin-Blanco, T.B. Kornberg and C.O. Pabo.** 1990. Crystal structure of an engrailed homeodomain-DNA complex at 2.8 Å resolution: a framework for understanding homeodomain-DNA interactions. *Cell* **63:** 579-90

60. **Schleif, R.** 1988. DNA binding by proteins. *Science* **241:** 1182-7
61. **Ingraham, H.A., S.E. Flynn, J.W. Voss, V.R. Albert, M.S. Kapiloff, L. Wilson and M.G. Rosenfeld.** 1990. The POU-specific domain of Pit-1 is essential for sequence-specific, high affinity DNA binding and DNA-dependent Pit-1-Pit-1 interactions. *Cell* **61:** 1021-33
62. **Hanes, S. and R. Brent.** 1989. DNA specificity of the bicoid activator protein is determined by homeodomain recognition helix residue 9. *Cell* **57:** 1275-1283
63. **Treisman, J., P. Gonczy, M. Visishtha, E. Harris and C. Desplan.** 1989. A single amino acid can determine the DNA binding specificity of homeodomain proteins. *Cell* **59:** 553-562
64. **Verrijzer, C.P., A.J. Kal and P.C. van der Vliet.** 1990. The oct-1 homeo domain contacts only part of the octamer sequence and full oct-1 DNA-binding activity requires the POU-specific domain. *Genes Dev* **4:** 1964-74
65. **Verrijzer, C.P., J.A.W.M. van Oosterhout, W.W. van Weperen and P.C. van der Vliet.** 1991. POU proteins bend DNA via the POU-specific domain. *Embo J* **10:** 3007-14
66. **Kapiloff, M.S., Y. Farkash, M. Wegner and M.G. Rosenfeld.** 1991. Variable effects of phosphorylation of Pit-1 dictated by the DNA response elements. *Science* **253:** 786-9
67. **O'Hare, P. and C.R. Goding.** 1988. Herpes simplex virus regulatory elements and the immunoglobulin octamer domain bind a common factor and are both targets for virion transactivation. *Cell* **52:** 435-45
68. **Preston, C.M., M.C. Frame and M.E. Campbell.** 1988. A complex formed between cell components and an HSV structural polypeptide binds to a viral immediate early gene regulatory DNA sequence. *Cell* **52:** 425-34
69. **Kristie, T.M. and P.A. Sharp.** 1990. Interactions of the Oct-1 POU subdomains with specific DNA sequences and with the HSV alpha-trans-activator protein. *Genes Dev* **4:** 2383-96
70. **Kristie, T.M., J.H. LeBowitz and P.A. Sharp.** 1989. The octamer-binding proteins form multi-protein--DNA complexes with the HSV alpha TIF regulatory protein. *Embo J* **8:** 4229-38
71. **Goding, C.R. and P. O'Hare.** 1989. Herpes simplex virus Vmw65-octamer binding protein interaction: a paradigm for combinatorial control of transcription. *Virology* **173:** 363-7
72. **Stern, S., M. Tanaka and W. Herr.** 1989. The Oct-1 homoeodomain directs formation of a multiprotein-DNA complex with the HSV transactivator VP16. *Nature* **341:** 624-30
73. **Treacy, M.N., L.I. Neilson, E.E. Turner, X. He and M.G. Rosenfeld.** 1992. Twin of I-POU - A Two Amino Acid Difference in the I-POU Homeodomain Distinguishes an Activator from an Inhibitor of Transcription. *Cell* **68:** 491-505
74. **Verrijzer, C.P., J.A.W.M. van Oosterhout and P.C. van der Vliet.** 1992. The Oct-1 Pou Domain Mediates Interactions Between Oct-1 and Other Pou Proteins. *Mol Cell Biol* **12:** 542-551
75. **Hope, I.A. and K. Struhl.** 1986. Functional dissection of a eukaryotic transcription activator GCN4 of yeast. *Cell* **46:** 885-894
76. **Müller, I.M., W. Schaffner and P. Matthias.** 1990. Transcription factor Oct-2A contains functionally redundant activating domains and works selectively from a promoter but not from a remote enhancer position in non-lymphoid (HeLa) cells. *Embo J* **9:** 1625-34
77. **Tanaka, M. and W. Herr.** 1990. Differential transcriptional activation by Oct-1 and Oct-2: interdependent activation domains induce Oct-2 phosphorylation. *Cell* **60:** 375-86
78. **Gerster, T., C.G. Balmaceda and R.G. Roeder.** 1990. The cell type-specific octamer transcription factor OTF-2 has two domains required for the activation of transcription. *Embo J* **9:** 1635-43
79. **Imagawa, M., A. Miyamoto, M. Shirakawa, H. Hamada and M. Muramatsu.** 1991. Stringent integrity requirements for both trans-activation and DNA-binding in a trans-activator, Oct3. *Nucleic Acids Res* **19:** 4503-8
80. **Meijer, D., A. Graus and G. Grosveld.** 1992. Mapping the transactivation domain of the Oct6 POU transcription factor. submitted for publication
81. **Theill, L.E., J.L. Castrillo, D. Wu and M. Karin.** 1989. Dissection of functional domains

of the pituitary-specific transcription factor GHF-1. *Nature* **342:** 945-8
82. **Mitchell, P.J. and R. Tjian.** 1989. Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. *Science* **245:** 371-8
83. **Lewin, B.** 1990. Commitment and activation at pol II promoters: a tail of protein-protein interactions. *Cell* **61:** 1161-4
84. **He, X., R. Gerrero, D.M. Simmons, R.E. Park, C.J. Lin, L.W. Swanson and M.G. Rosenfeld.** 1991. Tst-1, a member of the POU domain gene family, binds the promoter of the gene encoding the cell surface adhesion molecule Po. *Mol Cell Biol* **11:** 1739-44
85. **Monuki, E.S., R. Kuhn, G. Weinmaster, B.D. Trapp and G. Lemke.** 1990. Expression and activity of the POU transcription factor SCIP. *Science* **249:** 1300-3
86. **Lemke, G., E. Lamar and J. Patterson.** 1988. Isolation and analysis of the gene encoding peripheral myelin protein zero. *Neuron* **1:** 73-83
87. **Finney, M. and G. Ruvkun.** 1990. The unc-86 gene product couples cell lineage and cell identity in C. elegans. *Cell* **63:** 895-905
88. **Chalfie, M. and M. An.** 1989. Genetic control of differentiation of the *Caenorhabditis elegans* touch receptor neurons. *Science* **243:** 1027-1033
89. **Simmons, D.M., J.W. Voss, H.A. Ingraham, J.M. Holloway, R.S. Broide, M.G. Rosenfeld and L.W. Swanson.** 1990. Pituitary cell phenotypes involve cell-specific Pit-1 mRNA translation and synergistic interactions with other classes of transcription factors. *Genes Dev* **4:** 695-711
90. **Dolle, P., J.L. Castrillo, L.E. Theill, T. Deerinck, M. Ellisman and M. Karin.** 1990. Expression of GHF-1 protein in mouse pituitaries correlates both temporally and spatially with the onset of growth hormone gene activity. *Cell* **60:** 809-20
91. **Mangalam, H.J., V.R. Albert, H.A. Ingraham, M. Kapiloff, L. Wilson, C. Nelson, H. Elsholtz and M.G. Rosenfeld.** 1989. A pituitary POU domain protein, Pit-1, activates both growth hormone and prolactin promoters transcriptionally. *Genes Dev* **3:** 946-58
92. **Karin, M., J.L. Castrillo and L.E. Theill.** 1990. Growth hormone gene regulation: a paradigm for cell-type-specific gene activation. *Trends Genet* **6:** 92-6
93. **McCormick, A., H. Brady, L.E. Theill and M. Karin.** 1990. Regulation of the pituitary-specific homeobox gene GHF1 by cell-autonomous and environmental cues. *Nature* **345:** 829-32
94. **Chen, R.P., H.A. Ingraham, M.N. Treacy, V.R. Albert, L. Wilson and M.G. Rosenfeld.** 1990. Autoregulation of pit-1 gene expression mediated by two cis-active promoter elements. *Nature* **346:** 583-6
95. **Crenshaw III, E.B., K. Kalla, D.M. Simmons, L.W. Swanson and M.G. Rosenfeld.** 1989. Cell-specific expression of the prolactin gene in transgenic mice is controlled by synergistic interactions between promoter and enhancer elements. *Genes Dev* **3:** 959-72
96. **Rosenfeld, M.G.** 1991. POU-domain transcription factors: pou-er-ful developmental regulators. *Genes & Dev.* **5:** 897-907

The EMBO Journal vol.6 no.13 pp.4041–4048, 1987

Molecular characterization of the testis specific *c-abl* mRNA in mouse

Dies Meijer, André Hermans, Marieke von Lindern, Ton van Agthoven, Annelies de Klein, Petra Mackenbach[1], Anton Grootegoed[1], Daniela Talarico[2], Giuliano Della Valle[2] and Gerard Grosveld

Departments of Cell Biology and Genetics and [1]Biochemistry, Erasmus University, 3000 DR Rotterdam, The Netherlands and [2]Dipartimento di Genetica e Microbiologia, Universita di Pavia, 27200 Pavia, Italy

Communicated by F.G.Grosveld

The *c-abl* gene encodes a protein tyrosine kinase and is transcribed from at least two promoters giving rise to transcripts of two size classes of ~5 and 6 kb in length. These mRNAs only differ in their most 5′ exon and encode proteins of similar size but with different N-termini. In the mouse testis an additional abundant *c-abl* mRNA of 4 kb is detected. This mRNA was shown to be expressed in the haploid male germ cells of the adult mouse. Here we describe the cloning and molecular characterization of a cDNA representing the testis specific *c-abl* transcript. We show that the 4 kb *c-abl* mRNA arises from alternative polyadenylation of an RNA transcribed from the same promoter as the 5 kb mRNA. The site of polyadenylation is unusual in this shorter transcript as it is not preceded by the highly conserved hexanucleotide AAUAAA. The use of this polyadenylation site removes 1.2 kb of 3′ sequences present in the somatic *c-abl* mRNAs, but does not affect the main open reading frame of the transcript. Using *in situ* hybridization on whole testis sections it is shown that the 4 kb *c-abl* mRNA is most abundant in the elongating spermatids.

Key words: *c-abl*/spermatogenesis/polyadenylation/differential expression/cDNA

Introduction

The mouse cellular *c-abl* gene is the homologue of the *v-abl* oncogene, present in the genome of the acutely transforming Abelson murine leukaemia virus (AMuLV). The *c-abl* gene is highly conserved during evolution and *c-abl* related genes have been detected in the genomes of a large variety of organisms such as hamster, chicken, rabbit, man (Goff *et al.*, 1980), *Drosophila melanogaster* (Shilo and Weinberg, 1981) and *Caenorhabdites elegans* (Goddard *et al.*, 1986). The *c-abl* gene was shown to be a single copy locus in the mouse genome (Wang *et al.*, 1984). In all murine tissues examined, the gene is transcribed into mainly two RNA species of ~6 and 5 kb that are translated into a 150 kd protein, exhibiting tyrosine specific protein kinase activity (Konopka and Witte, 1985). The genomic organization and the relationship between the two mRNAs have now been described (Wang *et al.*, 1984, Ben-Neriah *et al.*, 1986). The *c-abl* gene consists of a body of 10 small exons and one large 3′ exon, spanning >30 kb of mouse DNA. Recent cDNA cloning of *c-abl* mRNAs from a mouse lymphoid cell line identified an additional series of four 5′ exons (Ben-Neriah *et al.*, 1986). Each of these exons can be spliced onto the common set of 11 body exons, producing four different *c-abl* mRNAs with heterogeneous 5′ ends. This is the basis for the difference in size between the two *c-abl* mRNAs, as the two predominantly used 5′ exons (called type I and type IV) differ by >1 kb in length. These type I and type IV exons are conserved in humans and they are both preceded by a transcriptional promoter (Shtivelman *et al.*, 1986; Bernards *et al.*, 1987). So the two major *c-abl* mRNAs are thus transcribed from two separate promoters. The remaining two transcripts, i.e. type II and type III mRNAs, are probably very rare. The type I and type IV transcripts encode proteins of 1123 and 1142 amino acids and differ only in their most N-terminal 26 and 45 amino acids, respectively. Both N-termini are relatively hydrophilic and do not resemble transmembrane or signal sequences. However, the type IV mRNA encoded protein initiates with the sequence Met-Gly-Gln, which is used as an acceptor site for myristic acid in the *v-abl* and *scr*-proteins. Such a hydrophobic modification might result in membrane anchorage of the protein.

None of the *c-abl* transcripts described above can account for an additional *c-abl* mRNA of 3.7–4.7 kb found in the mouse testis (Müller *et al.*, 1982). This mRNA was shown to be specifically expressed in the haploid germ cells, present in the mature testis (Ponzetto and Wolgemuth, 1985). Here we describe the cloning and characterization of this testis specific *c-abl*

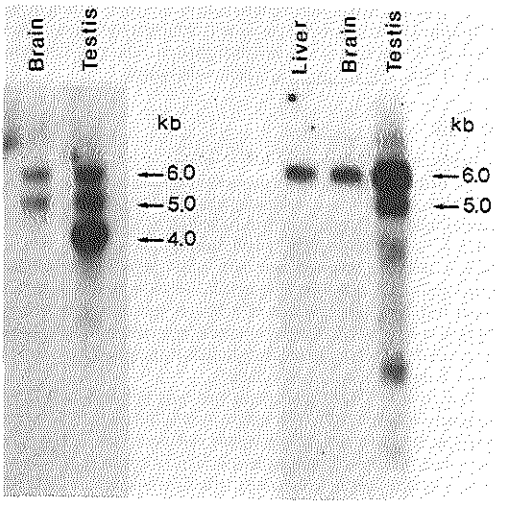


Fig. 1. Expression of the *c-abl* gene in different tissues. Equal amounts of poly(A)$^+$ RNA (10 μg) were run on denaturing 0.8% agarose gels, transferred to nitrocellulose and hybridized with (a) a *c-abl* cDNA probe (covering the protein tyrosine kinase domain) or (b) a type IV specific *c-abl* cDNA probe (Ben-Neriah *et al.*, 1986).

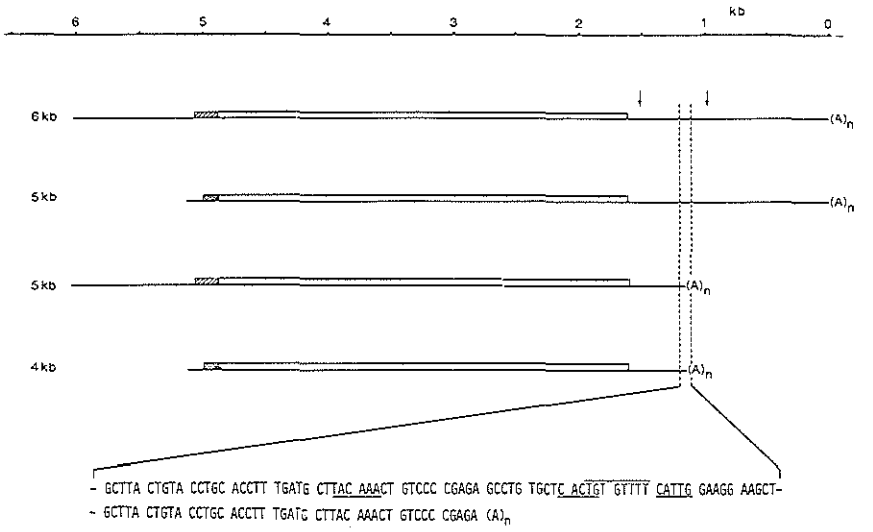


Fig. 2. *c-abl* mRNAs expressed in testis. Schematic representation of the *c-abl* mRNAs expressed in the mouse testis. The ORF of the mRNAs is indicated by a box. Differences in the ORF, contributed by the type I and type IV exons are dotted or lined, respectively. The sequence shows at which position polyadenylation takes place in the shorter transcripts. The overlined and underlined sequences are discussed in the text.

transcript. We show that this *c-abl* mRNA is transcribed from the type I promoter and is alternatively polyadenylated. RNAs transcribed from the type IV promoter can be alternatively polyadenylated as well, but they represent a minor species in testis. The encoded *c-abl* protein remains unaltered. Using *in situ* hybridization we show that the short *c-abl* transcript is most abundant in the later stages of spermatogenesis. The possible function of this mRNA in germ cells is discussed.

Results

Cloning of a testis specific c-abl cDNA

RNA extracted from BCBA mouse brain and testis was poly(A) selected, separated on a denaturing agarose/formaldehyde gel and blotted onto nitrocellulose. Hybridization of this Northern blot with a *c-abl* cDNA probe, encompassing the protein tyrosine kinase domain, showed that the *c-abl* gene expresses two transcripts of 5 and 6 kb (Figure 1a). In testis the predominant *c-abl* mRNA is a transcript of 4 kb, as was shown before by Müller *et al.* (1982). In order to characterize this mRNA in more detail cDNA molecules representing this transcript were cloned from a mouse testis cDNA library.

The poly(A)$^+$ RNA from testis was used to synthesize cDNA primed with oligo(dT). The cDNA was tailed with dGTP using terminal deoxynucleotidyl transferase and cloned into λgt10 using a synthetic *Eco*RI-dC adaptor according to LeBouc *et al.* (1986). Recombinant phages were plated onto *Escherichia coli* BNN102 and the resulting plaques (1×10^6) were screened with a 3′ *v-abl* *Sal*I–*Hin*dIII probe (Reddy *et al.*, 1983). More than 50 positive clones were obtained. Inserts from phages longer than 1.2 kb (12) were subcloned into the *Eco*RI site of pUC19 and mapped for the presence of known restriction enzyme cleavage sites. Based on their restriction map, the cDNAs could be grouped into two classes. One class (two clones) contained inserts, that were colinear with the previously described 1.2 kb 3′ cDNA clone, isolated from a mouse L-cell cDNA library by Wang *et al.* (1984). It contains a poly(A) signal and terminates in a poly(A) tail. This cDNA represents the 3′ end of both 6 and 5 kb *c-abl* mRNAs.

The second class of clones were colinear with *v-abl* sequences, but were lacking all sequences contained in the 3′ cDNA clone mentioned above. The restriction maps indicated that all cDNA inserts in this class were primed at exactly the same position, making it unlikely that they were artifacts of cloning.

To determine whether these clones arose from alternatively polyadenylated or spliced transcripts, the 3′ part of three independently derived cDNAs was sequenced. It was found that all three were colinear with the published *v-abl* sequence up to nucleotide 4469 (Reddy *et al.*, 1983). The 3′ sequence of these cDNAs is depicted in Figure 2. From these results we conclude that the clones arose from polyadenylated transcripts, containing a poly(A) tail 1171 nucleotides upstream of the normal poly(A) addition site, used in the generation of the somatic 5 and 6 kb transcripts (Wang *et al.*, 1984, Ben-Neriah *et al.*, 1986). S1 nuclease mapping further showed that this poly(A) addition site is abundantly used in testis, whereas it is not used to a detectable extent in brain (Figure 3B).

As the longest testis specific cDNA is only 2.6 kb, we constructed a full length clone (TS *c-abl*) with 2.5 kb of upstream genomic sequences, using two other independently derived partial cDNAs and a genomic DNA fragment containing the type I exon (Figure 3A). The latter was isolated from a mouse genomic library, using a type I exon probe. To check the integrity of this construct, we performed S1 nuclease mapping experiments on testis RNA using brain or cell line 70Z RNA as control. (70Z is a pre-B cell line of Balb/c origin, known to express relatively high levels of *c-abl* mRNA (Paige *et al.*, 1978; Wang *et al.*, 1983). The results of these experiments are shown in Figure 3A.

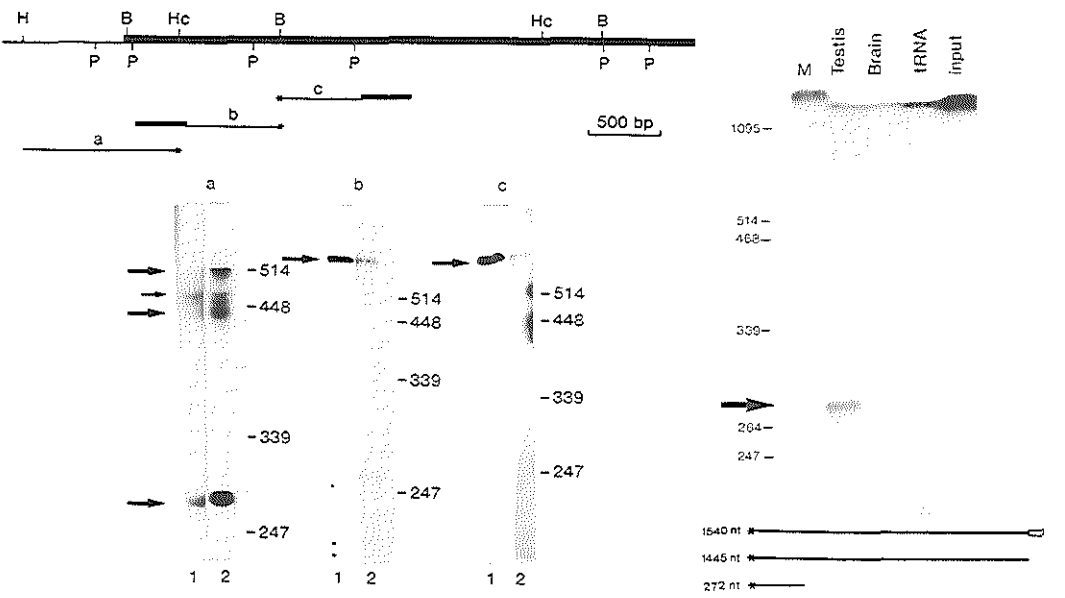


Fig. 3(A). Structure of the testis specific type I cDNA. The upper part of the figure shows the restriction map of the TS *c-abl* cDNA clone. The thick line represents cDNA sequences, while the thin line represents genomic DNA sequences. Restriction sites are indicated by letters. B = *Bgl*II, P = *Pvu*II, H = *Hin*dIII, Hc = *Hin*cII. S1 probes used to check the integrity of the construct are shown below the map (a,b,c). Hybridizations were done at 55°C using 20 μg of total RNA. Probes b and c were derived from subclones of the TS *abl* cDNA in pUC19. The thick lines represent plasmid sequences. Input bands run high up in the gel and are not visible in lanes a, b and c (a) S1 analysis with probe (a) with 1. 70Z RNA, 2. testis RNA (b) probe (b) with 1. testis RNA 2. brain RNA (c) probe c with 1. testis RNA 2. brain RNA. Different exposure times for lanes 1 and 2 are shown to allow a better comparison of the S1 protected bands. (B) Confirmation of the position of the testis specific poly(A) addition site. The S1 probe used is a 1540 bp *Pvu*II fragment derived from a 3' *c-abl* cDNA clone isolated from a brain cDNA library and covers 1445 nt of 3' UTR including the poly(A) consensus sequence used to generate the somatic *c-abl* transcripts. The open box indicate plasmid sequences. The probe was labelled using a 'replacement synthesis' protocol according to Maniatis *et al.* (1982). Protected fragments are indicated below the autoradiogram. Input (1540) and fully protected fragments (1450) are not separated on this gel. The 272 nt fragment protected by testis RNA (arrow) confirms the position of the poly(A) addition site as shown in Figure 2.

The first probe (a) covering the first (type I), the second and part of the third exon yielded protected fragments of 278 nucleotides (nt), a set of two minor fragments of 450 and 455 nt and two major fragments of 435 and 500 nt. The 278 nt fragment corresponds to the 5' common exon border (Ben-Neriah *et al.*, 1986) and represents non-type I mRNAs (mainly type IV). The two sets of major and minor bands probably correspond to different transcriptional start sites of the type I promoter. The two minor bands were also seen when 70Z RNA was used to protect probe (a), while the two major bands were absent. It is possible that the minor fragments represent start sites in diploid cells in the testis, while the major fragments represent the start sites in the haploid spermatids. We did not investigate this observation further.

Probes (b) and (c) derived from our TS *c-abl* clone were fully protected by RNA from testis and brain, (arrows in panels b and c) indicating that no alternative splicing takes place in testis *c-abl* mRNA downstream from the 5' *Hin*dII site (the 5' *Hin*dII site maps within the 3rd exon).

As the last 3' exon, in which alternative polyadenylation occurs, is also present in the 6 kb type IV transcript, the question arose whether this transcript is also alternatively polyadenylated in testis. To address this question, a Northern blot containing poly(A)$^+$ RNA from liver, brain and testis was hybridized with a type IV specific probe. As can be seen in Figure 1b this probe detects in liver and brain only the 6 kb mRNA, while in testis RNA an additional band of 5 kb is observed and one minor band of ~3.5 kb. We do not know the nature of this minor transcript. The 5 kb mRNA probably represents an alternatively polyadenylated type IV transcript. We conclude that both major *c-abl* transcripts are subject to alternative polyadenylation in testis.

Alternative polyadenylation does not affect the main open reading frame

The 3' part of the *c-abl* transcripts is encoded by one large exon of 3 kb (Wang *et al.*, 1984). This exon contains a long untranslated region (UTR, ± 1630 bp). As alternative polyadenylation of *c-abl* mRNAs takes place 1171 nt upstream of the normal poly(A) addition site, the use of this upstream polyadenylation site does not affect the main open reading frame (ORF) and therefore should encode the same protein as the somatic 5 kb mRNA. To obtain further evidence, we used full length cDNA clones representing both type I mRNAs to transcribe RNA *in vitro* in the presence of the Cap analogue 7m GpppG, using T7 RNA polymerase. Both RNAs were translated in a cell free translation system (rabbit reticulocyte lysate) in the presence of [^{35}S]methionine. It was found that both RNAs were translated into proteins with a mol. wt of 150 kd (Figure 4) which could be precipitated with anti *c-abl* antibodies (not shown). Although the *in vitro* transcribed RNA is translated into a 150 kd *c-abl*

Fig. 4. *In vitro* translation of *in vitro* transcribed RNA. RNA was transcribed from constructs representing the testis specific 4 kb *c-abl* mRNA and the somatic 5 kb mRNA. *In vitro* translation products were separated on a 7.5% SDS–polyacrylamide gel. 1. Protein translated from the testis specific 4 kb mRNA. 2. Protein translated from the somatic 5 kb mRNA.

protein this does not necessarily mean that the 4 kb is translated *in vivo* also.

The 3′ untranslated region of different genes has been shown to be involved in regulation of expression (Shaw and Kamen, 1986; Owen and Kühn, 1987). Regulation of the stability of a transcript can be a possible mechanism of control. This hypothesis was examined by measuring the half-life of the *c-abl* transcripts in the cell line 70Z and in isolated spermatids. Results of these experiments are shown in Figure 5a. RNA was extracted from 70Z cells and isolated spermatids before and after treatment with 5 μg/ml of actinomycin D at various times (0–6 h). The RNA samples were subsequently electrophoresed in agarose/formaldehyde gels and blotted onto nitrocellulose. The Northern blot containing 70Z RNA samples was hybridized with a *c-abl*, a *myc* and a glyceraldehyde-3-phosphate-dehydrogenase (GAPDH) probe. The half-life of the *myc* and GAPDH transcripts have been reported and were chosen as reference (Dani *et al.*, 1984). Figure 5a shows that the *myc* transcript is rapidly depleted from the RNA pool, while the GAPDH transcript seems almost constant over the 6 h time period. The half-life of the *myc* transcript was estimated to be ~20 min in accordance with earlier reports (Dani *et al.*, 1984; Linial *et al.*, 1985). The *c-abl* mRNAs of 5 and 6 kb seem to be relatively stable over the first 2 h of actinomycin D incubation, but then rapidly decay with an estimated half-life of ~30 min. The 6 kb *c-abl* transcript seems to be slightly more stable than the 5 kb transcript.

A similar experiment was performed using spermatids isolated from mouse testis. As shown in Figure 5b the results of this experiment clearly indicate the higher stability of the *c-abl* mRNAs in spermatids. In fact the 6 kb *c-abl* mRNA seems to be more stable in spermatids too. This blot was also probed with a *pim-1* probe. *pim-1* is a gene that is frequently activated by proviral insertion in murine leukaemia virus-induced T cell lymphomas (Selten *et al.*, 1986). The *pim-1* gene expresses a 2.8 kb mRNA in lymphoid cells and encodes a protein kinase. In mouse testis the *pim-1* gene is expressed as a 2.3 kb mRNA (M.von Lindern, unpublished results). This 2.3 kb *pim-1* transcript in spermatids is more stable than the 2.8 kb *pim-1* in 70Z cells, where we measured a half-life of ~30 min (data not shown). The *myc* gene is not expressed in spermatids. These results can be explained in two ways. Either all or most transcripts are more stable in isolated spermatids than in somatic cells, or the actinomycin D is not able to block transcription in spermatids. The latter explanation is ruled out since it was observed that 2.5 μg/ml actinomycin D essentially blocks [^{3}H]uridine incorporation in isolated spermatids (2.6×10^6 isolated spermatids incorporated ±1600 c.p.m. [^{3}H]uridine in 1 h while in the presence of 2.5 μg/ml actinomycin D only 20 c.p.m. were incorporated).

In situ hybridization on whole testis sections

Using isolated populations of male germ cells Ponzetto and Wolgemuth (1985) demonstrated that the shorter testis specific *c-abl* mRNA is present in spermatids and elongating spermatids but not in mid-pachytene spermatocytes. We used *in situ* hybridization on whole testis sections to study the expression of the shorter *c-abl* mRNA. The unique structure of the 3′ end of the 4 kb *c-abl* mRNA made it possible to synthesize an oligonucleotide that would only hybridize to the shorter *c-abl* transcripts under appropriate hybridization and washing conditions. The oligonucleotide synthesized was a 30-mer, covering the first 20 nucleotides preceeding the upstream polyadenylation site and 10 adenine residues of the poly(A) tail of the shorter transcript (see Figure 2). Hybridization and washing conditions were selective for a full 30-mer hybrid. The specific hybridization to the testis 4 kb *c-abl* mRNA was checked on Northern blots under these conditions (final stringency 2 × SSC, 56°C). Analysis of cross sections of different seminiferous tubules, containing different stages of maturation indicated that the short *c-abl* mRNA is most abundant in the elongating spermatids. One such cross-section of a tubule is shown in Figure 6a. To rule out the possibility that this oligonucleotide hybridizes non-specifically to structures present in the elongating spermatids, a series of control experiments were performed. We synthesized an oligonucleotide that covers the poly(A) addition site of β-major globin of the mouse and 10 A residues of the poly(A) tail. No labelling above background of any cell type present in the testis was observed using this oligonucleotide. In another experiment we used a DNA probe specific for rRNA. As can be seen in Figure 6b this probe labels all cells, but in particular those close to the tubular wall. No non-specific labelling of late spermatids was observed.

Discussion

Here we have shown that the 4 kb testis specific *c-abl* transcript results from the use of an alternative polyadenylation site 1172 nt upstream of the normal addition site. We have further demonstrated that this alternative polyadenylation site is specific for the haploid stage of spermatogenesis.

Polyadenylation of the *c-abl* transcripts in spermatids is unusual as the site of poly(A) addition is not preceeded by the highly conserved hexanucleotide AAUAAA (Proudfoot and Brownlee 1976). Many studies concerning 3′ processing of polyadenylated transcripts have demonstrated the importance of this hexanucleotide in cleavage/polyadenylation (reviewed in Birnstiel *et al.*, 1985). As shown in Figure 2 a sequence TACAAA is present ~11 nt upstream of the poly(A) addition site used in spermatids and remotely resembles this conserved hexanucleotide. However, this sequence contains two mutations one of which (U–C) was shown to prevent cleavage/polyadenylation in SV40 late transcription (Wickens and Stephenson, 1984). Moreover,

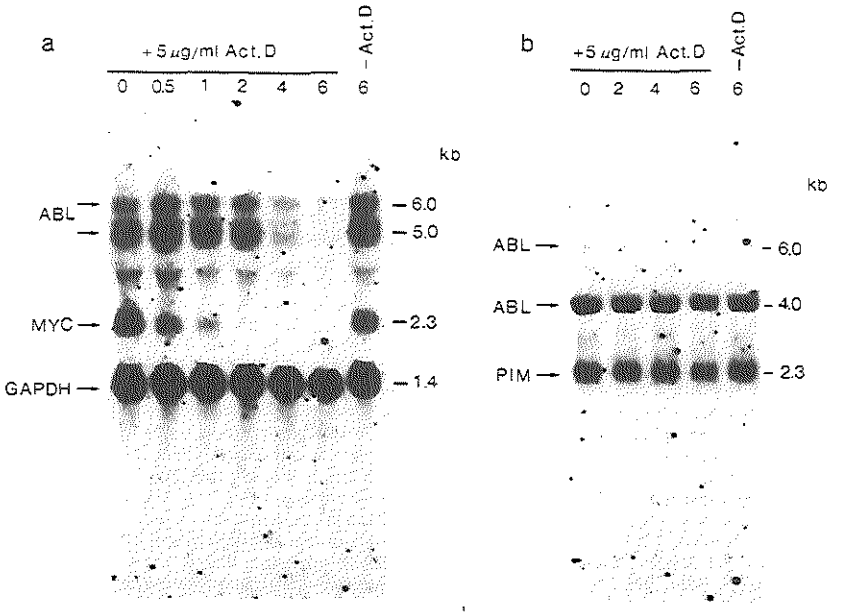


Fig. 5. Actinomycin D chase experiments. Cells and isolated spermatids were treated with or without actinomycin D for various times (0–6 h). RNA was prepared, separated on denaturing 0.8% formaldehyde gels and blotted onto nitrocellulose. The blots were subsequently probed with a *c-abl*, GAPDH and a *myc* probe in the case of 70Z cells (a) and with a *c-abl* and *pim-1* probe in the case of isolated spermatids (b). *myc* and GAPDH were not expressed spermatids. Numbers above the lanes indicate the time of actinomycin D treatment in hours.

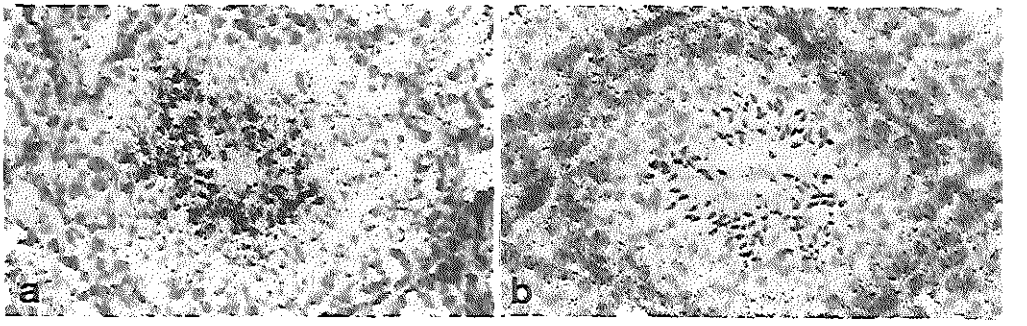


Fig. 6. *In situ* hybridization on whole testis sections using the testis specific *c-abl* oligonucleotide (a) or an rRNA probe (b). The pictures show one cross section of a semininiferous tubule.

the same transition in the β-globin polyadenylation sequence was found to be the sole cause of a β-thallassaemia in humans (Orkin *et al.*, 1985).

Although these studies show that the AAUAAA element is necessary for proper 3′ processing of pre-RNAs, it cannot be the only prerequisite, as this sequence frequently occurs in transcripts, where it does not direct cleavage/polyadenylation. It has been shown that in addition to the upstream AAUAAA signal, sequences downstream from the addition site are required as well (Kessler *et al.*, 1986; McDevitt *et al.*, 1986; Gil and Proudfoot, 1987). These sequence elements are very weakly conserved and are generally referred to as GT or T-rich sequences. McLauchlan *et al.* (1985) compiled a large number of downstream sequences from different genes and derived a consensus sequence YGTGTTYY (Y = pyrimidine residue) located ~30 nt downstream of the addition site in 67% of the genes examined. A perfect match to this consensus sequence is found at the appropriate distance downstream from the spermatid specific poly(A) addition site (overlined in Figure 2). Berget (1984) also provided evidence for the conservation of a sequence CAYUG, located immediately downstream and upstream from the poly(A) addition site in a number of genes. Two pentanucleotides, perfectly matching this consensus, are also present downstream of the testis specific addition site (underlined in Figure 2).

It is very likely that these sequence elements play a role in the formation of the shorter *c-abl* transcripts. Nevertheless, a cell type specific factor or mechanism must exist that discriminates between the two poly(A) addition sites that can be used. It has been suggested that small ribonucleoprotein particles (snRNPs) are involved in 3′ processing of mRNAs on the basis of partial complementarity between the small U4-RNA and sequences surrounding the poly(A) addition site, including the highly conserved hexanucleotide and the more weakly conserved CAYUG sequence (Berget, 1984). Although there is no direct evidence for a role of U-RNAs in 3′ processing of polyadenylated mRNAs, involvement of small RNAs has been demonstrated for the maturation of a non-polyadenylated histon mRNA in sea urchin (see Birnstiel *et al.*, 1985). It is tempting to speculate that different U-RNAs might direct the choice of different potential cleavage sites to be used.

Alternative polyadenylation has been demonstrated for a number of genes. In the calcitonin/CGRP gene (Amara *et al.*, 1984) and the mouse μ-immunoglobulin gene (Early *et al.*, 1980), polyadenylation can occur in two different coding exons, thereby directing the protein to be synthesized. In contrast, alternative polyadenylation in testis of the *c-abl* transcripts does not affect the coding capacity of the transcripts, as was also observed in the case of the mouse α-amylase (Tosi *et al.*, 1981) chicken vimentin, (Zehner and Paterson, 1983), chicken ovalbumin (LeMeur *et al.*, 1984) and mouse dihydrofolate reductase (DHFR) gene (Setzer *et al.*, 1980, 1982). For the latter genes, no tissue specificity in the use of poly(A) addition site was observed although Kaufman and Sharp (1983) demonstrated a correlation with the metabolic state of the cell in the case of DHFR polyadenylation. The minor polyadenylation sites utilized in the shorter ovalbumin and DHFR transcripts are not preceeded by the polyadenylation signal AAUAAA as well.

It is of interest to note that the *Drosophila c-abl* homologue, Dash, contains a large 3′ exon, that is processed in several alternative ways (Telford *et al.*, 1985). S1 mapping of the 3′ end revealed five different ends, presumably corresponding to different polyadenylation sites. Three of these sites are used in the ovaries of the fly while the other two are utilized during development. The material specific poly(A) addition sites generate shorter transcripts. In this respect the 3′ processing of the Dash mRNAs in *Drosophila* ovaries resembles the 3′ processing of the *c-abl* mRNAs in the mouse male germinal cells.

It was shown by Ponzetto and Wolgemuth, 1985 and illustrated in Figure 5b, that the 4 kb mRNA is the predominant *c-abl* transcript present in isolated spermatids. The high abundance of this mRNA, compared to the normal mRNAs could be due to several mechanisms: (i) The 4 kb *c-abl* mRNA is more stable and has a longer half-life than the normal mRNAs, due to deletion of sequences that make the transcript unstable. (ii) The truncated mRNA is the only transcript produced during the haploid expression of the *c-abl* gene and continues to a far later stage in the developing germ cells than expression of the normal mRNAS. (iii) A combination of possibilities (i) and (ii).

(i) The 3′ untranslated region of a number of genes have been shown to play an important role in gene expression. Shaw and Kamen (1986) demonstrated the existence of a specific AU-rich sequence (AUUUA) within the 3′ UTR of a number of transiently expressed genes, that renders the transcripts unstable. Owen and Kühn (1987) recently demonstrated that the non-coding 3′ part of the transferrin receptor gene is required for regulation by iron. The *c-abl* gene of the mouse contains a long 3′ UTR that contains sequences that are conserved in the human *c-abl* gene, as detected by cross hybridization at relatively high stringency (0.3 × SSC, 56°C, data not shown). These sequences are removed from the mRNAs, by the use of the upstream polyadenylation site in spermatids and might be an important mechanism of regulation. Part of the 3′ UTR of mouse *c-abl* and all of the 3′ UTR contained in *v-abl* has been sequenced (Wang *et al.*, 1984; Reddy *et al.*, 1983). Although the 3′ UTR does contain an AT-rich stretch, it does not contain a copy of the destabilizer motif described by Shaw and Kamen (1986). However, our results with the actinomycin D chase experiments indicate that the shorter TS transcript is more stable in spermatids than the regular *c-abl* mRNAs in 70Z cells. In the latter, the *c-abl* transcripts display a relatively high stability, but after 2 h they rapidly decay ($t_{½} \sim 30$ min). This initial relative stability of the *c-abl* transcripts can be explained by the presence of a pool of nuclear pre-RNAs that replenish the cytoplasmic pool of mRNA by RNA splicing after transcription has stopped, until the pre-RNA pool is exhausted. We are currently testing this hypothesis. However, alternative explanations are also possible.

In isolated spermatids all transcripts were found to be stable over the 6 h time course. This makes it impossible to test our hypothesis on a longer half-life of the 4 kb *c-abl* mRNA.

Also, the shorter *pim-1* transcript is more stable than the normal 2.8 kb transcript which has a half-life of ~30 min. Interestingly, the normal *pim-1* mRNA contains a tandem of AUUUA sequences (Selten *et al.*, 1986) in its 3′ UTR. Preliminary data suggest that these sequences are no longer present in the testis specific *pim* transcript, which might explain the higher stability of this transcript in spermatids.

The data discussed above provide circumstantial evidence for a higher stability of the testis specific shorter *c-abl* transcript. Proof for the involvement of the *c-abl* 3′ UTR in determining messenger stability must await mutational analysis and expression of the TS *c-abl* gene construct in normal cells. A higher stability of mRNAs might well be a prerequisite for genes expressed in spermatids. For instance it has been shown that the mouse protamine mRNAs are present in spermatids, days before they are translated during steps 12–15 of spermiogenesis, indicating that they are under translational control (Hecht, 1986). Although the exact time is not known, it is believed that transcription stops around step 12 of spermiogenesis. From this point on, the cell continues to change its morphology, which is probably mediated by the regulated translation of mRNA, transcribed at an earlier stage in spermiogenesis.

(ii) Alternatively, accumulation of the 4 kb *c-abl* transcript in the late stages of spermiogenesis is not due to increased stability of the mRNA but to late transcription of this mRNA. This raises the question of what function the truncation of the mRNA serves. Possibly, it affects the translation of the mRNA, by altering the conformation of the transcript, or changes its ability to bind cellular factors. As the short *c-abl* transcript is present in large quantities until the later stages of spermiogenesis, the *c-abl* protein could play a role in the terminal differentiation of the germ cells.

Materials and methods

Sources of tissue and cell lines

Adult BCBA and NIH Swiss mice were used as sources of tissue in all experiments. Animals were killed by cervical dislocation and the tissues were immediately frozen in liquid nitrogen and subsequently used to extract RNA. The 70Z cell line is a pre-B lymphoid cell of Balb/c origin (Paige *et al.*, 1978). This cell line was grown in RPMI medium, supplemented with 15% new born calf serum and 70 μM β-mercaptoethanol.

Cell separation

Spermatogenic cells were isolated from decapsulated testis from NIH Swiss mice, using collagenase and trypsin. The cells were subsequently fractionated, using sedimentation at unit gravity (Staput procedure), followed by density gradient centrifugation (Percoll gradients) as described by Grootegoed *et al.* (1986). The different cell populations were immediately frozen in liquid nitrogen for subsequent RNA extraction. Spermatids for actinomycin D chase experiments were incubated in Eagles MEM, supplemented with 5 mM sodium-L-lactate and 0.4% (w/v) BSA (fraction V) and used immediately. To measure the inhibition of transcription in isolated spermatids by actinomycin D, 2.6×10^6 spermatids were incubated for 1 h in medium (as above) containing 2.5 μCi/ml [^{3}H]uridine (5 Ci/mmol) with or without actinomycin D (2.5 μg/ml) at 37°C, 5% CO_2. Cells were washed twice with PBS/0.5 mM uridine before [^{3}H]uridine incorporated in TCA precipitated material was measured.

RNA isolation and Northern blot hybridization

RNA was isolated using either the LiCl/urea (see Maniatis *et al.*, 1982) or the guanidinium-isothiocyanate protocol (Collins *et al.*, 1984). Poly(A)$^+$ RNA was selected by one or two cycles of oligo(dT) cellulose (Boehringer) chromatography. RNA was denatured in formamide/formaldehyde at 55°C for 15 min prior to electrophoresis on denaturing 0.8% agarose/formaldehyde gels, blotted onto nitrocellulose for 4–14 h and baked for 1 h at 80°C. Prehybridization of filters was carried out for 1 h at 42°C in hybridization liquid containing 50% formamide, 10% dextransulfate, 3 × SSC, 0.1% SDS and 100 μg/ml denatured salmon sperm DNA. Probes for hybridization were labelled with ^{32}P by random oligonucleotide labelling (Feinberg and Vogelstein, 1983), denatured in boiling water for 5 min quenched on ice and directly added to the hybridization mix. Following overnight hybridization at 42°C the filters were washed to a final stringency of 0.3 × SSC, 65°C. Filters were sealed in a plastic bag and exposed to Kodak XAR film at −70°C for various lengths of time.

cDNA library preparation and screening

RNA was isolated from testes using the guanidinium-isothiocyanate method, poly(A)$^+$ selected twice, as described above and used to prepare a cDNA library in phage λgt10 (Huynh *et al.*, 1984). Oligo(dT) was used to prime the first strand synthesis using reverse transcriptase. Second strand cDNA was obtained with Klenow enzyme after tailing of the first strand with dGTP and TdT using oligo(dC) as a primer. Double stranded cDNA was tailed with dGTP and TdT and cloned into λgt10 using a synthetic *Eco*RI adapter (Leboue *et al.*, 1986). Recombinant phages were plated on *E.coli* strain BNN102 (1×10^6 p.f.u.) and grown to confluency. Duplicate filters were prepared and screened with a 3′ *v-abl* probe *Sall–HindIII* 1.1 kb, according to Maniatis *et al.* (1982). Positive clones were plaque-purified in a second screening and the inserts were subcloned into the *Eco*RI site of pUC19 (Yanisch-Perron *et al.*, 1985).

In situ hybridization

In situ hybridization on whole tissue sections was performed according to a protocol developed by two of us (D.T. and G.D.V.). This protocol is based on published methods, used for *in situ* hybridization on *Drosophila* sections (Akam, 1983, Hafen *et al.*, 1983). A detailed description of this method will be published elsewhere by G.D.V. and D.T. The following oligonucleotides were synthesized, using an Applied Biosystems DNA synthesizer. Testis specific *c-abl* oligonucleotide: GCTTACAAACTGCCCCGAGAAAAAAAAAAA β-globin major oligonucleotide: ATAAAAAGCATTTATGTTCACTGCAAAAAAAAAA. The melting temperature of RNA–DNA duplexes formed by these oligonucleotides was calculated using the Wallace rule (Suggs *et al.*, 1981). The oligonucleotides were labelled by priming, using [^{35}S]dATP and Klenow enzyme.

DNA sequencing

DNA fragments end labelled at appropriate sites were sequenced, using the chemical degradation method of Maxam and Gilbert (1980).

S1 mapping

Total RNA (20 μg) was hybridized with heat denatured ^{32}P. 5′ or 3′ end labelled DNA fragments according to established methods (see Maniatis *et al.*, 1982) at 55°C for 18 h. S1 digestion was carried out at 37°C with 1000 U/ml of the enzyme (Boehringer) in 300 μl digestion buffer (300 mM NaCl/50 mM Na-acetate pH 4.4; 2.5 mM $ZnSO_4$).

In vitro transcription/translation

Fragments to be transcribed were cloned in the polylinker of pTZ19 (Pharmacia) 3′ of the T7 promoter. RNA was transcribed with T7 polymerase (Biolabs) in the presence of 250 μM GpppG (Pharmacia), according to the manufacturers protocol (Biolabs). After removal of the template by DNaseI digestion, the RNA was purified by phenol/chloroform extraction and precipitated three times with 2M NH_4-acetate and 3 vol ethanol. RNA was subsequently translated in a cell free translation system (Rabbit reticulocyte lysate Promega), in the presence of [^{35}S]methionine. Samples were electrophoresed on 7.5% SDS–polyacrylamide gel. The gel was treated with enhancer (Amersham), dried and exposed to autoradiographic film (Kodak XAR-5).

Acknowledgements

We thank André Bernards for the type IV *c-abl* probe and Jan Hoeijmakers and Dirk Bootsma for critical reading of the manuscript. We are indebted to Tar van Os, Cor van Dijk and Rita Boucke for help with the preparation of the manuscript.

References

Akam,M.E. (1983) *EMBO J.*, 2, 2075–2084.
Amara,S.G., Evans,R.M. and Rosenfeld,M.G. (1984) *Mol. Cell. Biol.*, 4, 2151–2160.
Ben-Neriah,Y., Bernards,A., Paskind,M., Daley,G.Q. and Baltimore,D. (1986) *Cell*, 44, 577–586.
Berget,S.M. (1984) *Nature*, 309, 179–182.
Bernards,A., Rubin,C.M., Westbrook,C.A., Paskind,M. and Baltimore,D. (1987) *Mol. Cell. Biol.*, 7, 3231–3236.
Birnstiel,M.L., Busslinger,M. and Strub,K. (1985) *Cell*, 41, 349–359.
Collins,S.J., Kubonishi,I., Miyoshi,I. and Groudine,M. (1984) *Science*, 225, 72–74.
Dani,Ch., Blanchard,J.M., Piechaczyk,M., El Soubaty,S., Marty,L. and Jeanteur,Ph. (1984) *Proc. Natl. Acad. Sci. USA*, 81, 7046–7050.
Early,P., Rogers,J., Davis,M., Calame,K., Bond,M., Wall,R. and Hood,L. (1980) *Cell*, 20, 313–319.
Feinberg,A.P. and Vogelstein,B. (1983) *Anal. Biochem.*, 132, 6–13.
Gil,A. and Proudfoot,N.J. (1987) *Cell*, 49, 399–406.
Goddard,J.M., Weiland,J.J. and Capecchi,M.R. (1986) *Proc. Natl. Acad. Sci. USA*, 83, 2172–2176.
Goff,S.P., Gilboa,E., Witte,O.N. and Baltimore,D. (1980) *Cell*, 22, 777–785.
Grootegoed,J.A., Jansen,R. and van der Molen,H.J. (1986) *J. Reprod. Fert.*, 77, 99–107.
Hafen,E., Levine,M., Garber,R.L. and Gehring,W.J. (1983) *EMBO J.*, 2, 617–623.
Hecht,N.B. (1986) In Stefaninin,M., Conti,M., Geremia,R., Ziparo,E. (eds) *Molecular and Cellular Endocrinology of the Testis*. Elsevier Science Publishers, Amsterdam, pp. 199–213.
Huynh,T.V., Young,R.A. and Davis,R.W. (1984) in Glover, D. (ed.) *DNA Cloning Techniques: a practical approach*. IRL Press, Oxford, pp. 49–78.
Kaufman,R.J. and Sharp,P.A. (1983) *Mol. Cell. Biol.*, 3, 1598–1608.
Kessler,M.M., Beckendorf,R.C., Westhafer,M.A. and Nordstrom,J.L. (1986) *Nucleic Acids Res.*, 14, 4939–4952.
Konopka,J.B. and Witte,O.N. (1985) *Biochem. Biophys. Acta*, 823, 1–17.
Le Bouc,Y., Dreyer,D., Jaeger,F., Binoux,M. and Sondermeyer,P. (1986) *FEBS Lett.*, 196, 108–112.
LeMeur,M.A., Galliot,B. and Gerlinger,P. (1984) *EMBO J.*, 3, 2779–2786.
Linial,M., Gunderson,N. and Groudine,M. (1985) *Science*, 230, 1126–1132.
Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning, A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Maxam,A. and Gilbert,W. (1980) *Methods Enzymol.*, 65, 499–560.
McDevitt,M.A., Hart,R.P., Wong,W.W. and Nevins,J.R. (1986) *EMBO J.*, 5, 2907–2913.
McLauchlan,J., Gaffney,D., Whitton,J.L. and Clements,J.B. (1985) *Nucleic Acids Res.*, 13, 1347–1368.
Müller,R., Slamon,D.J., Tremblay,J.M., Cline,M.J. and Verma,I.M. (1982) *Nature*, 299, 640–644.
Orkin,S.H., Cheng,T.C., Antonarakis,S.E. and Kazazian,H.H., Jr (1985) *EMBO J.*, 4, 453–456.
Owen,D. and Kühn,L.C. (1987) *EMBO J.*, 6, 1287–1293.
Paige,C.J., Kincade,P.N. and Ralph,P. (1978) *J. Immunol.*, 121, 641–648.
Ponzetto,C. and Wolgemuth,D.J. (1985) *Mol. Cell. Biol.*, 5, 1791–1794.
Proudfoot,N.J. and Brownlee,G.G. (1976) *Nature*, 263, 211–214.
Reddy,E.P., Smith,M.J. and Srinivasan,A. (1983) *Proc. Natl. Acad. Sci. USA*, 80, 3623–3627.
Selten,G., Cuypers,H.T., Boelens,W., Robanus-Maandag,E., Verbeek,J., Domen,J., van Beveren,C. and Berns,A. (1986) *Cell*, 46, 603–611.
Setzer,D.R., McGrogan,M. and Schimke,R.T. (1982) *J. Biol. Chem.*, 257, 5143–5147.
Setzer,D.R., McGrogan,M., Nunberg,J.H. and Schimke,R.T. (1980) *Cell*, 22, 361–370.
Shaw,G. and Kamen,R. (1986) *Cell*, 46, 659–667.
Shilo,B.Z. and Weinberg,R.A. (1981) *Proc. Natl. Acad. Sci. USA*, 78, 6789–6792.
Shtivelman,E., Lifshitz,B., Gale,R.P., Roe,B.A. and Canaani,E. (1986) *Cell*, 47, 277–284.

Suggs,S.V., Hirose,T., Miyake,T., Kawashima,E.H., Johnson,M.J., Itakura,K. and Wallace,R.B. (1981) *ICN-UCLA Symp. Mol. Cell Biol.*, 23, 683–693.
Telford,J., Burckhardt,J., Butler,B. and Pirrotta,V. (1985) *EMBO J.*, 4, 2609–2615.
Tosi,M., Young,R.A., Hagenbüchle,O. and Schibler,U. (1981) *Nucleic Acids Res.*, 9, 2313–2323.
Wang,J.Y.J. and Baltimore,D. (1983) *Mol. Cell. Biol.*, 6, 773–779.
Wang,J.Y.J., Ledley,F., Goff,S., Lee,R., Groner,Y. and Baltimore,D. (1984) *Cell*, 36, 349–356.
Wickens,M. and Stephenson,P. (1984) *Science*, 226, 1045–1051.
Yanisch-Perron,C., Vieira,J. and Messing,J. (1985) *Gene*, 33, 103–119.
Zehner,Z.E. and Patterson,B.M. (1983) *Proc. Natl. Acad. Sci. USA*, 80, 911–915.

Received on August 4, 1987; revised on October 1, 1987

Chapter III

On the possible role of the *c-abl* 3'UTR in determining messenger RNA half life

Dies Meijer, Anneke Graus and Gerard Grosveld

MGC, Department of Cell Biology and Genetics,Erasmus University, Rotterdam, The Netherlands

ABSTRACT

Through the use of an alternative polyadenylation site, the testis specific c-abl (TSabl) transcript of 4kb lacks 1.2kb of 3'UTR sequences that are otherwise present in the 5.3 and 6.5kb c-abl mRNAs. The TSabl RNA is abundantly expressed in post-meiotic male germ cells where it is very stable ($t_{1/2} \gg$ 6hrs). In contrast, the somatic 5.3 and 6.5kb mRNAs have a much shorter half life ($t_{1/2}$ 30-60 min.). This suggested that the 3'UTR of c-abl is involved in the regulation of c-abl expression through influencing mRNA half life. Here we test this hypothesis by ectopic expression of the TSabl RNA and derivatives thereof, in COS cells. The results of these experiments show that the ectopically expressed TSabl RNA has a similar short half life as the somatic c-abl mRNAs, thereby weakening the hypothesis that the 3'UTR of c-abl is involved in determining messenger half life. These results further suggest that the higher stability of TSabl mRNA in round spermatids is not a result of its 3' truncation, but rather that the mechanism(s) that regulate c-abl mRNA degradation in somatic cells are not operational in the haploid male germ cells.

RESULTS AND DISCUSSION

The mouse c-abl gene is the cellular homologue of the v-abl gene, the transforming oncogene of Abelson murine leukemia virus. The gene encodes a non-receptor type tyrosine specific kinase enzyme of 150 kD (reviewed in [1]). Two c-abl mRNAs of 5.3 (TypeI) and 6.5kb (TypeIV) in length are present in all tissues and cell lines examined [2, 3, 4]. The genomic organization of the c-abl gene and the relationship between the two mRNAs have been described [2, 5, 6, 7]. The two c-abl RNAs arise through alternative splicing of two different 5' exons (called typeI and typeIV in mouse and 1A and 1B in human) to a common set of ten 3' exons. Transcription of the two different 5' exons is initiated from separate promoters [6, 7]). The difference in size of the two c-abl RNAs is due to the difference in size of the first 5' exons. Since translation initiates within the first 5' exon, the two mRNAs encode proteins with different NH_2 termini.

In addition to the two c-abl mRNAs of 5.3 and 6.5kb, a shorter abundant transcript of 4kb is present in mouse testis [3] . This transcript was shown to be uniquely present in the post-meiotic germ cells within the testis [8]. cDNA cloning of this 4 kb testis specific c-abl (TSabl) transcript revealed that it is 1.2kb shorter at its 3' end as compared to the 5.3 and 6.5kb c-abl RNAs through the use of an alternative polyadenylation site [9, 10]. This alternative polyadenylation site is unusual in that it is not preceded by the

highly conserved hexanucleotide AAUAAA but instead contains a sequence remotely resembling this motif (UACAAA). Since this alternative polyadenylation site is well within the 3′ UTR the open reading frame (ORF) is not affected. Furthermore the TSabl transcript contains a Typel 5′ exon and is probably initiated from the same promoter as the 5.3kb mRNA. The high level of the TSabl mRNA in transcriptionally inactive elongating spermatids and residual bodies suggested that the transcript is very stable. Indeed, direct measurement of the stability of the TSabl RNA in round spermatids revealed a long half life (>6hrs). In contrast the half life of the 5.3 and 6.5kb c-abl mRNAs were shown to be in the order of 30 minutes to one hour in the pre B-cell line 70Z [9].

We hypothesized, that the higher stability of the TSabl mRNA is a direct result of the lack of 1.2kb of 3′UTR sequences. A longer half life of the TSabl mRNA could be linked to the progressive transcriptional inactivation of the spermatid nucleus during spermiogenesis. The stable TSabl mRNA could serve as a template for translation at a stage at which transcription has ceased. The 3′UTR of many eukaryotic mRNAs have been shown to be involved in regulation of expression [11-14]. The sequences important for

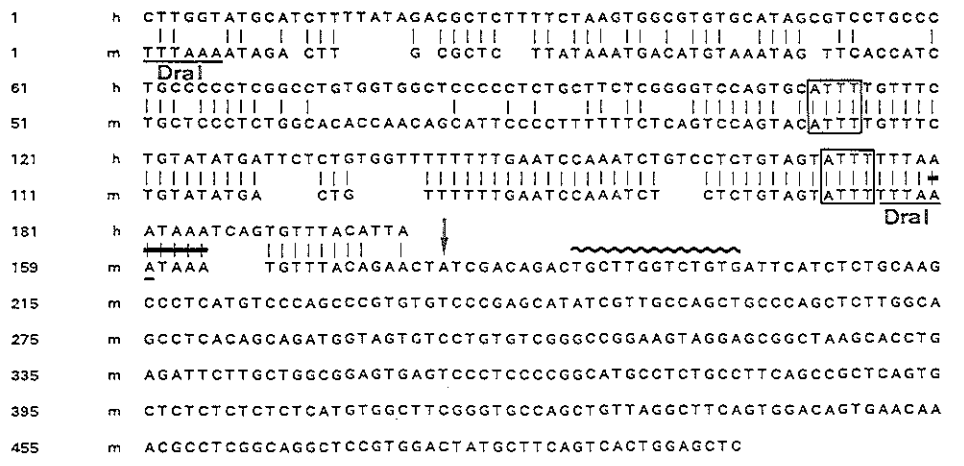


Figure 1. Partial sequence of the mouse c-abl 3′UTR plus flanking genomic fragment. The sequence is aligned with the last 200bp of the human c-abl 3′UTR (top strand) to show the regions of high homology in this region [15]. A SacI fragment of 1.1kb (fragment pSS09 in [2]) was subcloned from a large genomic clone in λEMBL3. This genomic phage which was isolated from a mouse Balb/C genomic library and contained the last large c-abl exon plus flanking sequences. The 5′ SacI site of the 1.1kb fragment maps within the 3′UTR of the last c-abl exon [2]. The fragment was sequenced on both strands after subcloning in the M13 phages mp18 and mp19. Only the last 525bp of this fragment are shown here. The DraI sites at position 3 and 156 were used to insert ClaI linkers. The conserved sequences resembling the destabilizer motif AUUUA are boxed. The poly(A) signal is underlined and the poly(A) addition site in the mouse sequence is indicated with an arrow. The GT-rich sequences 3′ of the polyadenylation/cleavage site are overlined.

regulation within these RNAs are likely to be phylogenetically conserved. A comparison between the human c-abl 3'UTR and the mouse c-abl 3'UTR revealed one region of high homology [15]. This region which is located at the extreme 3' end of the c-abl mRNAs includes two conserved elements resembling the "destabilizing element" AUUUA found in the 3'UTR of lymphokine and cytokine mRNAs and some proto-oncogenes like c-fos and c-myc (see Figure 1; [11, 16]). In general these elements are found in the 3'UTR of genes that can be rapidly and transiently induced by an external signal. Although the c-abl gene does not belong to this class of genes, it is possible that the AU rich sequences within the c-abl 3'UTR do play a role in determining c-abl mRNA stability.

In order to address the role of these sequences in mRNA stability we constructed four different c-abl expression plasmids (see Figure 2). The

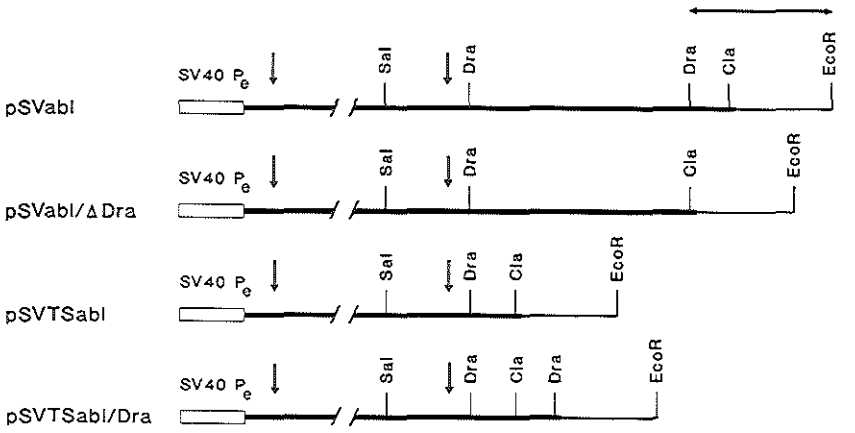


Figure 2. Schematic representation of the different c-abl expression plasmids. Transcription of c-abl mutants is initiated from the SV40 early promoter. The SV40 sequences are represented here by a box. Apart from the SV40 early promoter/enhancer this fragment also contains the SV40 origin of replication and a small intron [24]. The plasmid background is pTZ19. The thick line represents the c-abl cDNA sequence and the thin line the genomic sequence 3' of the poly(A) addition site (see figure 1). The position of the start and termination codons of the c-abl ORF is indicated by vertical arrows. The ClaI restriction sites were introduced by ligating ClaI linkers into the DraI sites of the 1.1kb genomic SacI fragment. Using the unique NcoI and SalI sites, full length clones were constructed (pSVabl and pSVabl/ΔDra). The ClaI restriction site at the testis specific poly(A) addition site was introduced by PCR. Two oligonucleotides were synthesized: CAGTGAGCAATCGATCTCGGGGAC and GACTGAGAAAGCCTGTGCC. The first (anti sense) oligonucleotide maps to the testis specific poly(A) addition site while the second (sense) oligonucleotide maps upstream of the unique SalI site. The typeI c-abl cDNA was used as a template and the resulting amplification product was cut with ClaI and SalI and subcloned to create pSVTSabl and pSVTSabl/Dra. All cloning manipulations were done as described [25]. The double headed arrow indicates the fragment of which the sequence is shown in figure 1.

proper 3' processing (cleavage and polyadenylation) of primary transcripts requires certain sequences 5' as well as 3' of the poly(A) addition site (reviewed by [17, 18]). Since cDNAs are derived from processed transcripts they do not contain the necessary downstream sequences. In order to be able to express the c-abl cDNA constructs in eukaryotic cells, the 3'end of the cDNAs were replaced by genomic sequences extending 340bp beyond the poly(A) addition site. The nucleotide sequence of the genomic fragment is shown in Figure 1. Downstream of the poly(A) addition site (indicated by an arrow) there is a GU rich region which was recognized as a second important cis element in 3'end processing [19]. Further experimentation showed, that this fragment supports the efficient 3'end processing of the c-abl transcripts. ClaI restriction sites were introduced by ligating ClaI linkers into the DraI site at position 3 or/and at position 178, directly flanking the poly(A) signal (see Figure 1). A ClaI restriction site was further created by PCR at the testis specific poly(A) addition site of the TSabl cDNA. Using the unique ClaI restriction sites, four different clones were constructed. These clones were placed under control of the SV40 early promoter. A small SV40 derived intron is present in the expression plasmids. The resulting constructs are schematically depicted in Figure 2. Clone pSVabl contains the entire typeI c-abl cDNA and contains a ClaI linker in the DraI site preceding the poly(A) signal sequence. The second construct pSVabl/ΔDra is derived from pSVabl but lacks the 175bp DraI fragment containing the conserved sequences (see Figure 1). The pSVTSabl construct represent the TSabl cDNA and the pSVTSabl/Dra construct in addition contains the 175bp DraI fragment. All four constructs were transiently expressed in COS cells and assayed for the stability of the c-abl RNAs. These experiments could not be done in isolated spermatids as a suitable transfection protocol for these cells does not exist as yet. The results of these experiments are presented in Figure 3. All four constructs expressed c-abl RNAs of the expected size indicating that the 3' genomic sequences are sufficient for proper 3'end processing of the primary transcripts. When transcription was blocked by ActinomycinD, all c-abl transcripts decay at a similar rate, while the endogenous GAPDH mRNA is stable. Therefore, removal of the 175bp DraI fragment does not result in a more stable RNA (compare panel A with panel D). Furthermore, the TSabl RNA is not more stable than the TypeI c-abl RNA in COS cells (panel B), and inclusion of the 175bp DraI fragment in the pSVTSabl does not result in a drastic destabilization of the transcript either. (panel C). As we did not quantitate the intensity of the hybridizing signal, we cannot rule out that the DraI fragment or the sequences from the testis specific polyA addition site to the DraI site at position 3 (Figure 1) have a subtle effect on messenger stability. However the qualitative picture is clear in that these sequences do not influence mRNA stability to the extent expected from the half life measurements of the TSabl RNA in round spermatids. These results argue against a role of these c-abl 3'UTR sequences in mRNA destabilization.

Instead these experiments suggest that the mechanism(s) that regulate c-abl degradation in COS cells are not operational in spermatids. However it should be noted that all four mutant c-abl RNAs share the last 20 nucleotides directly upstream of the polyA addition site, including the polyA signal.

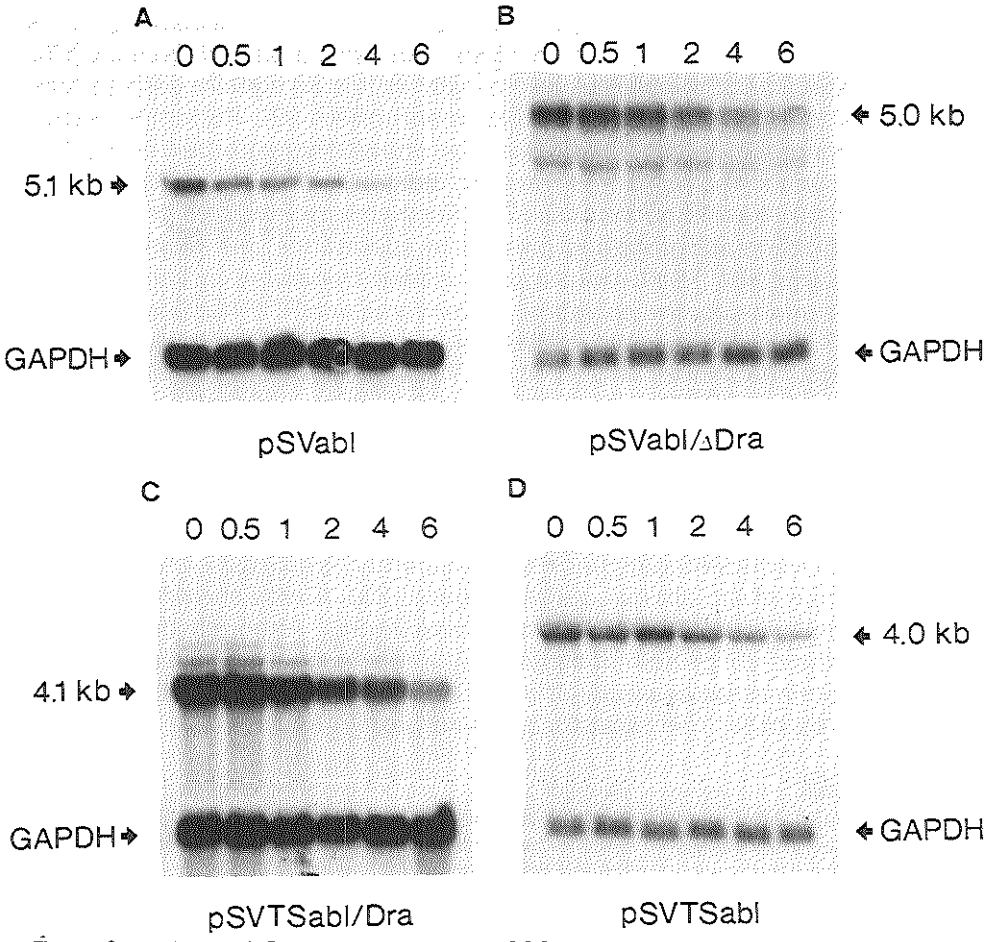


Figure 3. ActinomycinD chase experiments. COS cells were transiently transfected as described [26]. At day 3 of the transfection procedure the medium was changed for medium containing 10 μg/ml ActinomycinD. Cells were harvested at the indicated time (in hours) after addition of the actinomycinD containing medium and RNA was prepared using the Urea/LiCl method. Equal amounts of RNA was electrophoresed through denaturing 1% agarose gels and blotted onto nitrocellulose [27]. Blots were hybridized overnight at 42°C with a ^{32}P labelled mouse c-abl probe (a 2.2kb BglII cDNA fragment) and a human GAPDH probe in a hybridization buffer containing 50% formamide. Blots were washed to a final stringency of 0.1xSSC at 65°C and exposed for several hours. The human GAPDH probe detects the endogenous monkey GAPDH mRNA which is very stable [28]. The mouse c-abl probe detects the c-abl RNAs transcribed from the expression plasmids. Exposure times are too short to allow the detection of the endogenous monkey c-abl mRNAs.

Although it is unlikely that the polyA signal itself is involved in destabilizing the RNA we cannot rule out such a role for the 14 nucleotides separating the poly(A) signal from the poly(A) addition site. Destabilization of mRNAs have been shown to be linked to ongoing protein synthesis in a number of cases (for a review see [20]). If such a mechanism also operates on the TSabl RNA it is conceivable that the TSabl RNA is only poorly translated if at all. However the shorter 3'UTR of the TSabl transcript does not seem to affect the translatability of this mRNA intrinsically as it is an equally good template for translation *in vitro* as is the 5.3kb c-abl mRNA [9]. Polysome analysis experiments presented by Zakeri *et al.* has shown that a portion of the TSabl mRNA is found in the polysome fraction, indicating that at least some of the TSabl mRNA is translated in testis [21]. Furthermore Ponzetto *et al.* showed that the 150kD c-abl protein is present in spermatids and residual bodies, suggesting that the TSabl transcript is indeed translated *in vivo* [22]. The same distribution over the polysome, monosome and RNP fraction was found for the 5.3 and 6.5kb c-abl mRNAs as for the TSabl mRNA [21]. This suggests that the shorter 3'UTR of the TSabl mRNA does not influence the distribution of the RNA over these fractions either. Therefore, it remains unclear what function the alternative polyadenylation plays in TSabl RNA maturation and what factors cause the high stability of the transcript in spermatids. It has been suggested that the high level expression of the c-abl gene during spermiogenesis indicates an important role for the abl protein in the terminally differentiating germ cells. Recently Tybulewicz *et al.* described the generation of a c-abl null-mutant mouse [23]. It was shown that a male mouse homozygous for this mutation was fertile. Although this does not mean that the c-abl protein does not play a role in spermiogenesis, it shows that the protein is not essential. It's function could be redundant and may be taken over by other non-receptor tyrosine protein kinases.

REFERENCES

1. **Witte, O.N.** 1986. Functions of the abl oncogene. *Cancer Surv* **5**: 183-97
2. **Wang, J.Y., F. Ledley, S. Goff, R. Lee, Y. Groner and D. Baltimore.** 1984. The mouse c-abl locus: molecular cloning and characterization. *Cell* **36**: 349-56
3. **Muller, R., D.J. Slamon, J.M. Tremblay, M.J. Cline and I.M. Verma.** 1982. Differential expression of cellular oncogenes during pre- and postnatal development of the mouse. *Nature* **299**: 640-644
4. **Renshaw, M.W., M.A. Capozza and J.Y. Wang.** 1988. Differential expression of type-specific c-abl mRNAs in mouse tissues and cell lines. *Mol Cell Biol* **8**: 4547-51
5. **Ben-Neriah, Y., A. Bernards, M. Paskind, G.Q. Daley and D. Baltimore.** 1986. Alternative 5' exons in c-abl mRNA. *Cell* **44**: 577-86
6. **Bernards, A., M. Paskind and D. Baltimore.** 1988. Four murine c-abl mRNAs arise by usage of two transcriptional promoters and alternative splicing. *Oncogene* **2**: 297-304
7. **Shtivelman, E., B. Lifshitz, R.P. Gale, B.A. Roe and E. Canaani.** 1986. Alternative splicing of RNAs transcribed from the human abl gene and from the bcr-abl fused gene. *Cell* **47**: 277-84
8. **Ponzetto, C. and D.J. Wolgemuth.** 1985. Haploid expression of a unique c-abl transcript in the mouse male germ line. *Mol Cell Biol* **5**: 1791-4

9. **Meijer, D., A. Hermans, M. von Lindern, T. van Agthoven, A. de Klein, P. Mackenbach, A. Grootegoed, D. Talarico, G. Della Valle and G. Grosveld.** 1987. Molecular characterization of the testis specific c-abl mRNA in mouse. *Embo J* **6**: 4041-8
10. **Oppi, C., S.K. Shore and E.P. Reddy.** 1987. Nucleotide sequence of testis-derived c-abl cDNAs: implications for testis-specific transcription and abl oncogene activation. *Proc Natl Acad Sci U S A* **84**: 8200-4
11. **Shaw, G. and R. Kamen.** 1986. A conserved AU sequence from the 3' untranslated region of GM-CSF mRNA mediates selective mRNA degradation. *Cell* **46**: 659-67
12. **Owen, D. and L.C. Kuhn.** 1987. Noncoding 3' sequences of the transferrin receptor gene are required for mRNA regulation by iron. *Embo J* **6**: 1287-93
13. **Kruys, V., M. Wathelet, P. Poupart, R. Contreras, W. Fiers, J. Content and G. Huez.** 1987. The 3' untranslated region of the human interferon-beta mRNA has an inhibitory effect on translation. *Proc Natl Acad Sci U S A* **84**: 6030-4
14. **Braun, R.E., J.J. Peschon, R.R. Behringer, R.L. Brinster and R.D. Palmiter.** 1989. Protamine 3'-untranslated sequences regulate temporal translational control and subcellular localization of growth hormone in spermatids of transgenic mice. *Genes Dev* **3**: 793-802
15. **Andrews, D.3., C.K. Tompkins, S.L. Hendrickson and J.W. Singer.** 1990. Cloning and sequencing of the human c-abl 3' untranslated region. *Oncogene* **5**: 441-4
16. **Caput, D., B. Beutler, K. Hartog, R. Thayer, S.S. Brown and A. Cerami.** 1986. Identification of a common nucleotide sequence in the 3'-untranslated region of mRNA molecules specifying inflammatory mediators. *Proc Natl Acad Sci U S A* **83**: 1670-4
17. **Proudfoot, N.J. and E. Whitelaw.** 1988. Termination and 3' end processing of eukaryotic RNA. in "Transcription and Splicing", B. D. Hames and D. M. Glover IRL Press, Oxford
18. **Manley, J.L.** 1988. Polyadenylation of mRNA precursors. *Biochim Biophys Acta* **950**: 1-12
19. **McLauchlan, J., D. Gaffney, J.L. Whitton and J.B. Clements.** 1985. The consensus sequence YGTGTTYY located downstream from the AATAAA signal is required for efficient formation of mRNA 3' termini. *Nucleic Acids Res* **13**: 1347-68
20. **Cleveland, D.W. and T.J. Yen.** 1989. Multiple determinants of eukaryotic mRNA stability. *New Biol.* **1**: 121-126
21. **Zakeri, Z.F. and D.J. Wolgemuth.** 1987. Developmental-stage-specific expression of the hsp70 gene family during differentiation of the mammalian male germ line. *Mol Cell Biol* **7**: 1791-6
22. **Ponzetto, C., A.G. Wadewitz, A.M. Pendergast, O.N. Witte and D.J. Wolgemuth.** 1989. P150c-abl is detected in mouse male germ cells by an *in vitro* kinase assay and is associated with stage-specific phosphoproteins in haploid cells. *Oncogene* **4**: 685-90
23. **Tybulewicz, V.L., C.E. Crawford, P.K. Jackson, R.T. Bronson and R.C. Mulligan.** 1991. Neonatal lethality and lymphopenia in mice with a homozygous disruption of the c-abl proto-oncogene. *Cell* **65**: 1153-63
24. **Okayama, H. and P. Berg.** 1983. A cDNA cloning vector that permits expression of cDNA inserts in mammalian cells. *Mol Cell Biol* **3**: 280-9
25. **Sambrook, J., E. Fritsch and T. Maniatis.** 1989. "Molecular Cloning, a laboratory manual". Cold Spring Harbor Laboratory, New York
26. **Meijer, D., A. Graus, R. Kraay, A. Langeveld, M.P. Mulder and G. Grosveld.** 1990. The octamer binding factor Oct6: cDNA cloning and expression in early embryonic cells. *Nucleic Acids Res* **18**: 7357-65
27. **Fourney, R.M., J. Miyakoshi, R.S. Dayill and M.C. Paterson.** 1988. Northern Blotting: Efficient RNA staining and transfer. *Focus* **10**: 5-7
28. **Dani, C., J.M. Blanchard, M. Piechaczyk, S.S. El, L. Marty and P. Jeanteur.** 1984. Extreme instability of myc mRNA in normal and transformed human cells. *Proc Natl Acad Sci U S A* **81**: 7046-50

Chapter IV

Analysis of the mouse Type I *c-abl* promoter.

Dies Meijer and Gerard Grosveld

MGC, Department of Cell Biology and Genetics,Erasmus University, Rotterdam, The Netherlands

ABSTRACT

The 4.2kb testis specific c-abl (TSabl) transcript is transcribed from the same promoter region as the type I 5,3kb c-abl mRNA that is expressed in every cell type. The TSabl RNA is abundantly expressed in the post-meiotic male germ cells. The high level of TSabl in round and elongating spermatids could be due to the high stability of the transcript (see Chapters II & III) and/or to ongoing transcription of the gene during spermiogenesis up to the point of genome inactivation.S1 nuclease protection analysis of the c-abl RNAs expressed in testis revealed the presence of two major CAP sites within the type I promoter region (Chapter II). These two transcriptional initiation sites were distinct from the major CAP sites observed with pre-B cell RNA. This suggested the possible involvement of tissue specific factors in forming the initiation complex. In this chapter we describe an analysis of the c-abl Type I proximal promoter region using DNAseI footprinting and gel retardation assays. We show that there are no differences in footprinting patterns in the Type I promoter region up to position -250 (position 1 is here taken as the first nucleotide of the AUG start codon), using crude nuclear extracts from testis, spermatids, liver and pre-B cells. However in a gel retardation assay we observed a testis specific complex with a probe containing the octamer binding site (position -260). This testis specific octamer complex is shown to result from proteolytic degradation of the ubiquitous Oct1 protein.

INTRODUCTION

The mouse c-abl gene is the cellular homologue of the viral oncogene present in the genome of the acutely transforming Abelson murine leukemia virus (reviewed in [1]). The c-abl gene gives rise to two major transcripts of 5.3 and 6.5kb (termed TypeI and TypeIV respectively) [2, 3]. These two mRNAs originate from the alternative splicing of two different 5′ exons onto a set of 10 common exons [2]. Transcription of these 5′ exons is initiated from separate promoters [4, 5]. The two transcripts encode c-abl proteins which differ only in their amino termini. Both proteins have a tyrosine specific protein kinase activity and are thought to play an important role in an as yet unidentified signal transduction pathway. The gene is transcribed in all cell types examined and throughout embryogenesis[3, 6, 7].

Apart from the 5.3 and 6.5kb c-abl mRNAs a testis specific c-abl transcript of 4.2kb is present in the post-meiotic male germ cells [8]. The cloning of a cDNA corresponding to this mRNA revealed that it is 1.2kb shorter at its 3′ end as compared to the 5.3 and 6.5kb c-abl mRNAs through the use of an alternative polyadenylation site [9, 10]. The alternative polyadenylation site is located within the 3′ UTR and thus does not affect the ORF of the open reading frame of the mRNA. The TSabl mRNA contains

a TypeI 5' exon and is probably transcribed from the same promoter region as the 5.3kb mRNA.

The high level of this testis specific c-abl transcript could be due to different, not mutually exclusive mechanisms; a) the TSabl mRNA is more stable and has a longer half-life than the normal c-abl mRNAs, and b) the truncated 4.2 kb c-abl RNA is the only transcript produced during the haploid expression of the c-abl gene and continues to be transcribed to a far later stage in the developing germ cells than the normal c-abl mRNAs. The stability of the TSabl mRNA has been discussed in the previous chapters. Here we concentrate on the TypeI promoter region, from which the TSabl mRNA is transcribed, and the nuclear factors that interact with sequence elements within the promoter. Nuclease SI protection analysis of testis RNA revealed the presence of two major transcription initiation sites (CAP sites) within the TypeI promoter region [9]. These transcription initiation sites are distinct from the major CAP site used in the pre-B cell line 70Z/3. This suggested the involvement of tissue specific factors in forming the initiation complex. Here we describe DNAseI footprint and gel retardation experiments aiming at the identification of DNA elements, and nuclear factors that interact with these elements, that might play a role in the haploid specific expression of this promoter.

RESULTS AND DISCUSSION

The typeI c-abl promoter region and its human homologue 1A, have been cloned and sequenced [4, 5]. Using nuclease SI protection assays the transcription initition sites in 70Z/3 cells (a pre-B cell line) and testis have been mapped [9]. The chromosomal organization of the c-abl gene is schematically depicted in Figure 1. The approximate position of the testis specific and B cell specific CAP sites are also indicated in Figure 1 with arrows. It should be noted that the CAP sites in 70Z/3 cells have been mapped at the single nucleotide level, while the CAP sites in testis are approximate. The two major transcription initiation sites in testis are distinct from the transcription initiation sites in 70Z/3 cells. The TypeI c-abl promoter is rather rich in the dinucleotide CG and lacks a TATA-box and CCAAT box sequence. Transcription from TATA-less promoters is mostly promiscuous with respect to the initiation site used. The results of the nuclease SI experiments suggested the involvement of tissue specific factors in forming/positioning the initiation complex in 70Z/3 cells versus testis.

In order to identify DNA binding proteins that might be involved in the testis specific regulation of the TypeI c-abl promoter, we performed DNAseI footprint experiments using DNA probes derived from this promoter and crude nuclear extracts from testis, spermatids, 70Z/3 cells and liver. The DNA probes we used for this analysis contain the transcription initiation sites in testis and 70Z/3 cells. DNA fragments were labelled at the AvaI or the PvuII site using $\gamma(^{32}P)$-ATP and T4 polynucleotide kinase. Four areas were

identified that were protected from limited DNAseI digestion by nuclear proteins from testis, liver and 70Z/3 cells. The results of these experiments are schematically shown in Figure 1. The most proximal protected area maps over the cluster of CAP sites that are used in 70Z/3 cells. This footprint partially overlaps with a three times repeated motif GGCGGGA. The repetition of these motifs creates two junction motifs; GGGAGGC that resemble the concensus binding site for the nuclear factor SpI (GGGCGG).A mutational analysis by Letovsky and Dynan[11] of the central cytosine residue within this recognition sequence showed that a C→A mutation resulted in only a three fold reduction of the binding affinity of SpI for this site. It is of interest to note here that this part of the TypeI promoter is highly conserved between the human and mouse gene. However the A residue in the two distal GGCGGGA repeat elements is a C residue in the human Ia promoter sequence. The junction of these two elements create a concensus SpI binding site (GGGCGGC). It is likely that the GGGAGGC motif represents a true binding site for SpI and that protection results from binding of this factor. The second and third protected regions most likely also result from binding of SpI. The protected region around position -140 contains a concensus high affinity SpI binding site (GGGGCGGGGC) and a concensus

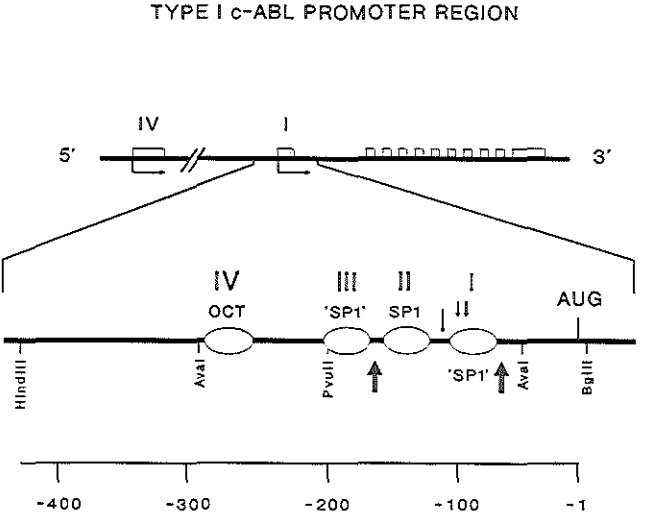


Figure 1. Schematic representation of the c-abl gene and its typeI promoter. The four regions that were protected from limited DNAse1 digestion by all four nuclear extracts used, are indicated by ovals. The DNAseI footprint analysis was performed with crude nuclear extracts from testis, spermatids, liver and 70Z/3 cells. The approximate position of the testis specific CAP sites are indicated by large arrows and the CAP sites in 70Z/3 cells by small arrows. CAP sites were mapped by S1 nuclease protection. The position of the protected fragments is indicated relative to the AUG start codon.

binding site for the factor GCF. The last factor is involved in the transcriptional repression of the EGF-R gene [12]. The third protected region around position -190 contains the lower affinity Spl binding site GGGGAGGGC. The major CAP sites in testis map 3' of the two lower affinity Spl binding sites present in protected region I and III.

The fourth protected region maps around position -275 and contains a perfect match with the concensus octamer binding site ATTTGCAT as present in the promoters of light and heavy chain immunoglobulin genes and the IgH enhancer [13, 14, 15]. This element is not conserved in the human 1A promoter sequence (but might be located in a more upstream position for which no sequence data are available yet). Interestingly, no abundant TSabl mRNA is observed in human testis. The octamer element was shown to be the major element that confers cell-type specificity on immunoglobuline gene promoters and enhancers [15, 16, 17]. This DNA motif was shown to be a binding site for the ubiquitous nuclear factor Oct1 (also called NFIII, OTF1, OBP100 etc) and for a large number of other nuclear proteins present in a more restricted set of cell types [18-23]. These tissue specific octamer binding factors establish tissue specific expression patterns through interaction with a common DNA element [14, 24]. The B-cell specific expression of immunoglobulin genes can at least in part be explained by the B-cell specific octamer binding factors Oct2A and Oct2B [14, 16, 22, 25]. These proteins cannot be distinghuised on the basis of their footprints over the octamer sequence as they produce similar footprints. However in a gel retardation assay the different proteins binding to the octamer element can be easily distinghuised on the basis of the electrophoretic mobility of the protein/DNA complexes they form. Thus although we could not detect differences in footprints over the Typel promoter region when using testis, spermatid, liver or 70Z/3 cell nuclear extracts it is possible that these prints resulted from binding of different tissue specific factors with the same DNA element. In this respect it is of interest that a testis specific octamer binding factor has been described in sea urchin [20, 26]. It is possible that a similar factor exists in mouse. Such a factor could be involved in the spermatid specific regulation of the the Typel c-abl promoter. We therefore decided to perform gel retardation experiments using a DNA probe containing the octamer motif and nuclear extracts from testis, spermatids, liver and a B-cell line (BMG). The results of this experiment are shown in Figure 2. All four extracts contain the Oct1 protein. B cells contain in addition to Oct1 the B-cell specific octamer binding factors Oct2A and Oct2B [22], while testis nuclear extracts produce a complex with a mobility in between that of Oct2A and Oct2B. Furthermore this complex was also observed when using nuclear extracts of spermatids, indicating that the protein is present in the germ cells. This complex was tentavily named TSoct.

The Oct1 and Oct2 proteins are encoded by different genes that belong to a large gene family [27]. This gene family is characterized by the presence

of a large stretch of homology which was first noted when the protein sequence of the Oct1 and Oct2 proteins was compared with the sequence of the rat pituitary specific transcription factor Pit1/GHF1 and the C. elegans gene Unc86 ([27] and references therein). This region of homology, the POU domain, was subsequently shown to constitute the DNA binding domain of the different proteins [28, 29, 30]. The POU domain consists of approximately 160 aminoacids and can be subdivided into two conserved regions, The aminoterminal POU-specific (POU_S) region of aproximately 75 residues is connected by a short diverged linker sequence to a second conserved domain of aproximately 65 aminoacids that shows extensive homology to the Drosophila homeobox proteins (POU homeodomain; POU_{HD}). As the testis specific octamer binding factor binds to the same sequence as the Oct1 and Oct2 proteins it is possible that the protein is encoded by one of these two genes or by another member of the POU domain gene family. (It is of interest to note here that we have found that the mouse Oct2 gene is highly expressed in testis, making this gene the prime candidate to encode the TSoct protein). The expression of the Oct2 gene in testis and the cloning of testis specific Oct2 cDNAs is described in the next chapter.

To further characterize the TSoct protein and to establish its relation with

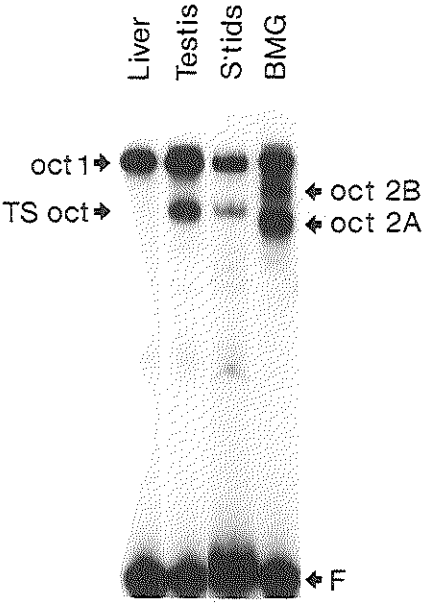


Figure 2. Bandshift experiment showing that distinct octamer binding factors are expressed in different cell types or tissues. A radiolabelled 30bp double stranded oligonucleotide containing the octamer motif was incubated with 5 μgram of crude nuclear extract from liver, testis, spermatids and a B-cell line (BMG). Protein/DNA complexes were resolved on a native 4% polyacrylamide gel.

the Oct1 and Oct2 proteins we performed protease clipping bandshift assays (PCBA; [22]). Nuclear extracts from testis, a B cell line (ROS) and a cell line that only expresses Oct1 (Epi7) were incubated with the octamer probe for 15 minutes at roomtemperature with different amounts of the endoprotease Arg-C. This protease hydrolysis peptide bonds at the carboxy end of arginine residues. Proteolytic degradation products that are still capable of forming a complex (ie fragments that retain the DNA binding domain) with the octamer probe are visualized after seperation on a native polacrylamide gel. In this way a protease specific degradation pattern of the octamer binding factor is produced. The results of these experiments are shown in Figure 3. Comparison of the degradation patterns of the testis complexes with the degradation pattern of the B-cell complexes reveal major differences. While the degradation pattern of the testis complexes is quite similar to the Oct1 degradation pattern (compare panel B and C in Figure 3). This indicates that the protein in the TSoct complex is more related to the Oct1 protein than to the Oct2 protein. To test whether the TSoct protein is also related antigenically to Oct1, we used antibodies directed against the Oct1 protein [31] in combination with a bandshift assay. As can be seen in Figure 4, preincubation of nuclear extracts with increasing amounts of Oct1 antibodies resulted in a progressive decrease in intensity of the Oct1 complex as well as the TSoct complex. The intensity of the Oct2 complex was only slightly decreased when B-cell nuclear extracts were preincubated with Oct1

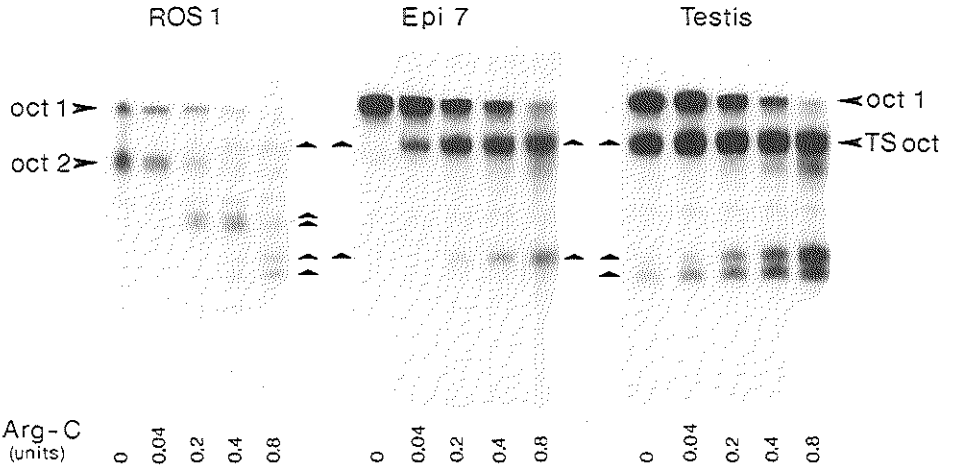


Figure 3. Protease clipping bandshift assay. Radiolabelled probe was incubated at roomtemperature with nuclear extracts of a B-cell line (ROS-1; left panel), Epi7 cells (middle panel) and testis (right panel). After ten minutes the indicated amount of the endoprotease Arg-C was added. Incubation at roomtemperature was continued for another ten minutes, before loading the samples on gel. The position of the degradation products is indicated with a triangle. Free probe is not shown.

antibodies (panel B in Figure 4). These results strongly suggest that the protein in the TSoct complex is related to Oct1 and not Oct2. The TSoct protein might therefore be encoded by a testis specific Oct1 mRNA, a related POU gene or the protein is a degradation product of Oct1. To test the latter possibility, the octamer probe was incubated with testis nuclear extracts for longer periods of time at roomtemperature, on ice and at roomtemperature in the presence of a wide spectrum cocktail of protease inhibitors. As is evident from Figure 5 the TSoct complex results from proteolytic degradation of the Oct1 protein. Degradation of Oct1 was never observed in nuclear extracts of brain, kidney, spleen, liver and a large number of cell lines. Despite the extensive precautions to prevent degradation of nuclear proteins during their isolation we have to conclude that a specific protease activity copurifies with the testis derived nuclear proteins. It is possible that this proteolytic activity resides in the acrosome or acrosomal membrane that is thightly associated with the spermatid nucleus. Thus, despite the fact that the Oct2 gene is highly expressed in testis, we must conclude that in mouse testis Oct1 is the only octamer binding factor detected and that binding of

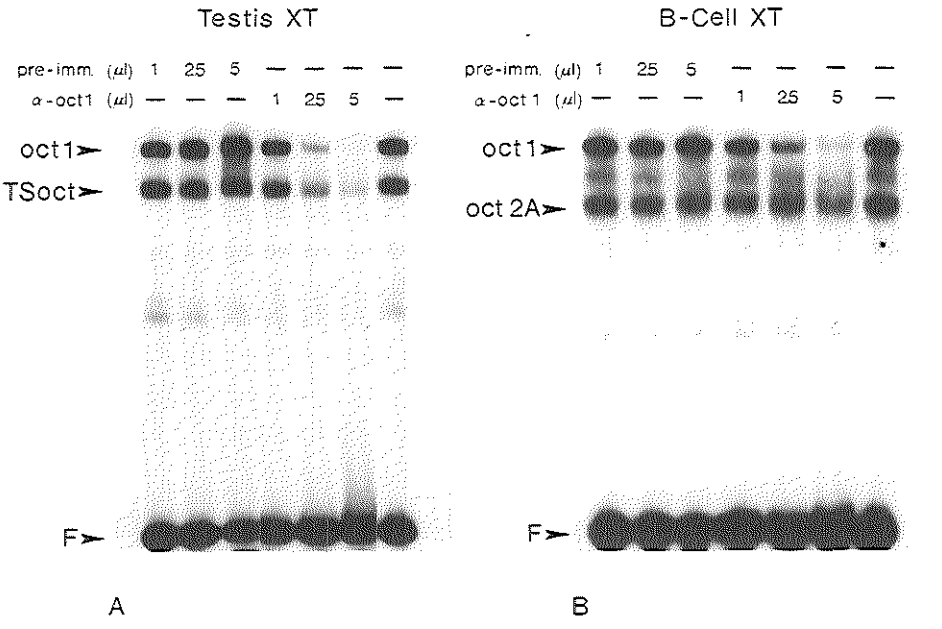


Figure 4. The TSoct complex contains Oct1 immunoreactive material. Testis nuclear extracts were incubated with the radiolabelled octamer probe in the presence of increasing amounts of a non-specific serum or Oct1 anti-serum as indicated above the lanes (panel A). Panel B shows a similar experiment with B-cell nuclear extracts. Note that the presence of Oct1 antibodies does not result in a supershift of Oct1 containing complexes but rather in a decrease in the number (intensity) of these complexes.

this protein results in the observed footprint IV.

CONCLUDING REMARKS

The proximal part of the typeI c-abl promoter contains at least four DNA elements that interact with nuclear factors present in testis, spermatids, 70Z/3 cells and liver. Three of these DNA elements (footprint I to III) probably interact with the transcription factor Sp1, while the fourth DNA element interacts with the transcription factor Oct1. These transcription factors are probably involved in the activation of transcription from the type I c-abl promoter region. As we did not positively identify Sp1 as the transcription factor interacting with the DNA elements I to III it is possible that other factors, like GCF [12], that have a high affinity for DNA elements related to the Sp1 binding site are responsible for the footprints I to III.

Two other DNA motifs that are conserved between the mouse typeI and human 1A promoter were not protected from limited DNAseI digestion (GGGGTTAAGG around position -160 and TGGGCCCTTTGTTA around position -120). These elements, which are located at either site of the concensus Sp1 site, do not match any of the DNA motifs known to be involved in transcriptional regulation. However, the high conservation of

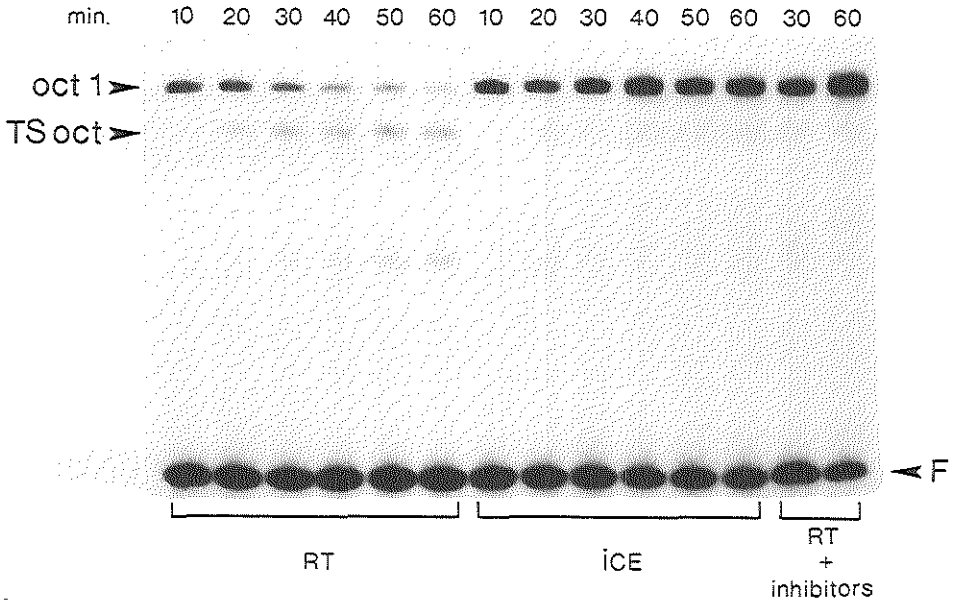


Figure 5. The TSoct complex results from proteolytic degradation of the Oct1 complex. Equal amounts of freshly prepared testis nuclear extracts were incubated with the radiolabelled octamer probe over an increasing time span either at roomtemperature or on ice (time indicated in minutes above the lanes). Two other samples were incubated at roomtemperature in the presence of a cocktail of protease inhibitors.

these elements in the otherwise diverged promoter region suggests that they may be important. It is possible that these DNA motifs do interact with nuclear proteins but that these interactions are not detected in our footprint assays as we used crude nuclear extracts.

It has been suggested by Zhu *et al.*[32] that the TTAA motif present in these conserved elements represents a functional TATA box. The TATA box was shown to be important in determining the exact site of transcription initiation in a number of cases. Deletion of the TATA element from the rabbit β-globin promoter was shown to affect the rate of transcriptional initiation *in vitro*. Furthermore initiation was shown to occur randomly over a stretch of DNA [33]. Given the observation that transcription initiation from the mouse typeI c-abl promoter occurs over a large stretch of DNA it is not likely that the TTAA element represents an efficient TATA box.

From the footprinting analysis presented here it is not clear if any of these DNA elements is involved in the activity of the typeI promoter in spermatids. It is clear that a functional analysis of the typeI c-abl promoter is needed to answer this question. Since a suitable transient transfection system for spermatogenic cells is missing, such an analysis must involve the generation of transgenic mice carrying different promoter deletion constructs.

ACKNOWLEDGEMENTS

We would like to thank Dr I. Mattaj for Oct1 antibodies, Anton Grootegoed and Petra Mackenbach for germ cell isolation, and Prof. Dirk Bootsma for continious support.

MATERIALS AND METHODS

Cell separation and nuclear extract preparation

Different populations of spermatogenic cells were isolated using a sedimentation method (Staput procedure) followed by density gradient centrifugation (Percoll gradients) as described by Grootegoed *et al.*[34]. This resulted in a highly purified population of round spermatids. Nuclear extracts from spermatids were prepared exactly as described [20]. Nuclear extracts from whole tissues was done as described by Lichtsteiner *et al.*[35] During the whole isolation procedure a mixture of protease inhibitors was included in the buffers (1mM PMSF, 5μg/ml leupeptin, 5μg/ml pepstatin, 5 mM Benzamidin, 5 μg/m' chymostatin). Protein concentrations of the nuclear extracts were determined with the direct photospectrometric method of Kalb and Bernlohr [36].

Bandshift and footprint assays

Electrophoretic mobility shift assays were performed as described [37] using a double stranded oligonucleotide that was labelled with $\gamma[^{32}P]$-ATP and T4 polynucletide kinase (Boehringer). Typically, 2 fmoles of probe was combined with 5 μg of nuclear extract in 20 μl bandshift buffer (10 mM Hepes-KOH pH7.6, 60mM KCl, 1 mM EDTA, 1 mM DTT, 4% Ficoll, 100 μg/ml poly[dI-dC]). and incubated at roomtemperature for 30 minutes. Samples were electrophoresed through a 4% (29:1) polyacrylamide gel in 0.25xTBE. After fixation in 10% acetic acid/10% methanol, the gel was dried and exposed to an autoradiographic film. DNAseI protection assays

were performed as described by Barberis *et al.*[20] Probes derived from the typel c-abl promoter were endlabelled on different restriction enzyme sites, either at the 5' end or 3' end, and purified from preparative low melting agarose gels. Probes were incubated on ice with 10 to 25 μg of crude nuclear extract in 25 μl of footprint buffer (20mM Hepes-KOH pH7.6, 60mM KCl, 1mM DTT, 0.8 mM $MgCl_2$, 0.2 mM PMSF, 8% glycerol,1 μg poly[dI-dC]). Then 1 to 10 μg of DNAsel was added and incubated for 90 seconds on ice. The reaction was terminated by adding 80 μl of stopmix (20mM Tris-HCl pH7.5, 10 mM EDTA, 0.5% SDS). The DNA was extracted by phenol/chloroform, ethanol precipetated, washed once with 70% ethanol and dissolved in sequence gel loading buffer. DNA samples were electrophoresed through a 8% polyacrylamide/7.5M Urea sequence gel together with a partially chemically degraded G+A sequence ladder of the probe as marker. After fixation and drying, the gel was exposed to autoradiographic film at -70°C using an intensifying screen.

REFERENCES

1. **Witte, O.N.** 1986. Functions of the abl oncogene. *Cancer Surv* **5:** 183-97
2. **Ben-Neriah, Y., A. Bernards, M. Paskind, G.Q. Daley and D. Baltimore.** 1986. Alternative 5' exons in c-abl mRNA. *Cell* **44:** 577-86
3. **Wang, J.Y., F. Ledley, S. Goff, R. Lee, Y. Groner and D. Baltimore.** 1984. The mouse c-abl locus: molecular cloning and characterization. *Cell* **36:** 349-56
4. **Bernards, A., M. Paskind and D. Baltimore.** 1988. Four murine c-abl mRNAs arise by usage of two transcriptional promoters and alternative splicing. *Oncogene* **2:** 297-304
5. **Shtivelman, E., B. Lifshitz, R.P. Gale, B.A. Roe and E. Canaani.** 1986. Alternative splicing of RNAs transcribed from the human abl gene and from the bcr-abl fused gene. *Cell* **47:** 277-84
6. **Muller, R., D.J. Slamon, J.M. Tremblay, M.J. Cline and I.M. Verma.** 1982. Differential expression of cellular oncogenes during pre- and postnatal development of the mouse. *Nature* **299:** 640-644
7. **Renshaw, M.W., M.A. Capozza and J.Y. Wang.** 1988. Differential expression of type-specific c-abl mRNAs in mouse tissues and cell lines. *Mol Cell Biol* **8:** 4547-51
8. **Ponzetto, C. and D.J. Wolgemuth.** 1985. Haploid expression of a unique c-abl transcript in the mouse male germ line. *Mol Cell Biol* **5:** 1791-4
9. **Meijer, D., A. Hermans, M. von Lindern, T. van Agthoven, A. de Klein, P. Mackenbach, A. Grootegoed, D. Talarico, G. Della Valle and G. Grosveld.** 1987. Molecular characterization of the testis specific c-abl mRNA in mouse. *Embo J* **6:** 4041-8
10. **Oppi, C., S.K. Shore and E.P. Reddy.** 1987. Nucleotide sequence of testis-derived c-abl cDNAs: implications for testis-specific transcription and abl oncogene activation. *Proc Natl Acad Sci U S A* **84:** 8200-4
11. **Letovsky, J. and W.S. Dynan.** 1989. Measurement of the binding of transcription factor Sp1 to a single GC box recognition sequence. *Nucleic Acids Res* **17:** 2639-53
12. **Kageyama, R. and I. Pastan.** 1989. Molecular cloning and characterization of a human DNA binding factor that represses transcription. *Cell* **59:** 815-25
13. **Singh, H., R. Sen, D. Baltimore and P.A. Sharp.** 1986. A nuclear factor that binds to a conserved sequence motif in transcriptional control elements of immunoglobulin genes. *Nature* **319:** 154-8
14. **Kemler, I. and W. Schaffner.** 1990. Octamer transcription factors and the cell type-specificity of immunoglobulin gene expression. *Faseb J* **4:** 1444-9
15. **Gerster, T., P. Matthias, M. Thali, J. Jiricny and W. Schaffner.** 1987. Cell type-specificity elements of the immunoglobulin heavy chain gene enhancer. *Embo J* **6:** 1323-30
16. **Müller, M.M., S. Ruppert, W. Schaffner and P. Matthias.** 1988. A cloned octamer transcription factor stimulates transcription from lymphoid-specific promoters in non-B cells. *Nature* **336:** 544-51
17. **Wirth, T., L. Staudt and D. Baltimore.** 1987. An octamer oligonucleotide upstream of a TATA motif is sufficient for lymphoid-specific promoter activity. *Nature* **329:** 174-8

18. **Landolfi, N.F., J.D. Capra and P.W. Tucker.** 1986. Interaction of cell-type-specific nuclear proteins with immunoglobulin VH promoter region sequences. *Nature* **323:** 548-51
19. **Cox, P.M., S.M. Temperley, H. Kumar and C.R. Goding.** 1988. A distinct octamer-binding protein present in malignant melanoma cells. *Nucleic Acids Res* **16:** 11047-56
20. **Barberis, A., F.G. Superti and M. Busslinger.** 1987. Mutually exclusive interaction of the CCAAT-binding factor and of a displacement protein with overlapping sequences of a histone gene promoter. *Cell* **50:** 347-59
21. **Schöler, H.R., A.K. Hatzopoulos, R. Balling, N. Suzuki and P. Gruss.** 1989. A family of octamer-specific proteins present during mouse embryogenesis: evidence for germline-specific expression of an Oct factor. *Embo J* **8:** 2543-50
22. **Schreiber, E., P. Matthias, M.M. Müller and W. Schaffner.** 1988. Identification of a novel lymphoid specific octamer binding protein (OTF-2B) by proteolytic clipping bandshift assay (PCBA). *Embo J* **7:** 4221-9
23. **Schreiber, E., K. Harshman, I. Kemler, U. Malipiero, W. Schaffner and A. Fontana.** 1990. Astrocytes and glioblastoma cells express novel octamer-DNA binding proteins distinct from the ubiquitous Oct-1 and B cell type Oct-2 proteins. *Nucleic Acids Res* **18:** 5495-503
24. **Schaffner, W.** 1989. How do different transcription factors binding the same DNA sequence sort out their jobs? *Trends Genet* **5:** 37-9
25. **Tanaka, M. and W. Herr.** 1990. Differential transcriptional activation by Oct-1 and Oct-2: interdependent activation domains induce Oct-2 phosphorylation. *Cell* **60:** 375-86
26. **Barberis, A., F.G. Superti and M. Busslinger.** 1989. The testis-specific octamer-binding protein of the sea urchin has a molecular weight of 85 kDa. *Nucleic Acids Res* **17:** 7114
27. **Herr, W., R.A. Sturm, R.G. Clerc, L.M. Corcoran, D. Baltimore, P.A. Sharp, H.A. Ingraham, M.G. Rosenfeld, M. Finney, G. Ruvkun and H.R. Horvitz.** 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:** 1513-6
28. **Ingraham, H.A., S.E. Flynn, J.W. Voss, V.R. Albert, M.S. Kapiloff, L. Wilson and M.G. Rosenfeld.** 1990. The POU-specific domain of Pit-1 is essential for sequence-specific, high affinity DNA binding and DNA-dependent Pit-1-Pit-1 interactions. *Cell* **61:** 1021-33
29. **Sturm, R.A. and W. Herr.** 1988. The POU domain is a bipartite DNA-binding structure. *Nature* **336:** 601-4
30. **Verrijzer, C.P., A.J. Kal and P.C. van der Vliet.** 1990. The oct-1 homeo domain contacts only part of the octamer sequence and full oct-1 DNA-binding activity requires the POU-specific domain. *Genes Dev* **4:** 1964-74
31. **Pruijn, J.M., d.V.P. van, N.A. Dathan and I.W. Mattaj.** 1989. Anti-OTF-1 antibodies inhibit NFIII stimulation of *in vitro* adenovirus DNA replication. *Nucleic Acids Res* **17:** 1845-63
32. **Zhu, Q.S., N. Heisterkamp and J. Groffen.** 1990. Characterization of the human ABL promoter regions. *Oncogene* **5:** 885-91
33. **Grosveld, G.C., E. de Boer, C.K. Shewmaker and R.A. Flavell.** 1982. DNA sequence necessary for transcrition of the rabbit β-globin gene *in vivo*. *Nature* **295:** 120-126
34. **Grootegoed, J.A., R. Jansen and d.M.H. van.** 1986. Effect of glucose on ATP dephosphorylation in rat spermatids. *J Reprod Fertil* **77:** 99-107
35. **Lichtsteiner, S., J. Wuarin and U. Schibler.** 1987. The interplay of DNA-binding proteins on the promoter of the mouse albumin gene. *Cell* **51:** 963-73
36. **Kalb Jr., V.F. and R.W. Bernlohr.** 1977. A new spectrophotometric assay for protein in cell extracts. *Anal. Biochem.* **82:** 362-371
37. **Meijer, D., A. Graus, R. Kraay, A. Langeveld, M.P. Mulder and G. Grosveld.** 1990. The octamer binding factor Oct6: cDNA cloning and expression in early embryonic cells. *Nucleic Acids Res* **18:** 7357-65

Chapter V

Oct2 gene expression and spermatogenesis

Dies Meijer, Anneke Graus, Petra Mackenbach*, Anton Grootegoed* and Gerard Grosveld

MGC, Department of Cell Biology and Genetics
*Department of Endocrinology and Reproduction.
Erasmus University, Rotterdam The Netherlands

ABSTRACT

Using Northern blot analysis we show that the mouse Oct2 gene is highly expressed in testis. Two major transcripts of 7.2 and 6.8 kb were observed in testis and one major Oct2 MRNA of 7kb in brain and spleen. Cell separation experiments indicate that the Oct2 gene is abundantly expressed in spermatocytes but not in spermatids. Oct2 cDNAs were isolated from a testis cDNA library and through PCR amplification of testis cDNA. Two cDNA constructs (1-tsOct2 and 2-tsOct2) were obtained that represent the ORF of the 7.2kb Oct2 mRNA. This mRNA results from alternative splicing of Oct2 exons and is possibly transcribed from a testis specific promoter. The tsOct2 cDNAs do not contain sequences encoded by the Oct2 exons 1 to 7a (1-tsOct2) or 1 to 7b (2-tsOct2). The Oct2 proteins potentially encoded by these cDNAs (44kD and 39kD respectively) are initiated from in frame methionine residues and are therefore truncated of the first 190 (1-tsOct2) or the first 249 (2-tsOct2) amino acids. This truncation of the 2-tsOct2 ORF includes most of the POU_S-A box, resulting in a severe reduction in DNA binding affinity of this protein. Alternative splicing of a small 74nt exon (exon13) results in a carboxyl terminus of the tsOct2 proteins that is identical to that of the B-cell Oct2B protein. Expression of the tsOct2 cDNAs and analysis of the produced proteins in a bandshift assay gave rise to complexes with a much higher mobility than the Oct1 and Oct2A complexes. However these tsOct2 generated complexes were not observed in testis nuclear extracts.

INTRODUCTION

Male germ cell formation in mammals, in particular rats and mice, is a well characterized developmental process. A self renewing population of spermatogonial stem cells exists within the seminiferous epithelium in the testis. These spermatogonia produce daughter cells that go through meiosis (spermatocytes) and give rise to four haploid round spermatids. The round spermatids then differentiate to elongated spermatids and finally to spermatozoa. During the post-meiotic period the germ cells undergo dramatic morphological changes as they acquire the flagellum, the acrosome and adopts its species specific head shape. The major part of the cytoplasm is shed as a residual body that is subsequently resorbed by Sertoli cells.

Underlying these processes is the temporally controlled expression of an as yet undefined number of genes. Some of these are uniquely expressed during spermatogenesis, while others express a testis specific form of a protein or transcript (for review see;[1-4]). The control of expression of these

genes can be at the transcriptional and/or post-transcriptional level. For example, the mouse protamine 1 gene (MP1) is expressed shortly after the second meiotic division but the mRNA is translated only days later during the later stages of spermiogenesis. Another example is the testis specific phosphoglycerate kinase 2 (PGK2) gene that is first transcribed during prophase of meiosis but translation of the mRNA occurs during spermiogenesis [5]. Both genes, MP1 and PGK2, are under transcriptional as well as translational control. In the case of MP1, sequences within the 3' untranslated region of the gene seem to be involved in the proper timing of translation [6]. Other genes such as the oncogenes int-1, c-mos and c-abl show temporally controlled expression during spermatogenesis as well (reviewed in [3]). It is very likely that specific transcription factors are involved in the temporally controlled expression of these genes. Identification and subsequent characterisation of these factors may give insight into the male germ cell differentiation pattern at the molecular level. These factors might be uniquely expressed in testis or might represent variations of factors that are also expressed in other cell types.

Many transcription factors are known to interact through common DNA elements. One such DNA element is the octamer sequence (ATTTGCAT), which is known to be a binding site for several transcription factors. This sequence was originally defined as a conserved element within the promoters and enhancer of the immunoglobulin genes [7, 8]. The octamer element was shown to confer B-cell specific expression to a non B cell specific promoter [9, 10]. This sequence was also found in the regulatory regions of genes that are not B cell specific, such as the histone H2b genes, the snRNA genes, and the mouse c-abl gene. Initially two proteins, called Oct1 and Oct2, have been described to interact with this element. Oct1 is a ubiquitous factor involved in cell cycle regulation of the histone H2b gene, the replication of adenovirus DNA, and expression of the snRNA genes [11-18]. The second factor Oct2 was originally described as B cell specific and as such assumed to be involved in the regulation of the immunoglobulin genes [19]. The two proteins have been purified and cDNAs encoding these proteins have been isolated and characterized [20-23]. Sequence comparison between these two cDNAs revealed a large stretch of homology which was also found in the rat pituitary specific factor Pit-1/GHF-1 and the C. elegans gene *unc86* ([24] and references therein). This region of homology (called the POU domain) was shown to be the DNA binding domain [25, 26, 27]. The POU domain is approximately 160 amino acids long and contains two subdomains. The amino terminal POU-specific (POU_S) region of approx. 75 amino acid residues is connected through a short linker sequence to the 65 amino acid POU-homeo domain (POU_{HD}). This last domain is related to the homeobox, a conserved region originally found in many Drosophila genes involved in development [28]. Over the past years, additional factors binding to the octamer motif have been identified [29-33]. Noteworthy with respect

to the work described here is the identification of a specific octamer factor in sea urchin testis [29]. It is possible that these distinct factors are involved in the tissue specific regulation of a large number of genes.

The presence of an octamer element within the mouse Type I c-abl promoter region [34], and our initial observation of a testis specific octamer binding factor (see previous chapter) suggested the involvement of this factor in the haploid specific expression of the c-abl gene. We therefore set out to clone cDNA molecules encoding octamer binding factors that are specifically expressed in the testis. Here we describe the cloning of cDNAs potentially encoding Oct2 proteins that are uniquely expressed during spermatogenesis. The possible involvement is discussed of these proteins in the processes described above.

RESULTS

Cloning of a CDNA encoding an octamer binding factor

The approach we took to isolate cDNAs from testis RNA potentially encoding octamer binding factors was based on the high homology within the POU homeo domain of the Oct1 and Oct2 cDNAs (Figure 1). Fully degenerated oligonucleotides were used to generate cDNA from testis RNA, that was subsequently amplified using Taq1 polymerase (PCR). cDNAs hybridizing with an internal degenerated oligonucleotide were subcloned in the plasmid

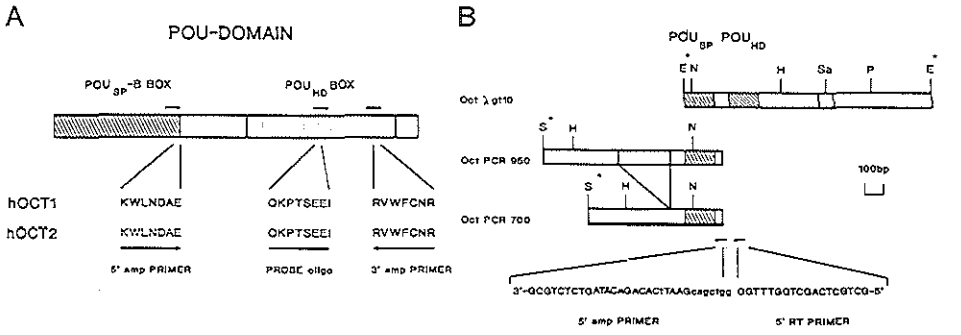


Figure. 1 cDNA PCR amplification of testis mRNAs encoding POU domain proteins. Fully degenerate oligonucleotides (see M&M) were synthesized based on amino acid sequences conserved between the human Oct1 and Oct2 proteins. Amplification products were separated on agarose gels and blotted onto nitrocellulose. The filters were probed with a POU_{HD} specific oligonucleotide. Positive bands were excised from the gel, cloned and sequenced.

B. Cloning of a testis specific Oct2 cDNA. A partial cDNA was isolated from a λgt10 testis cDNA library by use of a POU domain probe. Five prime extensions were cloned by exploiting the RACE protocol. The structural relationship between the partial Oct2 cDNA and the 5' extension clones is depicted. Restriction enzyme sites are: N = NcoI, H = HindIII, P = PstI, S = SalI, Sa = SacI, E = EcoRI. The SalI and EcoRI sites are artificial cloning sites. The 5' extension clones were joined at the NcoI site with the λgt10 Oct2 clone resulting in two clones: 1-tsOct2 (950 PCR) and 2-tsOct2 (750 PCR).

vector pTZ18. Sequencing of these clones showed them to be identical to the Oct2 POU domain. Using this cDNA as a probe, a testis cDNA library was screened. This yielded one positive clone (Octλgt10 in Figure 1B), and the cDNA insert was subcloned into pTZ19 and sequenced. Sequence analysis of this clone indicated that the cDNA was truncated at its 5' side. To obtain full length cDNA, 5' extensions were cloned using the RACE protocol ([35], see Materials and Methods) and distinct 5' cDNAs were obtained (Oct PCR950 and Oct PCR 700 in Figure 1B). Using the unique overlapping Nco1 site in the 5' extended cDNAs and the original octλgt10 cDNA two full length cDNAs were constructed; 1-tsOct2 representing the longer clone and 2-tsOct2 representing the shorter one. The relation between the two 5' PCR products is schematically depicted in Figure 1B.

The Oct2 gene is highly expressed in testis

Previously it has been reported that the Oct2 gene is exclusively expressed in lymphoid cells [20, 21, 22]. More recently, He *et al.* [36] also demonstrated expression of the Oct2 gene in rat brain, but not in rat testis. Therefore it is surprising that we isolated an Oct2 related cDNA from a mouse testis cDNA library. As we only isolated a single cDNA related to Oct2 from our testis cDNA library there is a possibility that the RNA, used to construct the library, contained some lymphoid RNA due to the presence of lymphoid cells in the testis.

We investigated the expression pattern of the Oct2 gene using Northern blot analysis of RNA from testis, spermatocytes, round spermatids and brain. As a probe we used the Oct2 cDNA from Octλgt10. In brain one major transcript of approximately 7kb was present while in testis two transcripts of approximately 7.2 and 6.8kb were detected (Figure 2A). These two RNAs

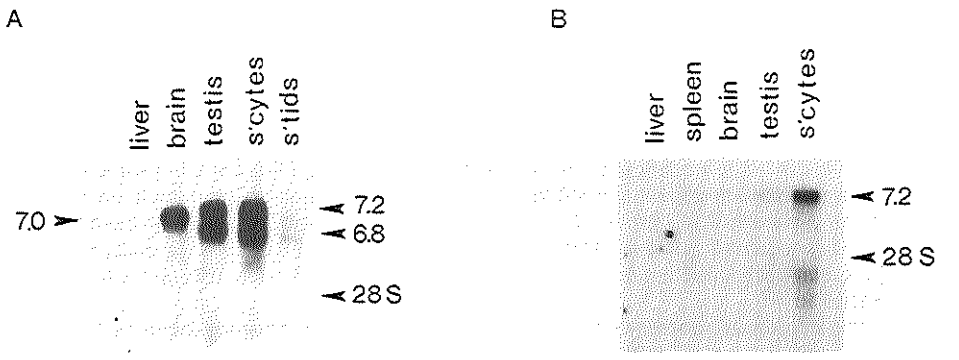


Figure 2. Northern blot analysis of Oct2 RNA from various sources as indicated. A. The probe consisted of a 490bp NcoI/HindIII Oct2 cDNA fragment (see figure 4) spanning most of the POU_S and all of the POU_{HD} domain. The position of 28S rRNA is indicated.
B. The probe used was a 430bp SacI-SacII fragment derived from the 5' end of the 1-tsoct2 cDNA. The amount of total RNA in each lane is 10 μg.

were also detected at a high level in RNA from spermatocytes, but at a much lower level in RNA from round spermatids. This clearly indicates that the Oct2 gene is actively transcribed during meiosis and is down-regulated after completion of the meiotic divisions. The probe used in the experiment described above only contains sequences that are also present in the cDNAs cloned from human and mouse B-cell libraries. However the 5' part of the tsOct2 cDNAs is distinct and fails to show any homology with B-cell derived Oct2 cDNAs (see below). To establish the relationship between the tsOct2 cDNAs and the two Oct2 RNAs detected in testis and spermatocytes, a probe derived from the unique 5' part of the cDNA (a 430bp SacI-SacII fragment) was used to screen a Northern blot containing RNA from spleen, brain, testis, spermatocytes and spermatids. This probe only detects the higher molecular weight transcript in testis and spermatocytes (Figure 2B). Therefore, the tsOct2 cDNAs represent the longer transcript observed in testis. As we selected two amplified 5' cDNA ends for further characterization, it is possible that the shorter of the two PCR products represents the other testis specific Oct2 RNA.

Sequence analysis

The nucleotide sequence of the 1-tsOct2 and 2-tsOct2 constructs was determined on both strands. The sequence is presented in Figure 3. Sequences that are not present in the 2-tsOct2 cDNA are underlined. Both cDNAs contain a long ORF encompassing the POU domain. In the 1-tsOct2 clone the reading frame starts at position 614 and terminates at position 1907 while in the 2-tsOct2 clone the reading frame starts at position 791 and terminates at position 1907. Thus the 2-tsOct2 ORF represents an amino terminal truncation of the 1-tsOct2 frame and lacks most of subdomain A of the POU specific box. The calculated molecular weights of the proteins potentially encoded by the 1-tsOct2 and 2-tsOct2 are 44kD and 39 kD, respectively. Both cDNAs contain a long 5'UTR that contains multiple short open reading frames. The nucleotide sequence of the Oct PCR750 clone that was used to construct 2-tsOct2 is identical to the Oct PCR950 sequence, except that nucleotides between nt432 and nt659 are missing (underlined sequence in Figure 3). This sequence has the hallmark of an intron. At the 5' side there is a GT dinucleotide and at the 3' side there is an AG dinucleotide preceded by a long stretch of pyrimidine residues. It is likely that these sequences represent an alternatively spliced intron within the PCR950 Oct2 cDNA. The sequence upstream of nt 432 that is shared by both PCR cDNAs, contain multiple splice donor/acceptor sites indicating that they may contain even more non-spliced sequences. The two independent PCR cDNA molecules start at the same position (nt1). This nucleotide could therefore represent the transcriptional initiation site (see below).

Downstream of nt 610 the cDNA sequence is homologous with the previously described Oct2 cDNAs cloned from B-cell libraries [20, 21, 22].

Only a few amino acid differences are found between the mouse and human sequence, none of them within the highly conserved POU domain. It should be noted that the methionine codon that opens the 1-tsOct2 reading frame is a valine codon in the human Oct2 sequence [20, 21].

Recently the genomic organization of the mouse Oct2 gene has been described [37, 38]. The gene consists of at least 14 exons spread over 35kb of genomic DNA. Six different B-cell specific Oct2 mRNAs have been characterized that arise through alternative splicing. Comparing the tsOct2 cDNA with the Oct2 exon sequences showed that the sequence up to nt610 is unique to the testis specific Oct2 cDNAs. Downstream of nt 611 the

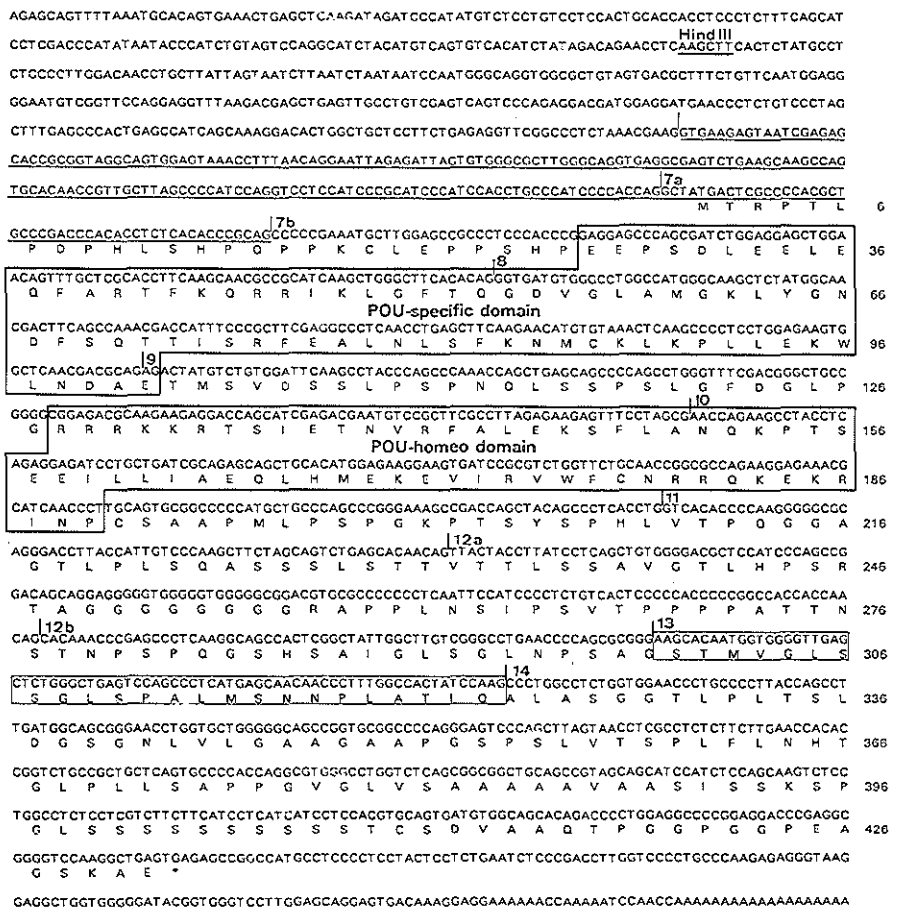


Figure 3. Nucleotide sequence and deduced amino acid sequence of the tsOct2 cDNA constructs. The sequence that is not present in the 2-tsOct2 cDNA is underlined. Exon assignment and numbering is according to Wirth *et al.* ([38]; see also Figure 7). The exon junctions are indicated by a vertical bar. The 74nt exon (exon13) that is subject to tissue specific alternative splicing is boxed (see also Figure 4).

tsOct2 sequence is composed of exons 7a to 14 (indicated in Figure 3, see also Figure 7). Interestingly, the tsOct2 cDNA contains exon 13. This 74nt exon was shown to be subject to alternative splicing in B cells. The splicing in of this exon results in a shift of reading frame, extending the ORF with 132 amino acids. B cell specific Oct2 mRNAs containing this exon encode the Oct2B protein, a second Oct factor in B cells [37, 38, 39]. Thus the putative tsOct2 proteins have the same carboxyl terminus as the Oct2B protein. This part of the protein is rich in serine residues, a property it shares with the human Oct1 protein [23]

Splicing of the 74nt 3'exon

To prove that the 74 nt sequence indeed represents a *bona fide* exon, genomic sequences were cloned from a phage library, mapped and sequenced. The result of this analysis is shown in Figure 4A. The exon is located on a 5kb HindIII genomic fragment and is flanked by consensus splice acceptor/donor sites. The two flanking exons are also located on this fragment. The 74nt exon is separated by a 122bp intron from its 5' neighbouring exon and by a 284bp intron from its downstream neighbour. This 74nt exon is conserved in human, since the cDNA sequence of one of the B-cell Oct2 cDNAs presented by Clerc *et al.* [20] contains this exon plus the 5' flanking intron and part of the 3' flanking intron. To demonstrate

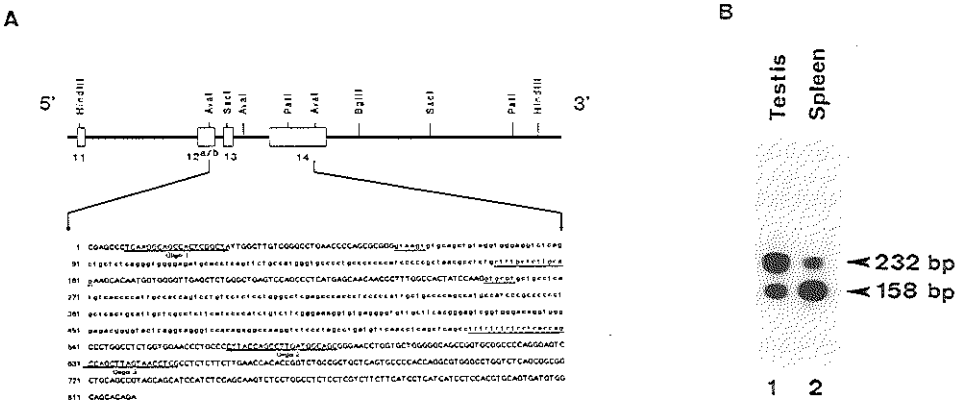


Figure 4. Tissue specific alternative splicing of a 3' exon, its genomic organization and sequence. Using a probe from the 3' part of the cDNA, a 5kb HindIII fragment was subcloned from a genomic clone in λ phage EMBL3. The fragment was mapped and partially sequenced (the two AvaI fragments were sequenced on both strands). Numbering of the exons is according to Wirth *et al.* [38]. Exons are not exactly drawn to scale. B. Sequence of the AvaI fragments. Exon sequences are represented by a capital lettertype, introns by a small lettertype. Splice donor and acceptor sequences are underlined. The oligonucleotides that were used to amplify Oct2 cDNAs are indicated (oligo1 and oligo3). Alternative splicing of exon13 will give rise to amplified cDNA products of 158bp and 232bp. These amplified products were detected with a third oligonucleotide(oligo2) on Southern blots of amplified material (C).

tissue specific alternative splicing of this exon, we employed the cDNA PCR technique. An Oct2 specific oligo was used to generate cDNA from mouse testis and spleen RNA. These cDNAs were subsequently amplified by PCR using the oligo's as indicated in Figure 4A. Amplified products were separated on a 2% agarose gel and blotted onto nitrocellulose. An internal oligo was used to detect the amplified products. As can be seen in Figure 4B the spliced-in product is the most abundant transcript in testis (232bp) while the spliced-out product is the most abundant transcript in spleen. In conclusion, the 74nt exon is subject to alternative splicing and this process seems to be regulated in a tissue specific fashion.

Is Oct2 gene expression in testis initiated from a tissue specific promoter? As the two 5'PCR Oct2 cDNA molecules start at the same nucleotide, it is possible that this nucleotide represents the transcription initiation site. Furthermore, a probe derived from the 5' part of these cDNAs only detects the largest of the two testis specific Oct2 mRNAs (Figure 2). Since the sequence upstream of nt 610 shows a perfect match with the splice acceptor consensus, and these sequences are not homologous to Oct2 exons 1 to 6, it is possible that transcription is initiated from a testis specific promoter located in intron 6. To map the position of the 5' part of the tsOct2 cDNAs, a genomic clone covering intron 6 was obtained (a kind gift of Dr T Wirth). Using a probe derived from the 5' part of the cDNA, it could be shown that these sequences mapped to intron 6 (not shown). Sequence analysis showed that the tsOct2 sequences are colinear with the genomic sequence upstream of exon 7a. A 440bp genomic HindIII fragment was subcloned and sequenced, covering the most 5' part of our PCR cDNAs (3' HindIII site at position 163 in Figure 3) and mapping 447bp upstream of exon

AAGCTTCTTATTTTTAATTTAAATTTATTTATATTTTGGTATTACTAGAG

AGTCAATCCAGAATTTTTAGTATACTAGGCAGGGGCTCTACCACTGAGCC

ACACCCCCAGCCCTCTATTCATATTTTATTTTGAGATGGTCTTTCTGAGT

TGCTCAGGCTGGCATTGAACTCACTCTGTAGTTCAAGCTGACCTTAAAAG

AATCCCTCTGCCTTGTTCTCTGGAGCAGCTATATGTTTGCTGTCAGGCCT

TAATTTTTTTTTTAGAGCAGTTTTAAATGCACAGTGAAACTGAGCTCAAG

ATAGATCCCATATGTCTCCTGTCCTCCACGTGCACCACCTCCCTCTTTCA

GCATCCTCGACCCATATAATACCCATCTGTAGTCCAGGCATCTACATGTC

AGTGTCACATCTATAGACAGAACCTCAAGCTT

Figure 5. Sequence of the 440bp HindIII fragment derived from intron 6. The sequence represented in the tsOct2 cDNAs is underlined. Overlined sequence motifs are discussed in the text.

7a. The sequence of this fragment is shown in Figure 5. The 3' part of the sequence is colinear with the tsOct2 cDNA sequences (underlined in Figure 5). Directly upstream of nt 1 of the tsOct2 cDNA sequence there is a stretch of 10 T residues. At position -30 (taking the first nucleotide of the tsOct2 cDNA as position 1) a TATA box sequence is present, and at position -95 there is an inverted CAAT box. These typical promoter elements might form the testis specific Oct2 promoter. However, an alternative explanation why the two PCR cDNA products start at the same nucleotide is provided by the presence of the 10 T residues upstream of nt 1. First strand cDNA molecules were tailed using TdT and dATP. Second strand synthesis was subsequently primed by a polylinker-oligodT oligonucleotide (see M&M). Obviously, this primer can anneal to internal A stretches (in the first strand cDNA) and generate truncated cDNAs. Thus, despite the presence of a TATA box and a CAAT box at typical distances, we cannot conclude that these sequences represent the actual promoter from which testis specific Oct2 gene expression is initiated.

The tsOct2 cDNAs encode functional octamer binding proteins

The 1-tsOct2 and the 2-tsOct2 cDNAs potentially encode proteins that differ in size considerably. However, in gel retardation experiments using testis nuclear extracts and the c-abl octamer element as a probe, only the

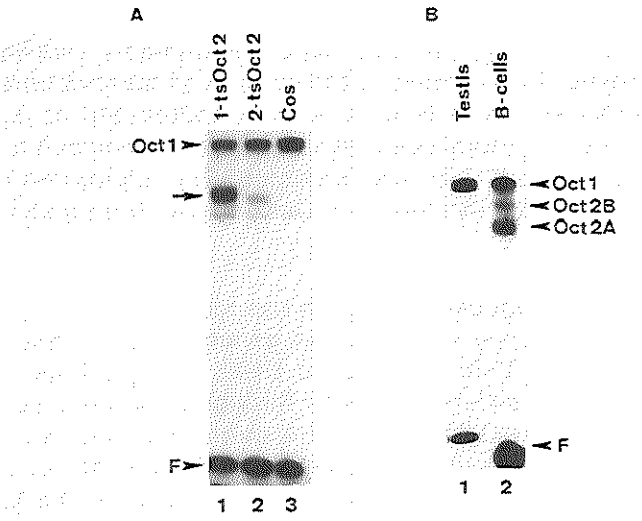


Figure 6. *In vivo* expression of tsOct2 cDNAs. The tsOct2 cDNA were expressed in COS cells under control of the SV40 early promoter/enhancer (see M&M). Mini nuclear extracts of transfected and mock transfected COS cells were used in a bandshift assay with a radiolabelled octamer probe (panelA). The arrow points at the tsOct2 encoded proteins and the free probe is indicated with an F. For comparison, Panel B shows the octamer complexes that are formed with testis and B-cell (ROS cell line) nuclear extracts.

ubiquitous Oct1 complex is observed (Figure 6B and the previous chapter). One explanation for the failure to detect testis specific Oct complexes could be that the tsOct2 mRNAs are poor templates for translation. Both 1-tsOct2 and 2-tsOct2 have long 5' untranslated regions containing multiple short open reading frames. It has been suggested that the presence of multiple potential start codons within the mRNA leader sequence severely reduces the efficiency of translation from a downstream start codon [40]. To test whether the tsOct2 cDNAs could encode functional octamer binding proteins, we performed bandshift assays using nuclear extracts of COS cells expressing the tsOct2 cDNAs from the SV40 promoter. The results of these experiments are shown in Figure 6. Clearly the tsOct2 RNAs are efficiently translated *in vivo*, giving rise to protein/DNA complexes that have a higher mobility than the Oct1 complex. These complexes are not observed with testis nuclear extracts (Figure 6B). Identical results were obtained with *in vitro* translated proteins. The weaker intensity of the 2-tsOct2 complex is probably not due to inefficient translation of the 2-tsOct2 mRNA, but to the fact that the protein misses most of subdomain A of the POU_S domain. Mutations and truncations of the POU_S domain of the Oct1 and Oct2 proteins impairs DNA binding to a different extent [26, 41]. Thus, the failure to detect these complexes in testis nuclear extracts is not due to an intrinsic poor translatability of these mRNAs.

DISCUSSION

In this paper we show that the mouse Oct2 gene is abundantly expressed in the germ cells of the testis. Two uniquely sized mRNAs of approximately 6.8 and 7.2 kb are detected at a high level in spermatocytes and at a much lower level in the post-meiotic round spermatids. In brain, spleen and thymus only one abundant transcript of 7 kb was detected. The expression of the Oct2 gene was further analyzed by cloning of Oct2 cDNAs from testis RNA.

Oct2 expression in testis

The genomic structure of the mouse Oct2 gene has recently been described [37, 38]. It is a single gene copy in the mouse genome and consists of at least 14 exons spread over 35kb of genomic DNA. The Oct2 gene was shown to generate at least 6 different mRNAs through alternative splicing, encoding different forms of the Oct2 protein. The exons that were shown to be subject to alternative splicing in B-cells are exon 4a and 4b, exon 7a, exon 12a and exon 13 (see Figure 7). From a comparison of the tsOct2 cDNAs and the B-cell Oct2 cDNAs, the following picture emerges. The 1-tsOct2 cDNA contains the exon sequences from exon 7a to 14, while 2-tsOct2 contains exon 7b to 14. Upstream of nt 611 the tsOct2 cDNAs don't show any homology to the published Oct2 sequence, nor to any other sequence in the EMBL DNA database. Consistent with this finding is that a probe derived from the 5' part of the tsOct2 cDNAs detected the longer of

the two testis Oct2 RNAs and none of the Oct2 transcripts in spleen, thymus or brain (Figure 2B). The tsOct2 cDNA sequence directly preceding exon 7a (at nt611) shows a perfect match to a splice acceptor site, suggesting that these sequences are derived from the intron between exon 6 and 7. Indeed, cloning, mapping and sequencing of genomic intron 6 sequences showed that the first 610 nt of the tsOct2 cDNAs are derived from intron 6. The 750 PCR oct2 cDNA molecule derives from a mRNA that arose through splicing of an intronic cryptic splice donor to the splice acceptor of exon 7b, skipping the exon 7a splice acceptor. It is of interest to note here that in B-cells the 7b splice acceptor is more frequently used than the 7a splice acceptor, resulting in a low frequency of Oct2 mRNAs containing exon 7a (5%; [38]). As both Oct2 cDNAs have an identical 5′ end that uniquely detects the longer of the two testis specific Oct2 mRNAs, it is possible that these mRNAs are transcribed from a unique promoter that is located in intron 6. Sequence analysis of the genomic region upstream of the PCR cDNA 5′ end reveals the presence of TATA and CAAT box-like elements at typical distances. This is not very surprising as the whole region is rather

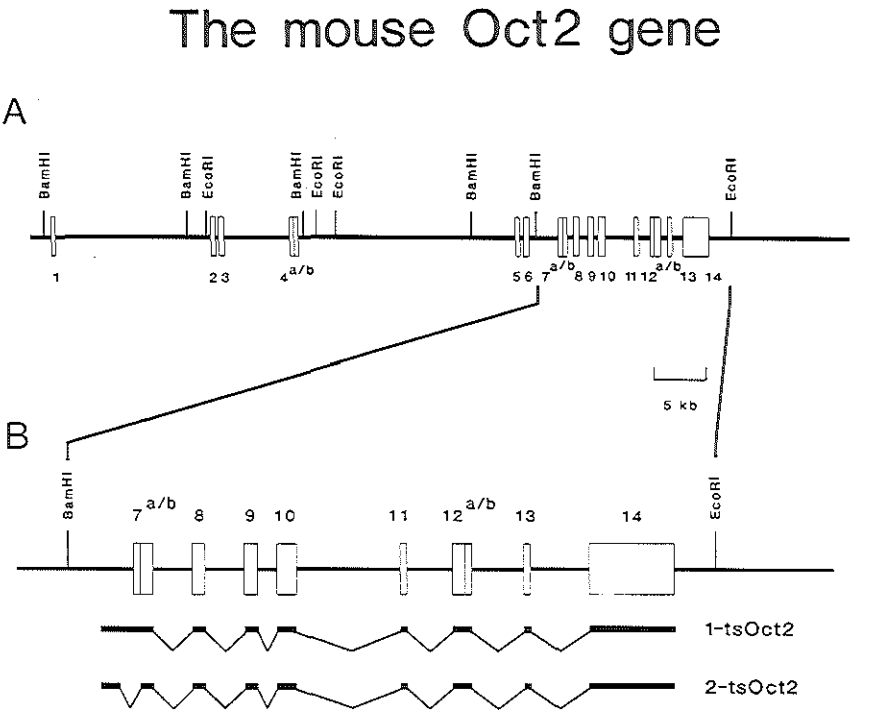


Figure 7. Schematic representation of the mouse Oct2 gene as described by Wirth *et al.* [38]. The way the tsOct2 cDNAs arose through a combination of alternative splicing and promoter usage from the Oct2 gene is schematically depicted underneath an enlarged portion of the Oct2 map.

A/T rich. It is more likely that the same 5' end of the two independent PCR cDNAs is caused by internal priming on the stretch of 10 A residues in the first strand cDNA molecules. Conventional primer extension experiments to determine the transcriptional start site of the tsOct2 mRNAs have not been successful so far. Further experiments need to be done to map the promoter region that generates Oct2 transcripts in testis. As stated above, the 5' ends of the characterized tsOct2 cDNAs detect the 7.2kb Oct2 mRNA in testis and not the 6.8kb mRNA. This indicates that this mRNA has a different 5' end. This mRNA is detected with a probe covering sequences encoded by exon 8 to exon 11 (Figure 2A).

It was shown by Wirth *et al.* [38] that exons downstream of the exons encoding the POU domain (exon 7b to exon 10) are also subject to alternative splicing (exon 12b and 13). We isolated one cDNA from a testis cDNA library that contained both exon 12b and exon13. Exon 12b was shown to be present in the majority of B-cell Oct2 mRNAs (>90%) while exon 13 was only present in 18% of the B-cell Oct2 mRNAs. We investigated the alternative splicing of exon 13 in testis and spleen (Figure7) using the cDNA PCR technique. We showed that exon 13 is present in the majority of testis Oct2 transcripts while it was only present in a minority of B-cell Oct2 RNAs. This finding is in agreement with the results obtained by Hatzopoulos *et al.* [37] who used a RNAse protection assay to study the alternative splicing of exon 13 in different tissues.

As we combined a cDNA cloned from a cDNA library with 5' extensions obtained through PCR the possibility exists that full length cDNAs are constructed that do not represent all of the different RNAs observed *in vivo*. It is possible that Oct2 RNAs are produced in testis that have the same 5'end as our tsOct2 cDNAs but lack exon13.

As shown by us and others the inclusion of exon 13 results in a shift of reading frame. This alternative reading frame terminates at a more downstream stopcodon. It was shown that B-cell Oct2 mRNA containing this exon encodes the low abundance Oct2B protein observed in B-cell nuclear extracts [37,38,39].

The tsOct2 cDNAs encode amino-terminal truncated Oct2 proteins

Because the tsOct2 cDNAs do not contain the Oct2 sequences encoded by exon 1 to 6, the ORFs in these cDNAs encode severely truncated versions of the B-cell Oct2 proteins. As the testis specific sequences in the Oct2 cDNAs do not provide a start codon, the open reading frames of the two tsOct2 cDNAs start at in frame methionine codons. As a consequence, the tsOct2 encoded proteins do not contain the three glutamine rich regions located in the amino terminal part of the B-cell Oct2 proteins. Recently, these glutamine rich regions have been shown to be involved in the transcriptional activation function of the Oct2 protein [41, 42, 43]. A second domain involved in transactivation was shown to be located carboxyl terminal of the

POU domain of Oct2. Furthermore, it was suggested that the cluster of acidic residues in the amino terminal part of the POU domain constitutes a third transactivation domain. This cluster of acidic residues (aa 27-36 in the 1-tsOct2 ORF) is not present in the 2-tsOct2 encoded protein. Whether the tsOct2 proteins can activate a promoter containing an octamer motif, in particular the c-abl type I promoter, remains to be determined. Even if the tsOct2 proteins lack an intrinsic transactivation domain, it is still possible that through interactions with other proteins the tsOct2 proteins can regulate responsive promoters via the octamer motif. Such combinatorial interactions have been studied in detail for the Oct1 protein and the Herpes Simplex Virus transactivating protein VP16 (also called VMW65 or α-TIF)[44-48]. A multiprotein complex of VP16, Oct1 and at least one other cellular factor can induce high level immediate early gene expression of HSV genes via the TAATGARAT motif, whereas Oct1 alone cannot. It is possible that similar interactions between the tsOct2 proteins and other cellular proteins exist. Such interactions might be established through the POU_{HD} domain,as was shown to be the case for Oct1/VP16 interaction [46], or through the putative leucine zipper that is located carboxyl terminal of the POU domain in the Oct2 proteins. However, it should be noted that the 2-tsOct2 encoded protein misses a part of the POU_{SP} domain resulting in a severely reduced binding affinity for the octamer sequence.

The possible role of Oct2 gene expression in the male germ line

The peculiar transcription of the Oct2 gene in male germ cells is not exceptional. Testis specific transcription of a large number of genes have been described, among which many oncogenes (for review see [2, 3, 4]). For a number of these genes it is not clear whether the testis specific transcripts they generate are translated, and if so what function these proteins play in spermatogenesis. For example the mouse fer gene, a member of the tyrosine kinase gene family, expresses a uniquely sized transcript in the germ cells of the testis [49, 50]. Analysis of cDNA clones representing a testis specific fer transcript (ferT) showed that the transcript encodes an amino terminal truncated fer protein. The ferT transcript was shown to contain a non spliced intron. The ORF is preceded by a long 5' UTR and starts within this non-spliced intron. The fer protein, encoded by this ferT transcript, still contains all of the characteristic features of a tyrosine kinase protein. Although *in vitro* transcription/translation indicates that ferT mRNA can be translated, it is not clear whether the RNA is actually translated *in vivo*. Another example is the proopiomelanocortin (POMC) gene. This gene is predominantly expressed in the pituitary gland, but testis specific transcripts have been described in rat, mouse and human [51]. In rat testis, the POMC transcripts do not contain exon1 and 2 of the POMC gene (which consists of three exons). The 5'ends of this testis specific POMC transcript were mapped using primer extension and S1 nuclease protection assays and were

shown to be heterogeneous. Depending on its exact 5'end, the testis specific transcript would encode a truncated version of the polyhormone precursor missing its first 52 to 100 amino acid residues. Furthermore no signal sequence would be present, making it unlikely that these mRNAs produce functional polypeptides.

As is the case for the ferT RNA, the tsOct2 RNAs have long 5'UTR sequences that contain multiple short open reading frames. It has been suggested that the presence of multiple start codons in 5'UTR sequences reduces the efficiency of translation of the downstream open reading frame [40]. Our *in vitro* translation and COS cell expression experiments clearly show that the tsOct2 RNAs can be translated. However, we could not detect complexes of the expected size in bandshift experiments using testis nuclear extracts (see Figure 6B) or extracts from spermatocytes or spermatids. Thus, it remains unclear whether the Oct2 mRNAs are actually translated in male germ cells. It is possible that Oct2 mRNAs are translated during a short interval in late pachytene and that they are then rapidly turned over as the cells go through the meiotic divisions. Stage specific regulation of translation has been demonstrated for a number of gene transcripts during spermatogenesis. If so, then the low overall concentration of the tsOct2 proteins in nuclear extracts of total testis would make them undetectable in our bandshift experiments. Another possible explanation is offered by the recent characterization of a new member of the POU domain family [52]. This protein called I-POU ('inhibitory POU') was isolated from *Drosophila*. The I-POU protein does not bind DNA itself but forms a complex with the *Drosophila* Cf1a POU protein, thereby inhibiting DNA binding of this protein to its recognition sequence. The characterization of the I-POU protein adds an additional level of regulation of gene expression. This situation is similar to the regulation of DNA binding activity of the transcription factors NFkB and AP1 for which inhibitors have been characterized [53, 54, 55]. It is tempting to speculate on the presence of an I-POU-like factor in male germ cells that would be able to regulate the DNA binding activity of the Oct2 proteins through complex formation. Thus, the inability to detect the tsOct2 proteins in bandshift assays could be explained by the presence of an abundant I-POU like protein that inhibits the binding of the tsOct2 proteins to the octamer sequence by complex formation. In this respect, it is noteworthy that several additional POU genes are expressed in mouse testis [56]. It could be that one of these genes, or an as yet unidentified POU gene, expresses an I-POU-like activity in testis. Thus, the question whether the tsOct2 proteins are present at any stage in male germ cell development, can only be answered with the use of antibodies directed against the Oct2 protein(s).

As long as the presence of Oct2 proteins in testis has not been proven, the possibility exists that the Oct2 transcripts are not translated at all. If so, their translation is actively prevented, as our COS cell expression

experiments show that the Oct2 RNAs can be translated *in vivo*. We can only speculate as to why the Oct2 gene is transcribed in spermatocytes, but its RNA not translated. During the pachytene stage of the meiotic prophase, the chromosomes are aligned, the synaptonemal complex is formed and recombination occurs. It could be that transcription plays an unanticipated role in any of these processes. Alternatively transcription might be a side effect of the complex organization of chromatin during meiotic prophase, or transcription is linked with recombination in an as yet unknown fashion. In any case, in testis a mechanism must have evolved that distinguishes transcripts that are to be translated at a later stage, from transcripts that are not translated at all.

MATERIALS AND METHODS

Cloning of POU domain cDNAs after amplification by PCR

Total RNA was extracted from testis by the LiCl/Urea method [57]. RNA (10 μg) was precipitated with ethanol, washed once with 70% ethanol and dried. The RNA was dissolved in 9 μl annealing buffer (250 mM KCl, 10 mM Tris-HCl pH8.3, 1mM EDTA) and 1 μl (10 pmol) of a fully degenerate anti sense oligonucleotide (5'-C(TG)(GA)TT(GA)CA(AG)AACCA(ACTG)-AC(ACTG)C-3';3'amp primer in Figure 2A) was added. The sample was heated for 3 minutes at 80 °C and subsequently incubated at 60°C for 30 minutes. The sample was cooled over a time span of approximately one hour to 42°C. After addition of 15 μl cDNA buffer (24mM Tris-HCl pH8.3, 16mM $MgCl_2$, 8mM DTT, 0.4mM dNTPs) and 10 units of AMV reverse transcriptase, the sample was incubated for 60 minutes at 42°C. Subsequently the cDNA was amplified by adding 55 μl of Taq polymerase buffer (100mM Tris-HCl pH8.8, 30mM $(NH_4)_2SO_4$, 6mM $MgCl_2$, 10 mM β-mercaptoethanol) 20 μl of dimethyl sulfoxide and 1 μl of a fully degenerate sense oligonucleotide (5'-AA(AG)TGGCT(AGCT)AA(TC)GA(CT)GC(AGCT)GA-3'; 5'amp primer in Figure 2A). The sample was denatured for 3 minutes at 93°C, and 1 unit of Taq polymerase was added. Twentyfour cycles were performed to amplify the cDNA: 5 min at 72°C (extension), 2 min at 42°C (annealing) and 1 min at 93°C (denaturation). The amplification was repeated with 10% of the material produced in the first amplification experiment. Finally the samples were desalted over a G-50 Sephadex spin-column, ethanol precipitated and dissolved in TE (10mM Tris-HCl pH7.1, 1mM EDTA) Ten percent of the material was electrophoresed through a 2% agarose gel, blotted onto a nylon membrane (Zeta probe) and hybridized to a degenerate oligonucleotide probe (5'-CA(AG)AAGCCTACCTC-(AG)GA(AG)GAGATC-3'; probe oligo in Figure 2A). DNA corresponding to the hybridizing bands was purified from a preparative agarose gel by phenol extraction and precipitation. Using Sal1 restriction enzyme sites incorporated in the amplification oligonucleotides the cDNAs were cloned in pTZ18. Positive clones were selected by colony hybridization using the probe oligonucleotide. Four positive clones were selected for further sequence analysis.

Construction and screening of a mouse testis cDNA library

The construction of the mouse testis λgt10 cDNA library has been described [58]. The library was screened with a radiolabelled probe covering the mouse Oct2 POU domain. This probe was obtained through PCR amplification of testis cDNA as described above. One positive clone was obtained. The insert of 1.4kb was subcloned in pTZ19. Five prime extensions of this cDNA were generated following the "rapid amplification of cDNA ends" (RACE) protocol as developed by Frohman *et al.*[35]. The following oligonucleotides were used;

5'RT oligo 5'-GCTGCTCAGCTGGTTTGG-3'

5'amp oligo 5'-GGTCGACGAATtCACAGACATAGTCTCTGCG-3'

PLdT15 oligo 5'-GTCGCGAATTCGTCGACGCGTTTTTTTTTTTTTTT-3'

adaptor oligo 5'-GTCGCGAATTCGTCGACGCG-3'.

Amplification products were electrophoresed through a 2% agarose gel and blotted onto

nitrocellulose filter. Five major bands, visualized by ethidium bromide staining of the gel, hybridized with a DNA probe representing the 5' part of the 1.4kb Oct2 cDNA. These bands ranged in size from 150 to 950bp. The two largest amplification products of 950bp en 750bp were subcloned and used for further analysis. The smaller amplification products have not been analyzed any further.

Subcloning and sequencing

The 1.4kb Oct2 cDNA isolated from the testis cDNA library as well as the 950bp and 750bp RACE products were further subcloned into the M13 derived phages mp18 and mp19 for dideoxysequencing with M13 universal primer. Sequencing was performed on single stranded recombinant phage DNA using T7 DNA polymerase (Pharmacia) or Sequenase (USB) according to the manufacturers specifications. Sequence data were handled using the Microgenie software package (Beckman).

Bandshift assays

Electrophoretic mobility shift assays were performed as described [29] using a double stranded oligonucleotide that was labelled with γ-(^{32}P) ATP and T4 polynucleotide kinase (Boehringer). Typically, 2 fmoles of probe were combined with 5 μg of nuclear extract in 20 μl bandshift buffer (10mM Hepes-KOH pH7.6, 60mM KCl, 1mM EDTA, 1mM DTT, 4% Ficoll, 100μg poly(dI-dC)) and the mixture was incubated at room temperature for 30 min. Samples were electrophoresed through a 4% (29:1) polyacrylamide gel in 0.25xTBE. After fixation in 10% acetic acid, 10% methanol, the gel was dried and exposed to an autoradiographic film.

***In vitro* transcription and translation**

The 1-tsOct2 and 2-tsOct2 cDNAs were subcloned as SalI-EcoRI fragments into pBluescript and linearized with SalI or EcoRI. RNA was synthesized using T3 RNA polymerase (sense RNA) or T7 RNA polymerase (α-sense RNA). *In vitro* translation was performed in rabbit reticulocyte lysates (Promega) according to the manufacturers specification. 1/50 of the *in vitro* translation mixture was used in a bandshift assay.

COS cell transfection

COS cells were maintained in DMEM/F10 (1:1), 5% FKS at 37°C and 5% CO2. Transfections were performed as described [59]. To obtain high level expression of the 1-tsOct2 and 2-tsOct2 cDNAs *in vivo*, the cDNAs were subcloned in the eukaryotic SV40 based expression vector pCDX [60] and transfected to COS cells. Mini nuclear extracts were prepared according to Schreiber *et al.* [61].

Isolation of RNA and Northern blotting

Total RNA from tissues or cells was isolated by the LiCL/Urea method [57]. Electrophoresis of RNA and blotting onto nitrocellulose membranes was done as described [62].

REFERENCES

1. **Hecht, N.B.** 1986. Regulation of gene expression during mammalian spermatogenesis. in "Experimental approaches to mammalian embryonic development", eds. J. Rossant and R. A. Pederson, Cambridge University Press, Cambridge
2. **Erickson, R.P.** 1990. Post-meiotic gene expression. *Trends Genet.* **6**: 264-269
3. **Propst, F., M.P. Rosenberg and G.F. Vande Woude.** 1988. Proto-oncogene expression in germ cell development. *Trends Genet* **4**: 183-187
4. **Willison, K. and A. Ashworth.** 1987. Mammalian spermatogenic gene expression. *Trends Genet* **3**: 351-355
5. **Gold, B., H. Fujimoto, J.M. Kramer, R.P. Erickson and N.B. Hecht.** 1983. Haploid accumulation and translational control of phosphoglycerate kinase-2 messenger RNA during mouse spermatogenesis. *Dev Biol* **98**: 392-399
6. **Braun, R.E., J.J. Peschon, R.R. Behringer, R.L. Brinster and R.D. Palmiter.** 1989. Protamine 3'-untranslated sequences regulate temporal translational control and subcellular localization of growth hormone in spermatids of transgenic mice. *Genes Dev* **3**: 793-802

7. **Falkner, F.G. and H.Z. Zachau.** 1984. Correct transcription of an immunoglobulin κ gene requires an upstream fragment containing conserved sequence elements. *Nature* **310:** 71-74
8. **Parslow, T.G., D.C. Blair, W.J. Murphy and D.K. Granner.** 1984. Structure of the 5' ends of immunoglobulin genes: a novel conserved sequence. *Proc. Natl. Acad. Sci. USA* **81:** 2650-2654
9. **Gerster, T., P. Matthias, M. Thali, J. Jiricny and W. Schaffner.** 1987. Cell type-specificity elements of the immunoglobulin heavy chain gene enhancer. *Embo J* **6:** 1323-30
10. **Wirth, T., L. Staudt and D. Baltimore.** 1987. An octamer oligonucleotide upstream of a TATA motif is sufficient for lymphoid-specific promoter activity. *Nature* **329:** 174-8
11. **Verrijzer, C.P., A.J. Kal and P.C. van der Vliet.** 1990. The DNA binding domain (POU domain) of transcription factor oct-1 suffices for stimulation of DNA replication. *Embo J* **9:** 1883-8
12. **Bohmann, D., W. Keller, T. Dale, H.R. Schöler, G. Tebb and I.W. Mattaj.** 1987. A transcription factor which binds to the enhancers of SV40, immunoglobulin heavy chain and U2 snRNA genes. *Nature* **325:** 268-72
13. **LaBella, F., H.L. Sive, R.G. Roeder and N. Heintz.** 1988. Cell-cycle regulation of a human histone H2b gene is mediated by the H2b subtype-specific consensus element. *Genes Dev* **2:** 32-9
14. **Roberts, S.B., N. Segil and N. Heintz.** 1991. Differential phosphorylation of the transcription factor Oct1 during the cell cycle. *Science* **253:** 1022-6
15. **O'Neill, E.A., C. Fletcher, C.R. Burrow, N. Heintz, R.G. Roeder and T.J. Kelly.** 1988. Transcription factor OTF-1 is functionally identical to the DNA replication factor NF-III. *Science* **241:** 1210-3
16. **Mattaj, I.W., S. Lienhard, J. Jiricny and E.M. De Robertis.** 1985. An enhancer-like sequence within the Xenopus U2 gene promoter facilitates the formation of stable transcription complexes. *Nature* **316:** 163-7
17. **Sive, H.L., N. Heintz and R.G. Roeder.** 1986. Multiple sequence elements are required for maximal *in vitro* transcription of a human histone H2B gene. *Mol Cell Biol* **6:** 3329-40
18. **Segil, N., S.B. Roberts and N. Heintz.** 1991. Mitotic Phosphorylation of the Oct-1 Homeodomain and Regulation of Oct-1 DNA Binding Activity. *Science* **254:** 1814-1816
19. **Staudt, L.M., H. Singh, R. Sen, T. Wirth, P.A. Sharp and D. Baltimore.** 1986. A lymphoid-specific protein binding to the octamer motif of immunoglobulin genes. *Nature* **323:** 640-3
20. **Clerc, R.G., L.M. Corcoran, J.H. LeBowitz, D. Baltimore and P.A. Sharp.** 1988. The B-cell-specific Oct-2 protein contains POU box- and homeo box-type domains. *Genes Dev* **2:** 1570-81
21. **Müller, M.M., S. Ruppert, W. Schaffner and P. Matthias.** 1988. A cloned octamer transcription factor stimulates transcription from lymphoid-specific promoters in non-B cells. *Nature* **336:** 544-51
22. **Scheidereit, C., J.A. Cromlish, T. Gerster, K. Kawakami, C.-G. Balmaceda, R.A. Currie and R.G. Roeder.** 1988. A human lymphoid-specific transcription factor that ctivates immunoglobulin genes is a homeobox protein. *Nature* **336:** 551-557
23. **Sturm, R.A., G. Das and W. Herr.** 1988. The ubiquitous octamer-binding protein Oct-1 contains a POU domain with a homeo box subdomain. *Genes Dev* **2:** 1582-99
24. **Herr, W., R.A. Sturm, R.G. Clerc, L.M. Corcoran, D. Baltimore, P.A. Sharp, H.A. Ingraham, M.G. Rosenfeld, M. Finney, G. Ruvkun and H.R. Horvitz.** 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:** 1513-6
25. **Ingraham, H.A., S.E. Flynn, J.W. Voss, V.R. Albert, M.S. Kapiloff, L. Wilson and M.G. Rosenfeld.** 1990. The POU-specific domain of Pit-1 is essential for sequence-specific, high affinity DNA binding and DNA-dependent Pit-1-Pit-1 interactions. *Cell* **61:** 1021-33
26. **Sturm, R.A. and W. Herr.** 1988. The POU domain is a bipartite DNA-binding structure. *Nature* **336:** 601-4

27. **Verrijzer, C.P., A.J. Kal and P.C. van der Vliet.** 1990. The oct-1 homeo domain contacts only part of the octamer sequence and full oct-1 DNA-binding activity requires the POU-specific domain. *Genes Dev* **4:** 1964-74
28. **Gehring, W.J.** 1987. Homeo boxes in the study of development. *Science* **236:** 1245-52
29. **Barberis, A., F.G. Superti and M. Busslinger.** 1987. Mutually exclusive interaction of the CCAAT-binding factor and of a displacement protein with overlapping sequences of a histone gene promoter. *Cell* **50:** 347-59
30. **Barberis, A., F.G. Superti and M. Busslinger.** 1989. The testis-specific octamer-binding protein of the sea urchin has a molecular weight of 85 kDa. *Nucleic Acids Res* **17:** 7114
31. **Cox, P.M., S.M. Temperley, H. Kumar and C.R. Goding.** 1988. A distinct octamer-binding protein present in malignant melanoma cells. *Nucleic Acids Res* **16:** 11047-56
32. **Lenardo, M.J., L. Staudt, P. Robbins, A. Kuang, R.C. Mulligan and D. Baltimore.** 1989. Repression of the IgH enhancer in teratocarcinoma cells associated with a novel octamer factor. *Science* **243:** 544-6
33. **Schöler, H.R., A.K. Hatzopoulos, R. Balling, N. Suzuki and P. Gruss.** 1989. A family of octamer-specific proteins present during mouse embryogenesis: evidence for germline-specific expression of an Oct factor. *Embo J* **8:** 2543-50
34. **Bernards, A., M. Paskind and D. Baltimore.** 1988. Four murine c-abl mRNAs arise by usage of two transcriptional promoters and alternative splicing. *Oncogene* **2:** 297-304
35. **Frohman, M.A., M.K. Dush and G.R. Martin.** 1988. Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. *Proc Natl Acad Sci USA* **85:** 8998-9002
36. **He, X., M.N. Treacy, D.M. Simmons, H.A. Ingraham, L.W. Swanson and M.G. Rosenfeld.** 1989. Expression of a large family of POU-domain regulatory genes in mammalian brain development. *Nature* **340:** 35-41
37. **Hatzopoulos, A.K., A.S. Stoykova, J.R. Erselius, M. Goulding, T. Neuman and P. Gruss.** 1990. Structure and expression of the mouse Oct2a and Oct2b, two differentially spliced products of the same gene. *Development* **109:** 349-62
38. **Wirth, T., A. Priess, A. Annweiler, S. Zwilling and B. Oeler.** 1991. Multiple Oct2 isoforms are generated by alternative splicing. *Nucleic Acids Res.* **19:** 43-51
39. **Schreiber, E., P. Matthias, M.M. Müller and W. Schaffner.** 1988. Identification of a novel lymphoid specific octamer binding protein (OTF-2B) by proteolytic clipping bandshift assay (PCBA). *Embo J* **7:** 4221-9
40. **Kozak, M.** 1989. The scanning model for translation: an update. *J Cell Biol* **108:** 229-41
41. **Müller, I.M., W. Schaffner and P. Matthias.** 1990. Transcription factor Oct-2A contains functionally redundant activating domains and works selectively from a promoter but not from a remote enhancer position in non-lymphoid (HeLa) cells. *Embo J* **9:** 1625-34
42. **Gerster, T., C.G. Balmaceda and R.G. Roeder.** 1990. The cell type-specific octamer transcription factor OTF-2 has two domains required for the activation of transcription. *Embo J* **9:** 1635-43
43. **Tanaka, M. and W. Herr.** 1990. Differential transcriptional activation by Oct-1 and Oct-2: interdependent activation domains induce Oct-2 phosphorylation. *Cell* **60:** 375-86
44. **O'Hare, P. and C.R. Goding.** 1988. Herpes simplex virus regulatory elements and the immunoglobulin octamer domain bind a common factor and are both targets for virion transactivation. *Cell* **52:** 435-45
45. **Preston, C.M., M.C. Frame and M.E. Campbell.** 1988. A complex formed between cell components and an HSV structural polypeptide binds to a viral immediate early gene regulatory DNA sequence. *Cell* **52:** 425-34
46. **Stern, S., M. Tanaka and W. Herr.** 1989. The Oct-1 homoeodomain directs formation of a multiprotein-DNA complex with the HSV transactivator VP16. *Nature* **341:** 624-30
47. **Kristie, T.M. and P.A. Sharp.** 1990. Interactions of the Oct-1 POU subdomains with specific DNA sequences and with the HSV alpha-trans-activator protein. *Genes Dev* **4:** 2383-96
48. **Goding, C.R. and P. O'Hare.** 1989. Herpes simplex virus Vmw65-octamer binding protein

interaction: a paradigm for combinatorial control of transcription. *Virology* **173:** 363-7
49. **Fischman, K., J.C. Edman, G.M. Shackleford, J.A. Turner, W.J. Rutter and U. Nir.** 1990. A murine fer testis-specific transcript (ferT) encodes a truncated Fer protein. *Mol Cell Biol* **10:** 146-53
50. **Keshet, E., A. Itin, K. Fischman and U. Nir.** 1990. The testis-specific transcript (ferT) of the tyrosine kinase FER is expressed during spermatogenesis in a stage-specific manner. *Mol Cell Biol* **10:** 5021-5
51. **Jeannotte, L., J.P. Burbach and J. Drouin.** 1987. Unusual proopiomelanocortin ribonucleic acids in extrapituitary tissues: intronless transcripts in testes and long poly(A) tails in hypothalamus. *Mol Endocrinol* **1:** 749-57
52. **Treacy, M.N., X. He and M.G. Rosenfeld.** 1991. I-POU: a POU-domain protein that inhibits neuron-specific gene activation. *Nature* **350:** 577-584
53. **Baeuerle, P.A. and D. Baltimore.** 1988. I kappa B: a specific inhibitor of the NF-kappa B transcription factor. *Science* **242:** 540-6
54. **Zabel, U. and P.A. Baeuerle.** 1990. Purified human I kappa B can rapidly dissociate the complex of the NF-kappa B transcription factor with its cognate DNA. *Cell* **61:** 255-65
55. **Auwerx, J. and C.P. Sassone.** 1991. IP-1: a dominant inhibitor of Fos/Jun whose activity is modulated by phosphorylation. *Cell* **64:** 983-93
56. **Goldsborough, A., A. Ashworth and K. Willison.** 1990. Cloning and sequencing of POU-boxes expressed in mouse testis. *Nucleic Acids Res* **18:** 1634
57. **Sambrook, J., E. Fritsch and T. Maniatis.** 1989. "Molecular Cloning, a laboratory manual" Cold Spring Harbor Laboratory, New York
58. **Meijer, D., A. Hermans, M. von Lindern, T. van Agthoven, A. de Klein, P. Mackenbach, A. Grootegoed, D. Talarico, G. Della Valle and G. Grosveld.** 1987. Molecular characterization of the testis specific c-abl mRNA in mouse. *Embo J* **6:** 4041-8
59. **Meijer, D., A. Graus, R. Kraay, A. Langeveld, M.P. Mulder and G. Grosveld.** 1990. The octamer binding factor Oct6: cDNA cloning and expression in early embryonic cells. *Nucleic Acids Res* **18:** 7357-65
60. **Okayama, H. and P. Berg.** 1983. A cDNA cloning vector that permits expression of cDNA inserts in mammalian cells. *Mol Cell Biol* **3:** 280-9
61. **Schreiber, E., P. Matthias, M.M. Müller and W. Schaffner.** 1989. Rapid detection of octamer binding proteins with 'mini-extracts', prepared from a small number of cells. *Nucleic Acids Res* **17:** 6419
62. **Fourney, R.M., J. Miyakoshi, R.S. Day III and M.C. Paterson.** 1988. Northern Blotting: Efficient RNA staining and transfer. *Focus* **10:** 5-7

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-HOMEO domain 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-HOMEO domain (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 μmolar all trans retinoic acid (RA;Sigma). These modifications of the original protocol prevents massive cell death, occuring 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 μg of plasmid DNA and 100 μg/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) gagaggaATTTGCATtccaccgaccttcc. Typically, 2 fmoles (10.000 cpm) of ds oligonucleotide was incubated with 5 μg of nuclear protein in 10 mM Hepes (pH 7.9), 60 mM KCl, 1 mM DTT, 1 mM EDTA, 4% Ficoll and 2 μgrams of poly dI-dC at room temperature, for twenty minutes. DNA-protein complexes were separated on a 4% polyacrylamide gel in 0.25×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	:	TAGGA*ATTTGCAT*TTCCGATC
U2	:	TGGTTGTGGCCGTCACAAAGAGGCGGGGCT *ATGCAAAT*AGGGTGTGCCGGGGCAGTCGGG
ad2	:	AGGCCAA*TATGATAAT*GAGGGGT
ad4	:	CACGCCTT*ATTTGCAT*ATTACT
ICP4	:	AGGGCGG*TAATGAGAT*GCCATGT
abl I(mut)	:	TAGGA*ATGTTCAG*TTCCGATC
hep$^+$oct$^+$	:	CGAGTG*CTCATGAATATGCAAAT*CAATTGG
hep$^+$oct$^-$	:	CGAGTG*CTCATGA*ATATCAGTCGCCATTGG

The 300 bp EcoRI fragment used in the binding competition experiment (Fig.6) were derived from the plasmids p6W$_3$TKCAT (FD) and p6W$_3$TKCAT(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 μg 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. Hybriziation 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×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 βgt10 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 neurectodermal 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, concomittant with a decreased intensity of the Oct4 complex (small arrows in Fig. 1A). At the same time, a weak complex appears that we refer

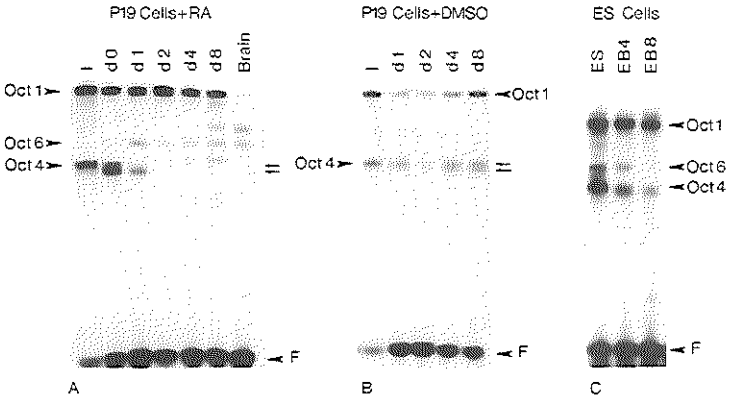


Figure 1. Octamer binding factors in nuclear extracts of differentiating P19 and ES cells. Radiolabelled probe was incubated with nuclear extracts of P19 or ES cells at different stages of differentiation. 3–5 μg of extract was used per lane. The position of the Oct1, Oct4 and Oct6 complexes are indicated. 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 then 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 then 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 neurectodermal 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.

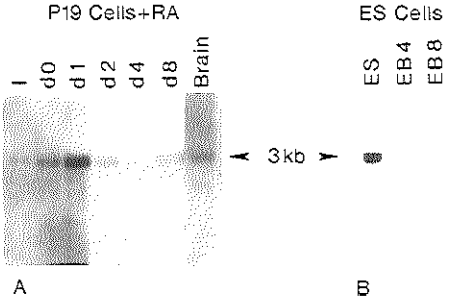


Figure 2. Expression of mb1 mRNA in differentiating P19 and ES cells. Cytoplasmic RNA and nuclear extracts were prepared from the same batch of cells. Nuclear extracts were used in the bandshift experiments presented in Fig. 1. Twenty μg of cytoplasmic RNA was denatured, separated on a 1% agarose gel 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×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,

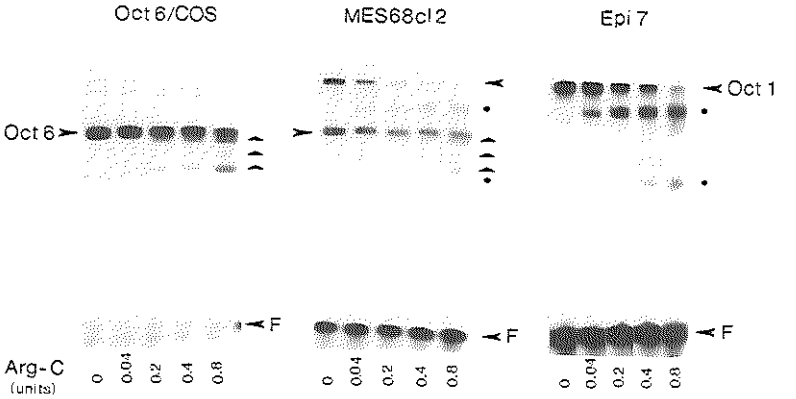


Figure 3. Protease clipping bandshift ssay of the Oct6 protein. Radiolabelled probe was incubated at roomtemperature with nuclear extracts of Oct6 transfected COS-1 cells (left), MES68cl2 (middle) and Epi7 cells (right). After ten minutes, different amounts of the endoprotease ArgC was added (amounts 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

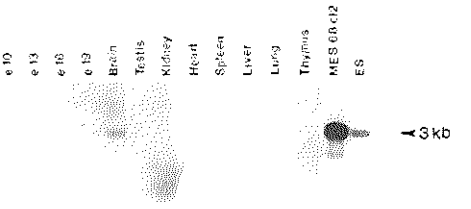


Figure 4. Expression of Oct6 mRNA in different mouse tissues and cell lines. Twenty μg of denatured total RNA was seperated on a 1% agarose gel in 0.66M formaldehyde and blotted onto Biotrans blotting membrane. Probe and hybridization conditions as in Fig. 2. Lane 1–4: total RNA from whole embryos at 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

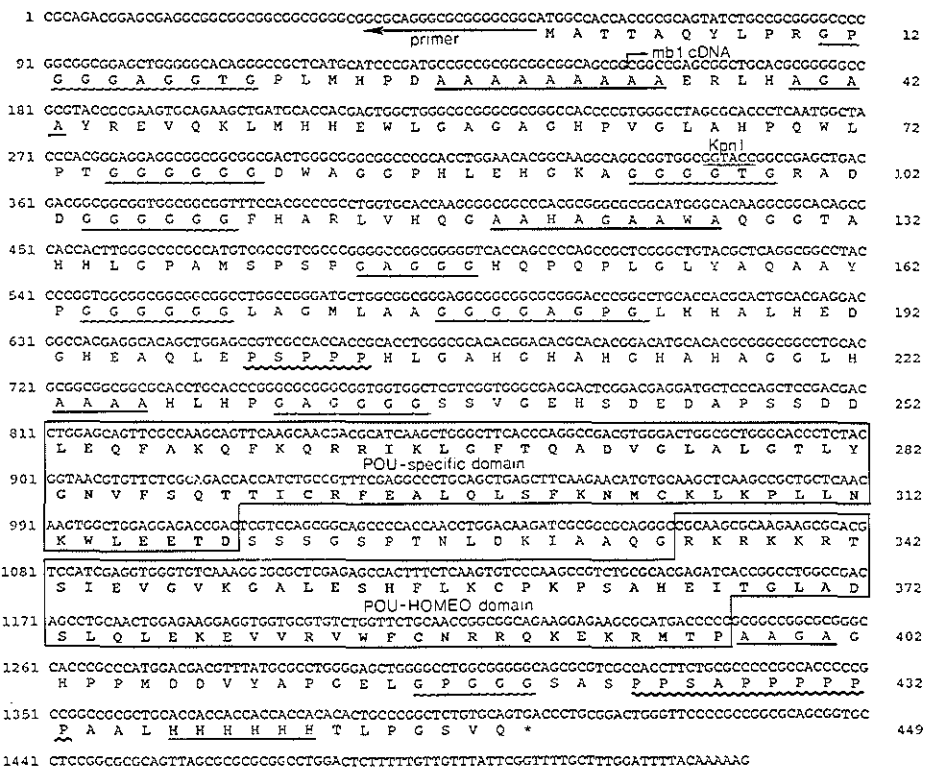


Figure 5. Nucleotide and predicted amino acid sequence of the extended mb1 cDNA encoding Oct6. Total length of the presented sequence is 1517nt. Nucleotide sequence of cDNA clone mb1 starts at position 155. The first 154 nt are derived from a genomic 510 bp SacI-KpnI fragment. Position 1 of the presented 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 aminoacid sequences. Homopolymeric, or quasi-homopolymeric amino acid sequences are underlined. The KpnI site that was used to link the genome Oct6 sequences to 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 (ablI, 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

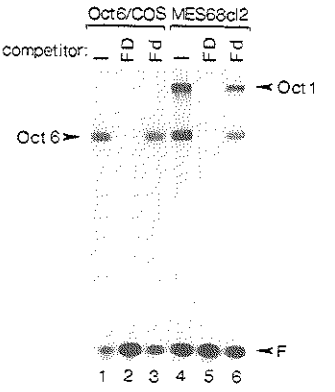


Figure 6. The Oct6 protein binds to the octamer motif. Radiolabelled probe was incubated with nuclear extracts of Oct6 transfected COS cells (lane 1 to 3) or MES68cl2 nuclear extract (lane 4 to 6) either in the absence (lane 1 and 4) or in the presence of a 1000 fold molar excess of cold competitor DNA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

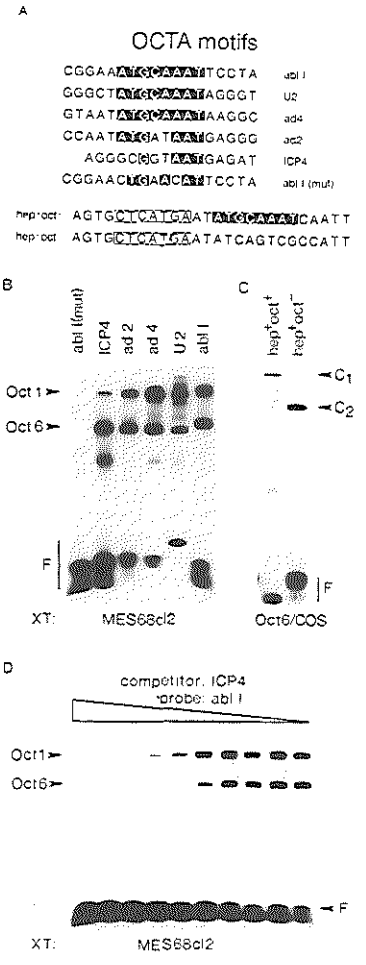


Figure 7. The octamer binding factors Oct1 and Oct6 bind with different affinities to sequences related to the consensus octamer motif. Panel A shows the sequence of the different OCTA motifs present in the ds oligonucleotides used. Nucleotides fitting the octamer concensus sequence are indicated in 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 serotyppe 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 tht relative affinities of Oct1 and Oct6 fore 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 ablI (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 ablI 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 comitted 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 appearrance of the Oct6 complex as it is one of the first complexes to be induced in P19 cells upon RA additon 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 proteinof 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 multiprotien complex on the TAATGARAT motif containing Oct1, the viral protein Vmw65 and a third cellular factor×(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 prelimenary experiments indicate that the Oct6 protein is capable of activating a minimal promoter linked to a synthetic enhancer containing a mutimerized 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 Sjozef van Baal for computer assistance.

REFERENCES

1. Owen, D. and Kühn, L.C. (1987). EMBO J. 6, 1287–1293.
2. Rothenberger, S., Müllner, E.W. and Kühn, L.C. (1990) Nucl.Acid Res. 18, 1175–1179.
3. Baker, B.S. (1989) Nature 340, 521–524.
4. Bingham, P.M., Ch u, T.B., Mims, I. and Zachar, Z. (1988) Trends in Gen. 4, 134–138.
5. Hatzopoulos, A.K., Schlokat, U. and Gruss, P. (1988) in: Transcription and Splicing eds. B.D. Hames and D.M. Glover (IRL Press, Oxford).
6. Mitchell, P.J. and Tjian, R. (1989) Science 245, 371–378.
7. Kemler, I. and Schaffner, W. (1990) FASEB J. 4, 1444–1449.
8. Labella, F., Sive, H.L., Roeder, R.G. and Heintz, N. (1988) Genes & Dev. 2, 32–39.
9. Schöler, H.R., Hatzopoulus, A.K., Balling, R., Suzuki, N. and Gruss, P. (1989) EMBO J. 8, 2543–2550.
10. Ko, H.S., Fast, P., McBride, W. and Staudt, L.M. (1988) Cell 55, 135–144.
11. Sturm, R.A., Das, G. and Herr, W. (1988) Genes & Dev. 2, 1582–1599.
12. Müller, M.M., Ruppert, S., Schaffner and Matthias, P. (1988) Nature 336, 544–551.
13. Scheidereit, C., Cromlish, J.A., Gerster, T., Kauvakamic, K., Balmaceda, C.G., Currie, R.A. and Roeder, R.G. (1988) Nature 336, 551–557.
14. Clerc, R.G., Corcoran, L.M., LeBowitz, D. and Sharp P.A. (1988) Genes & Dev. 2, 1570–1581.
15. Schöler, H.R., Ruppert, S., Suzuki, N., Chowdhary, K. and Gruss, P. (1990) Nature 344, 435–439.
16. Okamoto, K., Okazawa, H., Okuda, A., Sakai, M., Muramatsu, M. and Hamada, H. (1990) Cell 60, 461–472.
17. 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.
18. Monuki, E.S., Weinmaster, G., Kuhn, R. and Lemke, G. (1989) Neuron 3, 783–793.
19. He, X., Treacy, M.N., Simmons, D.M., Ingraham, H.A., Swanson, L.W. and Rosenfeld, M.G. (1989) Nature 340, 35–42.
20. Robertson, E., Bradley, A., Kuehn, M. and Evans, M. (1986) Nature 33, 445–448.
21. Rudnicki, M.A. and McBurney, M.W. (1987) in: Teratocarcinomas and Embryonic Stem Cells: A Practical Approach, E.J. Robertson, ed. (IRL Press, Oxford).
22. Robertson, E.J. (1987) in: Teratocarcinomas and Embryonic Stem Cells: A Practical Approach. E.J. Robertson, ed. (IRL Press, Oxford).
23. Schreiber, E., Matthias, P., Müller, M.M. and Schaffner, W. (1989) Nucl. Acid Res. 17, 6419.
24. Cough, N.M. (1988) Anal.Biochem. 173, 93–95.
25. Kalb, V.R. and Bernlohr, R.M. (1977) Anal.Biochem. 82, 362–371.
26. Barberis, A., Superti-Furga, G. and Busslinger, M. (1987) Cell 50, 347–359.
27. Bernards, A., Paskind, M. and Baltimore, D. (1988) Oncogene 2, 297–304.
28. Schreiber, E., Matthias, P., Müller, M.M. and Schaffner, W. (1988) EMBO J. 7, 4221–4229.
29. Lichtsteiner, S., Wuarin, J. and Schibler, U. (1987) Cell 51, 963–973.
30. Gerster, T. Matthias, P., Thali, M., Jiricny, J. and Schaffner, W. (1987) EMBO J. 6, 1323–1330
31. Auffray, A and Rougeon, F. (1980) Eur.J.Biochem. 107, 303–314.
32. Fourney, R.M., Miyakoshi, J., DayIII, R.S. and Paterson, M.C. (1988) Focus 10, 5–7.
33. Feinberg, A.P. and Vogelstein, B. (1983) Anal.Biochem 132, 6–13.
34. 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.
35. Sambrook, J., Fritsch, E.F. and .Maniatis, T. (1989) Molecular cloning: A laboratory manual (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory).
36. Sanger, F.G., Nicklen, S. and Coulson, A.R. (1977) Proc.Natl.Acad.Sci. USA 74, 5463–5467.
37. Jones-Villeneuve, E.M.V., Rudnicki, M.A., Harris, J.F. and McBurney, M.W. (1983) Mol.Cell.Biol. 3, 2271–2279.
38. McBurney, M.W., Reuhl, K.R., Ally, A.I., Nasipuri, S., Bell, J.C. and Craig, J. (1988) J.Neurosci. 8, 1063–1073.
39. Okayama, H. and Berg, P. (1983) Mol.Cell Biol. 3, 280–289.
40. Johnson, W.A. and Hirsh, J. (1990) Nature 343, 467–470.
41. Schöler, H.R., Balling, R., Hatzopoulos, A.K., Suzuki, N. and Gruss, P. (1989) EMBO J. 8, 2551–2557.
42. Kozak, M. (1989) J.Cell.Biol. 108, 229–241.
43. Elsholtz, H.P., Albert, V.R., Treacy, M.N. and Rosenfeld, M.G. (1990) Genes & Dev. 4, 43–51
44. 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.
45. Mermod, N., O'Neill, E.A., Kelly, T.J. and Tjian, R. (1989) Cell 58, 741–753
46. Bohmann, D., Keller, W., Dale, T., Schöler, H.R., Tebb, G. and Mattaj, I.W. (1987) Nature 325, 268–272.
47. O'Hare, P. and Goding, C.R. (1988) Cell 52, 435–445.
48. Verrijzer, C.P., Kal, A.J. and van der Vliet, P.C. (1990) Genes & Dev. in the press.
49. Baumruker, T., Sturm, R. and Herr, W. (1988) Genes & Dev. 2, 1400–1413.
50. Kemler, I., Schreiber, E., Müller, M.M., Matthias, P. and Schaffner, W. (1989) EMBO J. 8, 2001–2008.
51. Gerster, T. and Roeder, R.G. (1988) Proc.Natl.Acad.Sci.USA 85, 6347–6351.
52. Stern, S., Tanaka, M. and Herr, W. (1989) Nature 341, 624–630.
53. Roizman, B. and Sears, A.E. (1987) Ann.Rev.Microbiol. 41, 543–571.

Chapter VII

Mapping the transactivation domain of the Oct6 POU transcription factor.

Dies Meijer, Anneke Graus and Gerard Grosveld
Department of Cell Biology and Genetics, Erasmus University, Rotterdam, the Netherlands

ABSTRACT

The POU transcription factor Oct6 is expressed in embryonic stem cells, glial progenitor cells and in a restricted set of neurons in the CNS. The protein has been shown to act as a transactivator as well as a repressor. Here we show that the Oct6 protein activates transcription from three different promoters in Hela cells. The ability to activate a minimal tk promoter via a multimerized IgH enhancer octamer motif relies on a domain within the amino terminal third of the protein. Parts of this domain can be deleted without abolishing transactivation, suggesting that there is functional redundancy within this region. The transactivation domain of the Oct6 protein is different from other described activation domains in that it is highly glycine and alanine rich.

INTRODUCTION

The octamer motif (ATGCAAAT) is a well studied DNA motif present in enhancers and promoters of both ubiquitous and cell type specific expressed genes [1, 2, 3]. This DNA element was shown to be a binding site for a family of nuclear proteins present in different cell types and at different stages of mammalian development [4, 5]. A number of these proteins are believed to be involved in determining cell fate through selective regulation of target genes. The cloning of different cDNAs encoding several of these octamer binding factors revealed that they belong to the POU domain gene family [6-13]. This family was defined by a region of extensive sequence homology (the POU domain) between three mammalian transcription factors (Oct1, Oct2 and Pit1/GHF1) and one nematode regulatory protein (Unc86; [14] and references therein). The POU domain constitutes the DNA binding domain of these proteins and can be subdivided in two regions, separated by a short linker [10, 12, 15, 16]. The carboxyl terminal part shows homology with the classical homeobox proteins (the POU_{HD} Domain), while the amino terminus contains a homology specific for this class of proteins (the POU_S domain).

Oct6 was originally defined as an embryonic stem cell specific octamer binding factor [4]. Differentiation of these cells *in vitro* leads to a downregulation of the Oct6 protein. However in undifferentiated P19 EC cells the Oct6 gene is expressed at very low levels. The gene is transiently upregulated when these cells differentiate into neuronal cell types after aggregation and addition of retinoic acid to the culture medium. Expression increases again after several days of induction, indicating that the protein

plays a role in different stages or cell types during neuronal differentiation [7].

Cloning and sequencing of cDNAs encoding the Oct6 protein revealed that the Oct6 gene is the mouse homologue of the rat SCIP\Tst-1 gene [17]. The SCIP gene is highly expressed during glial cell development in the peripheral and central nervous system. It has been shown that the SCIP protein functions as a repressor of meylin specific genes during a period of rapid cell division that seperates a premyelinating from a myelinating phase in Schwann cell development [18]. Furthermore, using a PCR based approach, a partial Oct6/SCIP was cloned from neonatal rat testis which was named Testes-1 (Tst-1 [19]). Using *in situ* hybridization it was shown that the Oct6/SCIP/Tst-1 gene is also highly expressed in discrete regions of the developing nervous system [13, 19]. The Oct6 protein was shown to function as a positive as well as a negative regulator of transcription depending on the exact promoter architecture [13, 18, 20].

Here we show that high level ectopic expression of the Oct6 protein in HELA cells can activate transcription from three different promoters. Using a series of Oct6 deletion mutants we show that the transactivation domain of the protein is located in the first 157 amino acids and is distinct from its DNA binding domain. This domain, which is extremely glycine and alanine rich, can be subdivided in at least two active subdomains.

RESULTS

Oct6 transactivates a variety of promoters when ectopically expressed in HELA cells.

To study the transactivation potential of the Oct6 protein we used three different promoter constructs. A simplified promoter in which the octamer is located 20 bp upstream of the rabbit β-globin TATA box ($p\beta O^{+}Cat$ in Figure 1A) driving a CAT reporter gene. An almost identical promoter construct was shown to constitute a B-Cell specific promoter [8, 21]. This promoter could be readily activated by ectopically expressed Oct2A, Oct4 and Oct6 proteins [8, 13, 22]. Transcription from this promoter was shown to be dependent on an intact octamer motif. A second promoter construct consists of a minimal HSV-1 TK gene promoter flanked by a multimerized IgH enhancer octamer/μE4 motif (6WtkCat; [23, 24]). This type of enhancer was shown to be at least 1000 fold more active than its mutated counterpart in EC cells (6FdtkCAT in which only the octamer box was mutated;[24]). In this particular arrangement the closest octamer motif is 147 bp separated from the tk TATA box. A third promoter used in this study is the Herpes Simplex Virus 1 ICP4(IE175K) gene promoter [25]. This promoter contains one TAATGARAT motif at position -260 and a second related motif (TAATGGAAT) at position -110. The distal TAATGARAT motif of the ICP4 promoter has been shown to be a strong binding site for Oct6 wheras it is a weak binding site for Oct1 [7]. The reporter constructs are schematically

depicted in Figure 1A. In order to express the Oct6 protein in HELA cells the Oct6 cDNA was cloned in a CMV promoter/enhancer based expression vector ([26]; see M&M for details). The Oct6 expression vector and the different CAT reporters were co-transfected in different combinations in HELA cells. Reporter gene expression was measured by CAT assays [27]. Expression of the Oct6 protein was monitored by a bandshift assay (not shown).

As can be seen in Figure 2A Oct6 activates the simple βO^+ promoter confirming earlier reports [13]. Furthermore this activation is dependent on an intact octamer motif (compare lane 2 with lane 5). The same promoter

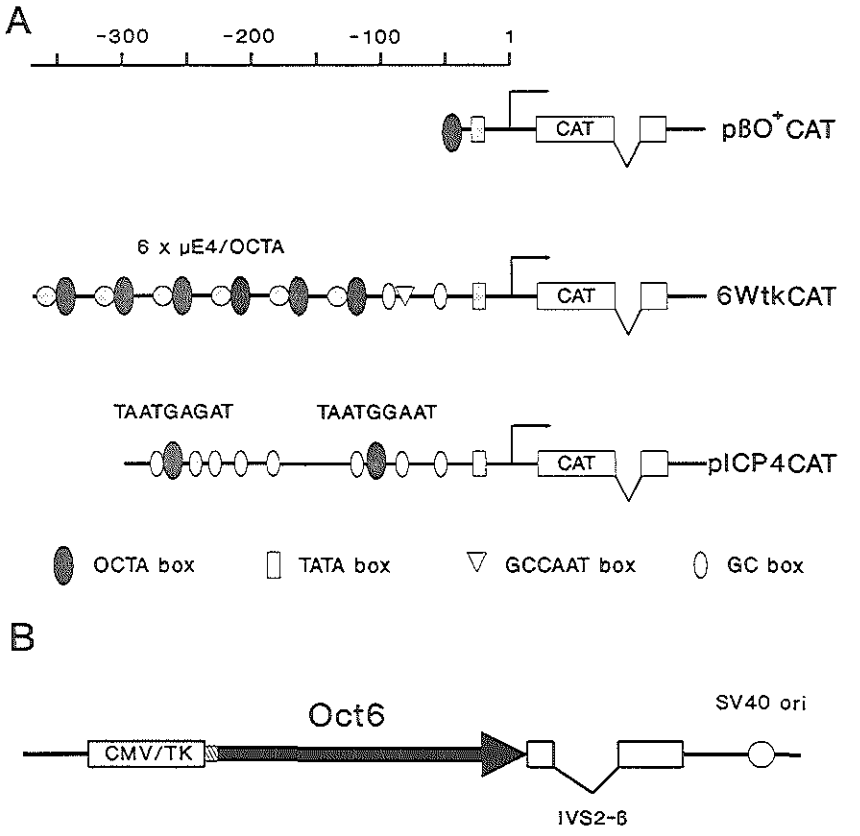


Figure 1. A. Schematic representation of the reporter plasmids used in this study. The reporter gene is the bacterial Cloramphenicol Acetyl Transferase gene (CAT). The 6WtkCAT and the ICP4CAT construct do not contain additional enhancers downstream of the CAT gene. The scale bar above the drawing indicates the position (in basepares) of the cis-acting elements relative to the transcriptional start site (arrow).
B. Structure of the Oct6 expression vector.

could be activated by Oct2A (lane 4 and [8]) and a chimaeric protein POU/VP16. This chimaeric protein consists of the Oct6 DNA binding domain (POU domain) coupled to the VP16 transactivation domain ([28, 29];see M&M) and is a strong transcriptional activator (see also Figure 4, lane 15). Although this promoter construct is clearly responsive, the absolute levels of expression are rather low.

In a second set of transactivation experiments we used the 6WtkCAT reporter and its mutated counterpart 6FdtkCAT. Clearly Oct6 is able to transcativate the tk promoter via the IgH octamer motif (fig 2B, lane 1 and 2) This results contrasts with an earlier report by Suzuki *et al.* [13] who reported that the Oct6 protein is not able to activate this reporter (see discussion). Furthermore, the HSV1 ICP4 promoter is also strongly induced by Oct6. Thus the Oct6 protein is able to activate transcription via octamer

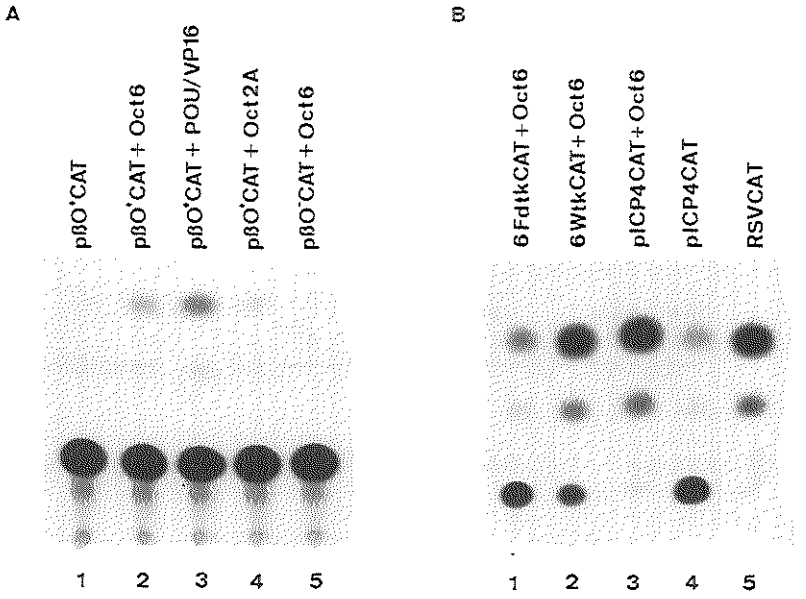


Figure 2. A. Oct6 activates transcription from a simple octamer/TATA promoter. The expression vectors pOct6, pPOU/VP16 and Oct2A and reporter plasmids were cotransfected into Hela cells as indicated. The structure of the Oct6 expression vector is outlined in figure 1. The structure of the POU/VP16 vector is described in Materials and Methods. The Oct2A vector is described in [8]. When no expression vector is indicated the reporter plasmid is cotransfected with an equal amount of the empty expression vector pEVRF0. B. Oct6 activates also more complex promoters.
The Oct6 expression vector (see figure1) was cotransfected with CAT reporter plasmids as indicated. Reporter constructs are as outlined in figure 1. The 6Fdtk promoter construct is the mutated counterpart of the 6Wtk promoter [24]. The RSVCAT construct serves as positive control, in which the CAT gene is driven by the strong Rous Sarcoma Virus promoter.

and octamer related sequences in different promoter settings.

Transactivation by ectopically expressed Oct6 is not restricted to Hela cells. In an earlier report by Suzuki *et al.* it was shown that Oct6 is not able to activate transcription from a distal position (with respect to the transcription initiation site) when expressed in Hela cells [13]. Similar findings were reported for the B cell specific Oct2A and the stem cell specific Oct4 proteins [22, 30]. One explanation for the ability of Oct6 to activate transcription from a more distal position in the experiments presented here is that our Hela cell line contains factors that are missing in their particular line or that our cell line misses factors that otherwise would prevent transactivation. To test this possibility we performed transactivation experiments using the 6WtkCAT reporter in cell lines of different origin. The MES cell line is stable derivative of the P19 EC cell line. and CHO is a Chinese Hamster Ovarian cell line. As can be seen in Figure 3 Oct6 activate the 6WtkCAT reporter gene in all three cell lines tested, indicating that there are no qualitative differences between the cell lines with respect to

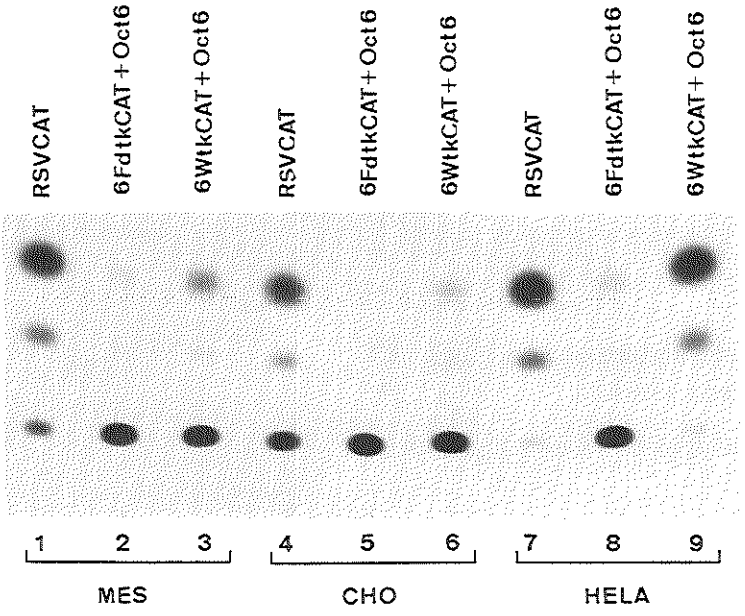


Figure 3. Oct6 transactivates the 6Wtk promoter in different cellular backgrounds. MES, CHO and Hela cells were transfected with the expression and reporter plasmids as indicated. $CaPO_4$/DNA precipetates were prepared as described and split into three portions and applied to the different cells.

supporting transactivation. However there are clear quantative diferences in the level of activation in the different cellular backgrounds. Transfection efficiency and levels of activation are highest in Hela cells. Therefore this cell line was chosen for mapping of the transactivation domain of the Oct6 protein.

Deletion mapping of the Oct6 transactivation domain.

In order to map the domain(s) of Oct6 involved in transactivation, a set of progressive amino terminal and carboxyl terminal deletions was constructed, using natural restriction sites in the Oct6 gene [7]. These truncated cDNAs were cloned in frame into CMV/enhancer based expression vectors [26]. All mutant Oct6 proteins have an intact POU domain. To test whether the different deletion constructs encode stable proteins that bind to the octamer motif and are localized in the nucleus, plasmids were transfected into COS-1 cells. The expressed mutant Oct6 proteins were assayed in a bandshift experiment using nuclear extracts from the transfected COS-1 cells and a radiolabelled OCTA probe. All constructs express Oct6 mutant proteins of the expected size at high levels. These proteins appear to be stable and are located in the nucleus (data not shown).

The transcriptional activation potential of the mutant Oct6 proteins was tested in cotransfection experiments with the 6WtkCAT reporter plasmid in Hela cells and quantative CAT assays (Figure 4, see also Figure 6). To exclude the possibility that differences in CAT activity are due to different levels of Oct6 mutant proteins, the amount of protein was estimated in a

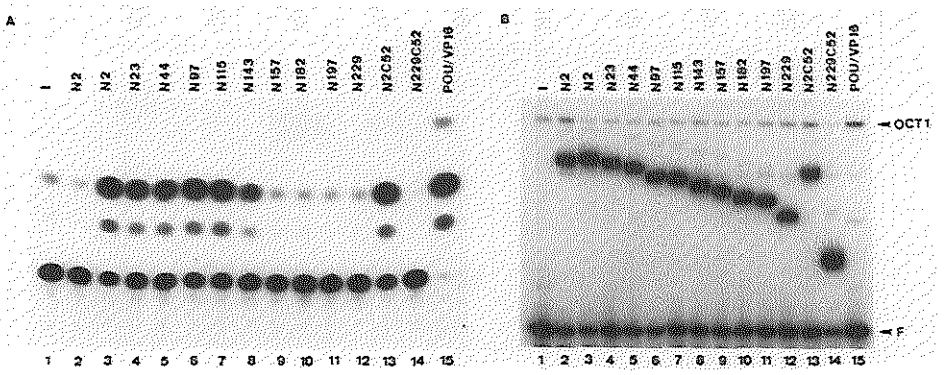


Figure 4. A. Transactivation of the 6Wtk promoter by truncated Oct6 proteins. Hela cells were transfected with the 6WtkCAT reporter plasmid and the different expression vectors as indicated except for lane 2 were the reporter plasmid is the 6FdtkCAT construct. Lane1 shows the basal level of expression of the 6WtkCAT construct (cotransfected with the empty expression vector pEVRF0). The structure and the relative transactivation potentials of the different Oct6 mutant proteins is outlined in figure 6. B. Bandshift assays with whole cell extracts of a fraction of Hela cells transfected with the different truncated Oct6 proteins.

bandshift assay using whole cell extracts of transfected Hela cells. As can be seen in Figure 4B all mutant proteins are expressed at a comparable high level except for POU/VP16. The fusion of the VP16 transactivation domain to the Oct6 POU domain appears to result in a labile protein or alternatively influences its DNA binding affinity. Nevertheless, this chimaeric protein is a strong transcriptional activator. As can be seen in Figure 4A high level expression of Oct6 results in 30-50 fold induction of the 6Wtk promoter (compare lane 1 and 3; the percentages conversion obtained in these experiments are given in Figure 6). This activation is dependent on an intact Octa motif within the 6W enhancer as this stimulation of CAT expression is not observed with the mutant version of this promoter (lane 2). Deleting the first 115 amino acids does not reduce the ability to activate the 6Wtk promoter. Although expression of mutants N23 and N44 resulted in a slightly lower transcactivation level, the mutants N97 and N115 showed wild type activity. Further deleting the Oct6 protein to amino acid 143 resulted in a drop of activity to approximately 30% of the wild type level. Deleting beyond a.a. 157 completely abolished transactivation. This maps a minimal region required for transactivation between amino acid 115 and 157, which is rich in glycine and alanine residues (19 Gly + Ala). However, this feature is not characteristic for this domain, as the entire Oct6 protein is extremely rich in glycine and alanine residues, apart from its POU domain. Therefor other features than just a high glycine and alanine content must be involved in mediating transactivation.

The Oct6 protein domain carboxyl terminal of the POU domain is rich in proline residues (13 out of 52). Proline rich transactivation domains have been described for a number of transcription factors. However no such function can be ascribed to this part of the Oct6 protein as the amino terminal deletion mutants N157 to N229 all contain this domain but fail to transactivate. Furthermore, the N2C52 deletion mutant in which the entire carboxyl terminal domain is removed shows a wild type level of transactivation. The POU domain by itself (N229C52) even seems to repress basal level of expression (compare lane 14 with lane1).

The experiments described above mapped a minimal transcriptional activation domain between amino acid 115 and 157. To check whether this domain is required and sufficient for transactivation or whether there is some functional redundancy within the amino terminal third of the protein, a small set of internal deletion mutants was constructed and tested for its ability to transactivate the 6Wtk promoter. The results of this experiment is shown in Figure 5A, while the internal deletion mutants are schematically depicted in Figure 6. Surprisingly all internal deletion mutants transactivate the 6Wtk promoter albeit at a lower level than the full length Oct6 protein. The level of the Δ94/232 and Δ116/230 proteins is lower than that of the other deletion mutants. This was consistently observed in a number of transfection experiments, which might indicate a lower stability of these mutant proteins

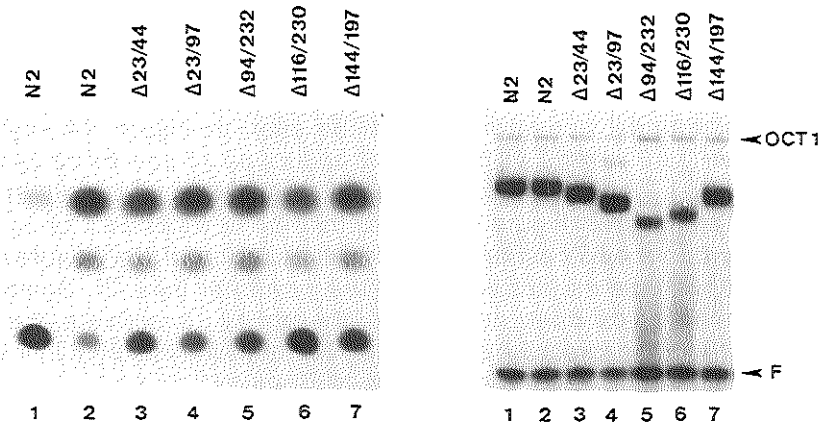


Figure 5. A. Transactivation of the 6Wtk promoter by Oct6 internal deletion mutant proteins. Hela cells were cotransfected with the 6WtkCAT reporter plasmid and the different expression plasmid as indicated except for lane 1 were the reporter is the 6FdtkCAT plasmid. The structure and the relative transactivation potential of the different Oct6 proteins is depicted in figure 6. B. Bandshift assays with whole cell extracts of Hela cells transfected with the different Oct6 mutant proteins.

(note the smear in lanes 5 and 6 in Figure 5, right panel) leading to underestimating the transactivation potential of the mutant proteins. As expected, internal deletions in the first 98 amino acids (Δ23/44; lane 3 and Δ23/97; lane 4) did not abolish transactivation as these proteins retain the domain between a.a. 115 and 157. However, neither deletion of aa 95 to aa 231 (lane 5) nor smaller deletions in this region (Δ116/230; lane6 and Δ144/197; lane 7) resulted in a complete loss of transactivation.

Thus, the combined analysis of the N terminal and internal deletion mutants suggest that the entire amino terminal third of the protein is involved in transactivation and can be split up in functionally redundant subdomains.

Interestingly, this amino terminal region contains a long stretch of alanine residues (9) that is polymorphic. The Oct6 cDNA cloned by Suzuki *et al.* encodes an Oct6 protein with a stretch of 8 ala residues and the rat Oct6 homologue SCIP has 11 ala residues at this position [13, 18]. The alanine stretch is located in a large region that is likely to adopt an α helical structure.

DISCUSSION

The six times repeated IgH enhancer octamer element (6W) was shown to strongly enhance transcription from the linked HSV tk promoter in ES and F9 EC cells. These cells express three octamer binding factors; Oct1, Oct4 and

Oct6 [24]. The Oct4 and Oct6 factors are downregulated upon differentiation. At the same time the 6W enhancer is extinghuised in these differentiating cells, implicating either of the two or both factors in enhancer function. However ectopic expression of Oct4 in Hela cells failed to activate the tk promoter via the 6W enhancer. In contrast with the results presented here, Suzuki *et al.* reported that also Oct6 is not able to activate the 6W enhancer in Hela cells whereas it activates a simple OCTA/TATA promoter (this promoter is similar to the pβO$^+$ promoter used here; Figure 2;[13]). It is unlikely that this discrepancy is due to intrinsic differences in the cell line used, as we show that high level expression of Oct6 results in the activation of the 6Wtk promoter in different cell lines. More likely these differences are due to differences in the levels at which the effector protein is expressed. It is possible that only high levels of Oct6 effectively compete for binding with the endogenous Oct1 protein and stimulates transcription in Hela cells. Thus our results suggest that the Oct6 protein could play a role in stimulation of transcription of several gene promoters in EC/ES cells.

One of the aims of this study was to localize the domain(s) in the Oct6 protein involved in the transactivation function of this protein. First we tested the responsiveness of three different promoters to high levels of Oct6. The three promoters used are different with respect to the number and the exact sequence of Oct6 binding sites, the distance of these sites to the TATA box, and the nature, number and position of other cis-acting elements (see Figure 1). All three promoters can be activated by coexpression of Oct6. Thus in our test system Oct6 can activate transcription from octamer and octamer related sequences in different promoter/enhancer settings. To determine which part of the Oct6 protein is involved in this function we constructed a series of deletion mutants of the Oct6 protein that were tested for their capicity to stimulate transcription from the 6Wtk promoter. All deletion mutants have an intact POU domain and thus retain the ability to bind the octamer motif. The data from the Oct6 deletion analysis indicate that the transactivation function is located in the amino terminal third of the protein. The entire activation domain is extremely rich in glycine and alanine residues and is composed of multiple, functionally redundant, subdomains.

Protein domains have been determined, involved in the transactivation function of a large number of transcription factors. There are no apparent structural similarities between these domains, although they can be roughly classified as being particularly rich in certain amino acids (see [31] and references therein). As far as data are available, the transactivation domains of the POU proteins are either glutamine- (Oct1) glutamine/leucine/proline- (Oct2), proline- (Oct3/Oct4) or serine/threonine-rich (Pit1/GHF1) [9, 16, 30, 32-35]. The transactivation domain of the Oct6 protein would form a new class, being glycine- and alanine-rich.

Interestingly, the proline rich region carboxyl terminal of the POU domain is dispensible for Oct6 to activate the 6Wtk promoter, suggesting that a high

concentration of a particular amino acid residue is not enough to make up an activation domain. This is also illustrated by the fact that the region between aa 157 and 229 of the Oct6 protein is also rich in glycine and alanine residues but does not activate transcription. A direct approach to address

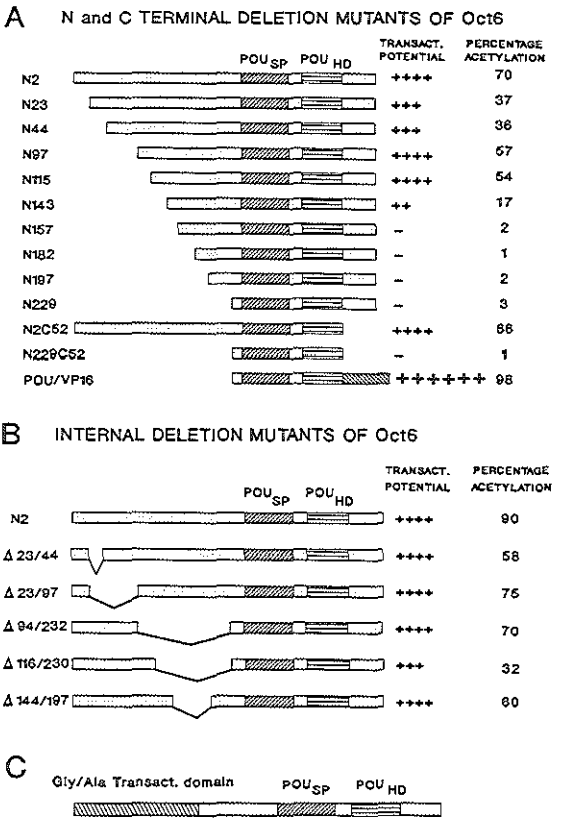


Figure 6. Structure and transactivation potential of the Oct6 deletion mutants used in this study. The naming of the different mutants reflects the number of codons removed from the Oct6 ORF either from the aminoterminus (N) or the carboxyterminus (C). The construction of the different expression plasmids is detailed in Materials and Methods. The N2 Oct6 protein is considered to be the wild type protein. The POU-specific and the POU-homeodomain are indicated by shaded boxes. The relative transcriptional activation of the 6Wtk promoter by the different mutant proteins is indicated. The percentage acetylation refers to the percentage of acetylated forms of chloramphenicol in the experiments presented in figure 4 and 5. CAT activity in extracts of the 6WtkCAT transfected cells is 2% (lanes 1 in figure 4 and 5). A. Structure of the N- and C-terminal Oct6 mutant proteins. B. Structure of internal deletion mutant Oct6 proteins. The portion of the Oct6 protein deleted is indicated by the number of the codons at which the ORF is fused. C. Structure of the Oct6 protein, highlighting the domain involved in transcriptional activation and DNA binding (the POU domain; POU_S and POU_{HD}).

this issue for the Oct2 protein has been reported by Gerster *et al.*[35]. The amino terminal transactivation domain of the Oct2 factor is characterised by a high content of glutamine, leucine and proline residues. The analysis of substitution mutants of subdomain I in which the glutamine residues were replaced by asparagine or the leucine residues by isoleucine showed a drastic reduction in transactivation potential. Thus the particular arrangement and interaction with other residues are critical for the activation function of this domain.

The transactivation domains of the Oct6 and Oct4 factors belong to different classes (being glyc/ala rich and pro rich respectively). This might indicate that Oct4 and Oct6 interact with different components of the basic transcription apparatus or alternatively that they interact with different socalled coactivators associated with the TATA box binding protein(TBP;[36] see also [37]. Thus taking into account the interaction with other DNA binding proteins and the differences in affinity between the two proteins for different octamer binding sites it is likely that the two octamer factors regulate different sets of genes in ES/EC cells.

The importance of interactions between octamer factors and other transcription factors is illustrated by the observation that the octamer motif can also mediate repression of transcription in F9 EC cells [38]. Obviously the sequence context of a cis-acting element is highly critical. Several mechanisms can be envisaged by which transcription factors can function as repressors (see [39]), one of which is competetive binding. For instance the SV40 B element consist of two directly repeated Sph motifs of which the junction forms an octamer binding site. Both Sph motifs are required for enhancer function in Hela cells, while high level expression of Oct1 or Oct2 represses the activity of the Sph motif via competetive binding [33].

The rat homologue of Oct6, SCIP is expressed at high levels in proliferating Schwann cells [17]. The expression of SCIP/Oct6 seems to antagonize the expression of glial specific genes like MBP and P_0 [18]. Transfection of a SCIP/Oct6 expression vector into cultured Schwann cells results in the downregulation of cotransfected MBP, P_0 and NGF-R promoters. These results indicate that the SCIP/Oct6 protein serves as a negative regulator of Schwann cell specific genes. Significantly downregulation of the P_0 promoter by SCIP was also observed when cotransfected in the monkey cell line CV-1, indicating that no Schwann cell specific components are needed to downregulate this promoter [20]. The P_0 promoter contains five SCIP/Oct6 binding sites that are remotely related to the octamer motif. One of the weaker binding sites overlaps with the P_0 TATA box. It was suggested that this binding sites serves as a negative element for SCIP/Oct6 regulation of P_0 gene transcription through a competition mechanism [20]. It is not known whether the POU domain of SCIP/Oct6 suffices for repression of the P_0 promoter or that in addition other domains of the protein are required for this function. It is of interest to note

here that a domain capable of mediating repression in Hela and CV-1 cells has been identified in the *Drosophila* Krüppel protein. It was shown that a N-terminal alanine rich domain of this protein fused to the DNA binding domain of the lac repressor can repress transcription of target genes containing lac operator sequences [40]. Such an alanine rich region can also be found in the *Drosophila* transcriptional repressors engrailed and evenskipped. Similarly, alanine rich domains are present in the Oct6 protein interwoven with the transcriptional activation domains as determined in this study. It is possible that these regions in the Oct6 protein are also involved in mediating repression of glial cell specific genes.

Acknowledgement

We would like to thank Drs P. O'Hare and H. Schöler for the different reporter plasmid DNAs and Sandro Rusconi and Katja Seipel for the CMV expression vectors and the Oct2A clone. Frank Grosveld for criticism, Judith Boer for constructing the POU/VP16 clone and Fried Zwartkruis for his help with the Phosphorimager, Sjozef van Baal for having a solution to every computer problem and Professor Dirk Bootsma for continuous support.

MATERIALS AND METHODS

Construction of Oct6 deletion mutants.

Deletions of the Oct6 cDNA were created using naturally occuring restriction sites. The different deletion constructs were cloned in frame into the CMV based expression vectors pEVRF0, pEVRF1, pEVRF2 or pEV3S [26]. These vectors allow expression of N terminally deleted proteins as an in frame fusion product with the first four amino acids of the HSV tk protein. Cloning of the full length Oct6 cDNA in pEVRF0 using the Ball restriction site in effect results in a protein that has 7 amino acid residues inserted between aa 2 and aa3 as compared to the wildtype protein. The first two aa encoded by the tk reading frame are the same as in the Oct6 reading frame. This expression plasmid is referred to as clone pN2Oct6 or Oct6 in Figure 2. The N stands for a N terminal truncation and the number following the N refers to the number of codons removed from the Oct6 open reading frame. The C stands for a carboxyl terminal truncated protein and the number following the C indicate the number of codons removed from the carboxyl terminus of the Oct6 protein. The translated linker sequence of the truncated clones is as follows: N2, MASWGSGTP; N23, MASWGSGTL; N44, MASWGSGTH; N97, MASWGSGVP; N115, MASWGSGTL; N143, MASWGSGVP; N157, MASWGSGTH; N182, MASWGSGT; N197, MASWGSGVP; N229, MASWGSGY. The chimaeric clone POU/VP16 was constructed by replacing the 3' part Oct6 cDNA from the SacII to the XbaI site (encoding the 51 aa carboxyl terminus of Oct6) in clone N229 for the 3' part of the HSV1 Vmw65 (VP16) gene. In this way the Oct6 POU domain is fused to the 80 aa carboxyl terminus of the VP16 protein. This part of the VP16 protein is a strong transactivator [29]. Internal deletion mutants were made by excising part of the Oct6 open reading frame in clone pN2Oct6. The numbers represent the codons at which the Oct6 ORF is fused. All fusion points were checked by dideoxy sequencing. All deletion constructs were tested by expression in COS cells. Using nuclear extracts of transfected COS cells in a bandshift assay indicated that all mutants were stable, bind to the octamer sequence with high affinity and were located in the nucleus.

Reporter plasmids

The reporter plasmids 6WtkCAT and 6FdtkCAT are described in detail in [22] and were a kind gift of Dr Hans Schöler. The βOCTA promoter was constructed by PCR amplification using a 5' sense oligo containing an OCTA box or a mutated version thereof and a 3' anti-sense oligo mapping in front of the rabbit β-globin start codon. A genomic subclone in pBR327 containing

the entire rabbit β-globin gene was used as a template. 5' sense oligo, GCGGATCCATTTGCAT-TACATAGTTCAGGACTTGG; 3' anti-sense oligo, GCGCTCGAGTCTGTTTTGGGGGATTGC The amplification products were cloned in front of the CAT gene in pBLCAT3 [23]. The integrity of the βO^+ and βO^- promoter constructs were checked by dideoxy sequencing. The full sequence of these promoters is as follows; GGATCCATTTGCATTACATAGTTCAGGACTTGGGCATAAAA-GGCAGAGCAGGGCAGCTGCTGCTTACA

The ICP4 CAT (IE175CAT) construct was a kind gift of Dr Peter O'Hare. This construct contains the IE175 (ICP4) promoter from -330 to +20 [25, 41].

cell transfections

Using the $CaPO_4$ precipetate method, Hela, MES or CHO cells were transfected with 5 μg of reporter plasmid and 2.5μg of Oct6 expression vector (or with 2.5μg of the empty expression vector pEVRF0) and 12.5 μg of pTZ19 plasmid DNA as carrier [27]. Transfected cells were harvested 48 hrs after removal of the $CaPO_4$ precipetate. Eighty percent of the cells were used to prepare cell extracts for a CAT assay and the remaining 20 % was used to prepare whole cell extracts for monotoring Oct6 protein expression in a bandshift assay. CAT assays were performed with equal amounts of protein (20 μg). Acetylated and non-acetylated forms of ^{14}C labeled chloramphenicol were seperated using standard thin layer chromatography. The ratio between acetylated and non-acetylated forms of chloramphenicol were calculated after quantitation of the signals using a Molecular Dynamics Phosphorimager.

Bandshift experiments and whole cell extracts

Whole cell extracts of transfected Hela cells (20%; see above) were prepared by resuspending the cells in 40 μl of 20mM Hepes-KOH pH7.9, 400mM KCl, 1mM EDTA, 10% glycerol, 10mM DTT, 1mM PMSF, 5 μg/ml leupeptin, 5μg/ml pepstatin and 10μg/ml chymostatin.Resuspended cells were subject to 4 cycles of freezing in liquid nitrogen and thawing on ice. The cellular debris was removed by centrifugation at 14000g for 5 minutes at 4°C. Equal amounts of cellular extract were used in a bandshift assay using 3 fmol of a ^{32}P endlabelled double stranded OCTA probe. Probe and protein were incubated on ice for 20 minutes in 20 mM Hepes-KOH ph7.9, 1mM EDTA, 1 mM EGTA, 4% Ficoll in a total volume of 20 μl. Complexed and free probe were seperated on a 4% polyacrylamide gel in 0.25 x TBE.

REFERENCES

1. **Kemler, I. and W. Schaffner.** 1990. Octamer transcription factors and the cell type-specificity of immunoglobulin gene expression. *Faseb J* **4:** 1444-9
2. **La Thangue, N.B. and P.W.J. Rigby.** 1988. Trans-acting protein factors and the regulation of eukaryotic transcription. in "Transcription and Splicing", B. D. Hames and D. M. Glover eds. IRL Press, Oxford
3. **Schaffner, W.** 1989. How do different transcription factors binding the same DNA sequence sort out their jobs? *Trends Genet* **5:** 37-9
4. **Schöler, H.R., A.K. Hatzopoulos, R. Balling, N. Suzuki and P. Gruss.** 1989. A family of octamer-specific proteins present during mouse embryogenesis: evidence for germline-specific expression of an Oct factor. *Embo J* **8:** 2543-50
5. **Schreiber, E., K. Harshman, I. Kemler, U. Malipiero, W. Schaffner and A. Fontana.** 1990. Astrocytes and glioblastoma cells express novel octamer-DNA binding proteins distinct from the ubiquitous Oct-1 and B cell type Oct-2 proteins. *Nucleic Acids Res* **18:** 5495-503
6. **Clerc, R.G., L.M. Corcoran, J.H. LeBowitz, D. Baltimore and P.A. Sharp.** 1988. The B-cell-specific Oct-2 protein contains POU box- and homeo box-type domains. *Genes Dev* **2:** 1570-81
7. **Meijer, D., A. Graus, R. Kraay, A. Langeveld, M.P. Mulder and G. Grosveld.** 1990. The octamer binding factor Oct6: cDNA cloning and expression in early embryonic cells. *Nucleic Acids Res* **18:** 7357-65
8. **Müller, M.M., S. Ruppert, W. Schaffner and P. Matthias.** 1988. A cloned octamer

transcription factor stimulates transcription from lymphoid-specific promoters in non-B cells. *Nature* **336**: 544-51

9. **Okamoto, K., H. Okazawa, A. Okuda, M. Sakai, M. Muramatsu and H. Hamada.** 1990. A novel octamer binding transcription factor is differentially expressed in mouse embryonic cells. *Cell* **60**: 461-72
10. **Schöler, H.R., S. Ruppert, N. Suzuki, K. Chowdhury and P. Gruss.** 1990. New type of POU domain in germ line-specific protein Oct-4. *Nature* **344**: 435-9
11. **Rosner, M.H., M.A. Vigano, K. Ozato, P.M. Timmons, F. Poirier, P.W. Rigby and L.M. Staudt.** 1990. A POU-domain transcription factor in early stem cells and germ cells of the mammalian embryo. *Nature* **345**: 686-92
12. **Sturm, R.A. and W. Herr.** 1988. The POU domain is a bipartite DNA-binding structure. *Nature* **336**: 601-4
13. **Suzuki, N., H. Rohdewohld, T. Neuman, P. Gruss and H.R. Schöler.** 1990. Oct-6: a POU transcription factor expressed in embryonal stem cells and in the developing brain. *Embo J* **9**: 3723-32
14. **Herr, W., R.A. Sturm, R.G. Clerc, L.M. Corcoran, D. Baltimore, P.A. Sharp, H.A. Ingraham, M.G. Rosenfeld, M. Finney, G. Ruvkun and H.R. Horvitz.** 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**: 1513-6
15. **Verrijzer, C.P., A.J. Kal and P.C. van der Vliet.** 1990. The oct-1 homeo domain contacts only part of the octamer sequence and full oct-1 DNA-binding activity requires the POU-specific domain. *Genes Dev* **4**: 1964-74
16. **Ingraham, H.A., S.E. Flynn, J.W. Voss, V.R. Albert, M.S. Kapiloff, L. Wilson and M.G. Rosenfeld.** 1990. The POU-specific domain of Pit-1 is essential for sequence-specific, high affinity DNA binding and DNA-dependent Pit-1-Pit-1 interactions. *Cell* **61**: 1021-33
17. **Monuki, E.S., G. Weinmaster, R. Kuhn and G. Lemke.** 1989. SCIP: a glial POU domain gene regulated by cyclic AMP. *Neuron* **3**: 783-93
18. **Monuki, E.S., R. Kuhn, G. Weinmaster, B.D. Trapp and G. Lemke.** 1990. Expression and activity of the POU transcription factor SCIP. *Science* **249**: 1300-3
19. **He, X., M.N. Treacy, D.M. Simmons, H.A. Ingraham, L.W. Swanson and M.G. Rosenfeld.** 1989. Expression of a large family of POU-domain regulatory genes in mammalian brain development. *Nature* **340**: 35-41
20. **He, X., R. Gerrero, D.M. Simmons, R.E. Park, C.J. Lin, L.W. Swanson and M.G. Rosenfeld.** 1991. Tst-1, a member of the POU domain gene family, binds the promoter of the gene encoding the cell surface adhesion molecule Po. *Mol Cell Biol* **11**: 1739-44
21. **Wirth, T., L. Staudt and D. Baltimore.** 1987. An octamer oligonucleotide upstream of a TATA motif is sufficient for lymphoid-specific promoter activity. *Nature* **329**: 174-8
22. **Schöler, H.R., G.R. Dressler, R. Balling, H. Rohdewohld and P. Gruss.** 1990. Oct-4: a germline-specific transcription factor mapping to the mouse t-complex. *Embo J* **9**: 2185-95
23. **Luckow, B. and G. Schutz.** 1987. CAT constructions with multiple unique restriction sites for the functional analysis of eukaryotic promoters and regulatory elements. *Nucleic Acids Res* **15**:
24. **Schöler, H.R., R. Balling, A.K. Hatzopoulos, N. Suzuki and P. Gruss.** 1989. Octamer binding proteins confer transcriptional activity in early mouse embryogenesis. *Embo J* **8**: 2551-7
25. **Mackem, S. and B. Roizman.** 1982. Differentiation between alpha promoter and regulator regions of herpes simplex virus 1: The functional domains and sequence of a movable alpha regulator. *Proc. Natl. Acad. Sci. USA* **79**: 4917-4921
26. **Matthias, P., M.M. Müller, E. Schreiber, S. Rusconi and W. Schaffner.** 1989. Eukaryotic expression vectors for the analysis of mutant proteins. *Nucleic Acids Res* **17**:
27. **Gorman, C.** 1985. High efficiency gene transfer into mammalian cells. in "DNA Cloning ", D. M. Glover IRL Press, Oxford
28. **Sadowski, I., J. Ma, S. Triezenberg and M. Ptashne.** 1988. GAL4-VP16 is an unusually

potent transcriptional activator. *Nature* **335**: 563-4
29. **Triezenberg, S.J., R.C. Kingsbury and S.L. McKnight.** 1988. Functional dissection of VP16, the trans-activator of herpes simplex virus immediate early gene expression. *Genes Dev* **2**: 718-29
30. **Müller, I.M., W. Schaffner and P. Matthias.** 1990. Transcription factor Oct-2A contains functionally redundant activating domains and works selectively from a promoter but not from a remote enhancer position in non-lymphoid (HeLa) cells. *Embo J* **9**: 1625-34
31. **Mitchell, P.J. and R. Tjian.** 1989. Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. *Science* **245**: 371-8
32. **Theill, L.E., J.L. Castrillo, D. Wu and M. Karin.** 1989. Dissection of functional domains of the pituitary-specific transcription factor GHF-1. *Nature* **342**: 945-8
33. **Tanaka, M. and W. Herr.** 1990. Differential transcriptional activation by Oct-1 and Oct-2: interdependent activation domains induce Oct-2 phosphorylation. *Cell* **60**: 375-86
34. **Imagawa, M., A. Miyamoto, M. Shirakawa, H. Hamada and M. Muramatsu.** 1991. Stringent integrity requirements for both trans-activation and DNA-binding in a trans-activator, Oct3. *Nucleic Acids Res* **19**: 4503-8
35. **Gerster, T., C.G. Balmaceda and R.G. Roeder.** 1990. The cell type-specific octamer transcription factor OTF-2 has two domains required for the activation of transcription. *Embo J* **9**: 1635-43
36. **Dynlacht, B.D., T. Hoey and R. Tjian.** 1991. Isolation of coactivators associated with the TATA-binding protein that mediate transcriptional activation. *Cell* **66**: 563-76
37. **Lewin, B.** 1990. Commitment and activation at pol II promoters: a tail of protein-protein interactions. *Cell* **61**: 1161-4
38. **Lenardo, M.J., L. Staudt, P. Robbins, A. Kuang, R.C. Mulligan and D. Baltimore.** 1989. Repression of the IgH enhancer in teratocarcinoma cells associated with a novel octamer factor. *Science* **243**: 544-6
39. **Levine, M. and J.L. Manley.** 1989. Transcriptional repression of eukaryotic promoters. *Cell* **59**: 405-8
40. **Licht, J.D., M.J. Grossel, J. Figge and U.M. Hansen.** 1990. *Drosophila* Kruppel protein is a transcriptional repressor. *Nature* **346**: 76-9
41. **O'Hare, P., C.R. Goding and A. Haigh.** 1988. Direct combinatorial interaction between a herpes simplex virus regulatory protein and a cellular octamer-binding factor mediates specific induction of virus immediate-early gene expression. *Embo J* **7**: 4231-8

Chapter VIII

The mouse Oct6 gene and its regulatory sequences.

Dies Meijer, Robert Kraay, Anneke Graus and Gerard Grosveld.
MGC, Department of Cell Biology and Genetics, Erasmus University Rotterdam The Netherlands

ABSTRACT

The gene encoding the POU factor Oct6 is expressed in embryonic stem cells, during early development of the nervous system, in a subset of neuronal cells and in glial cells of the central and peripheral nervous system. Here we describe the genomic organization of the Oct6 gene and a preliminary characterization of the promoter of the gene. The Oct6 gene is a single exon gene that generates a 3 kb mRNA. The promoter and the entire protein coding part of the gene are located in a CpG island.The chromatin structure of the Oct6 gene promoter region is DNAaseI hypersensitive in ES cells, but not in non-expressing MES cells. The Oct6 promoter is a typical cell type specific promoter in that it contains a TATA and CCAAT box at canonical distances from the transcriptional initiation site. From transient transfection experiments it is tentatively concluded that the promoter region does not contain all the information for cell type specific expression but that additional downstream sequences are required for this.

INTRODUCTION

The POU domain gene family is defined by a bipartite DNA binding domain which is composed of a class specific homeo box and a POU specific domain ([1] and references therein). Members of this class of proteins have been implicated in cell fate determination through selective regulation of target genes [2, 3, 4]. For example the mouse Pit-1 gene has been shown to be mutated in two different alleles of the mouse dwarf locus [5]. These mutations interrupt the normal development of the anterior pituitary gland which results in the loss of expression of growth hormone, prolactin and thyroid-stimulating hormone and hypoplasia of their respective cell types. Other members of the POU domain family for which no genetic data are available are supposed to be involved in cellular differentiation processes mainly on the basis of a restricted expression pattern during development and by association. A number of POU proteins bind with high affinity to the octamer sequence (ATTTGCAT), which constitutes a strong enhancer element in embryonic stem cells and B lymphocytes [6, 7]. One of these POU proteins, Oct6; also called SCIP and Tst-1, is expressed in undifferentiated embryonic stem cells, in early embryonic and adult neuronal cells and in glia cells of both the central as well the peripheral nervous system [8-12]. In Schwann cells the Oct6 gene is highly expressed during a phase of rapid cell division that separates a premyelinating from a myelinating state [13]. Expression of the Oct6 gene correlates with the

negative regulation of a number of Schwann cell specific genes. Indeed cotransfection analysis has shown that the Oct6 protein can repress expression from the P_0 promoter in non glial cells, while other promoters can be activated [12, 14, 15]. Thus the protein can either act as a positive or negative regulator of gene expression. An important question is how these cell type specific transcription factors themselves are regulated during development and homeostasis. As a first step towards understanding the regulation of the Oct6 gene we describe here the genomic organization of the gene and its promoter. The Oct6 gene is a single copy, single exon gene in the mouse genome. The protein coding part and the promoter are part of a CpG island. The promoter region of the gene is associated with a DNAasel hypersensitive chromatin structure in ES cells but not in the non-expressing MES cells. Transient transfection assays using 5' truncated promoter deletion mutants driving the bacterial CAT reporter gene identified an important element upstream of the CCAAT box. However these promoter constructs were equally active in cells expressing or not expressing Oct6, suggesting that additional regulatory elements are missing from these promoter constructs.

RESULTS

Cloning of the mouse Oct6 gene.

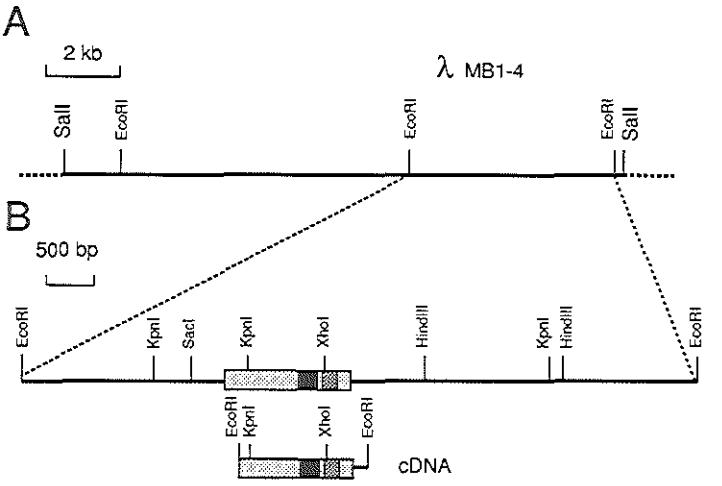


Figure 1. Genomic map of the mouse Oct6 gene.
A. Schematic representation of the one lambda clone out of ten that was further analysed. This clone contained a 18 kb genomic fragment generated by partial Sau3A digestion of total genomic mouse DNA cloned in the BamHI site of the λEMBL3 phage. The SalI sites are part of the phage and flank the BamHI cloning site.
B Detailed restriction map of the 6.4kb EcoRI fragment derived from λMB1-4. The cDNA sequence is aligned with the genomic clone to show that they are colinear over the entire length of the cDNA. The EcoRI sites in the Oct6 cDNA represent artificial cloning sites.

Clones were isolated by screening of an EMBL3 mouse genomic library at high stringency with an Oct6 cDNA probe (300bp PvuII fragment covering the POU domain). Ten strongly hybridizing clones were plaque purified in a second screening. Two of these clones were further analyzed by restriction mapping (Figure 1A). Hybridization of restriction fragments of the larger of the two genomic clones with a 1.4kb cDNA clone (covering 95% of the Oct6 open reading frame) revealed that the entire cDNA sequence was present on a 6.3kb EcoRI fragment and is not interrupted (data not shown, see Figure 1B). This fragment was subcloned in the plasmidvector pTZ18. The size of this EcoRI fragment corresponds with that of the Oct6 cDNA hybridization signal on genomic Southern blots, indicating that we have isolated the Oct6 gene and not a pseudogene or other closely related POU domain gene (see Figure 4).

The 6.3kb EcoRI fragment contains all of the Oct6 coding sequences.
The Oct6 gene expresses a 3 kb mRNA in ES cells, brain and Schwann cells. As the longest cDNA molecule we isolated is 1.4 kb, additional coding information must be present in the 6.3kb EcoRI fragment or flanking fragments. Previously we have shown that the 5' part of the open reading frame plus 5 ' UTR is contained within a 450bp SacI/KpnI fragment

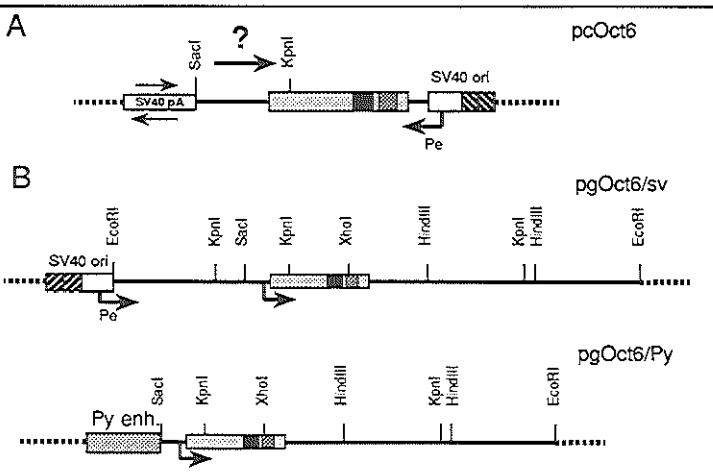


Figure 2. Schematic representation of Oct6 expression vectors
A. pcOct6. The Oct6 cDNA (see figure 1) was extended at its 5' end by linking the genomic SacI/KpnI fragment to the KpnI site. This extended cDNA was cloned as a SacI fragment in the pcDX vector in the opposite orientation with respect to the SV40 early promoter (Pe)/origin of replication (SV40 ori). Transcription termination and polyadenylation signals on both strands are derived from the SV40 genome (SV40 pA).
B. pgOct6/sv and pgOct6/Py. The genomic 6.4 kb EcoRI fragment was cloned downstream of the SV40 promoter/ori fragment that is derived from pcDX. No SV40 termination and polyadenylation sequences are present. The pgOct6/Py clone was constructed by linking the genomic SacI/EcoRI fragment to the polyoma virus PyF441 enhancer fragment. The enhancer fragment is derived from the pMC1neoPolyA plasmid [31].

immediately flanking the cDNA homologous region. To prove that this fragment also contains the minimal promoter region of the Oct6 gene we designed a functional assay. The 1.4kb Oct6 cDNA (mb1 in chapter VI) was extended by fusing the cDNA with the genomic fragment on the KpnI site, and this construct was cloned in the SV40 based expression vector pcDX ([16], called pcOct6 in Figure 2 and 3). The orientation of the Oct6 insert is antisense with respect to the SV40 early promoter. Thus expression of a functional Oct6 protein only occurs when the SacI/KpnI fragment contains promoter activity (see Figure 2 for details). Two transcripts are expected to be generated; one antisense RNA of 1.9kb transcribed from the SV40 promoter and one sense transcript of approximately 5kb transcribed from the putative Oct6 promoter. The pcOct6 construct was transfected into COS cells and expression was monitored in a gel retardation assay and by northern blotting (Figure 3A and B). As can be seen in Figure 3A, lane 1, transfected COS cells express high levels of Oct6 indicating that the SacI/KpnI fragment indeed contains promoter activity. Northern blot analysis using an Oct6 cDNA fragment as probe revealed two major and one minor

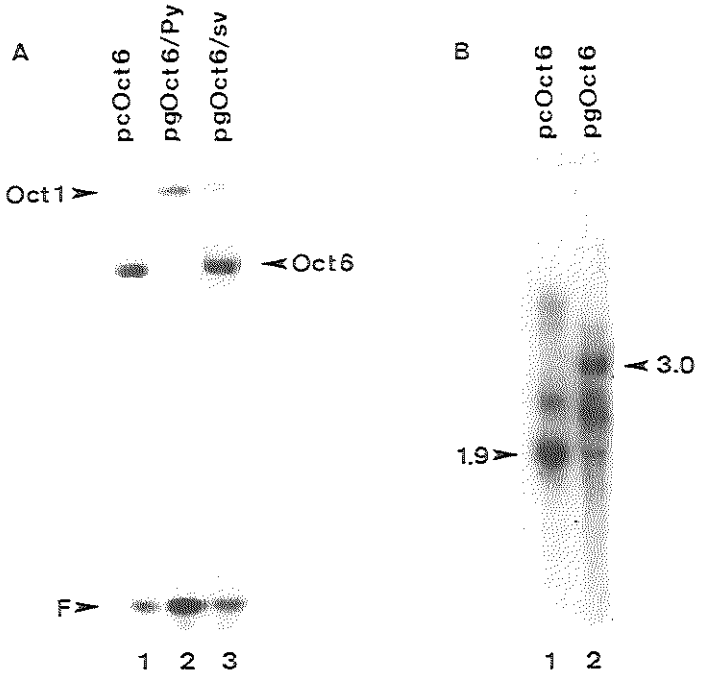


Figure 3. Bandshift and Northern blot analysis. Nuclear proteins and total cytoplasmic RNA was extracted from transfected COS cells. The presence of the Oct6 protein in nuclear extracts was determined in a bandshift asssay using the radiolabelled OCTA probe (see chapter V). The Oct6 mRNA is detected with the 200bp Oct6 PvuII probe (see chapter IV).

RNA species. As outlined above, the abundant transcript of 1.9kb most likely corresponds to the Oct6 antisense RNA generated from the SV40 early promoter. The low abundance 4.8kb transcript probably corresponds to the sense Oct6 RNA transcribed from the Oct6 promoter. The origin of the intermediately sized RNA is not clear. Thus despite the high level of antisense RNA the sense mRNA seems to be translated efficiently enough to generate high levels of Oct6 protein. To exclude the possibility that sense transcripts are generated from a cryptic promoter within the SacI/KpnI fragment we performed primer extension experiments on RNA from transfected COS cells and from an Oct6 expressing cell line (MES68; data not shown). Extension products of exactly the same length were observed in both cells indicating that transcription in the transfected COS cells is initiated from the correct CAP site.

To test whether the 3' end of the gene is also present on the 6.3kb EcoRI fragment a functional assay was used, similar to the one described above. The generation of stable RNA requires proper termination and polyadenylation of primary transcripts. We asked whether the 6.3kb fragment contained all the neccassry sequences by introducing the gene in COS cells and monitoring expression by gel retardation assay and northern blots. Two constructs were made (Figure 2). The 6.3 EcoRI fragment was cloned in a nonreplicating (in mammalian cells) vector and the fragment was flanked by a polyoma virus enhancer element to boost transcription from the Oct6 promoter (pgOct6/Py). A second construct was made by cloning the 6.3 kb fragment 3' of the SV40 early promoter/ori (pgOct6sv). This plasmid will replicate in COS cells. Both constructs were transfected in COS cells. Bandshift analysis of transfected cells showed that only pgOct6/sv expresses high levels of Oct6 protein. It has been shown previously by Yamaguchi and Matsukage [17] that DNA replication can overrule cis-acting silencing elements. The Oct6 gene is normally not expressed in COS cells. Therefore, the most likely explanation for the fact that pgOct6/sv expresses Oct6 and pgOct6/Py not, is that only pgOct6/sv replicates in COS cells. Northern analysis of pgOct6/sv transfected COS cells showes that the plasmid expresses an abundant Oct6 transcript of 3kb and several shorter RNAs (Figure 3B). Probably the shorter RNAs result from the use of cryptic polyadenylation sites. As we used total RNA we do not know whether all the detected RNAs are actually polyadenylated. Thus the pgOct6/sv clone expresses an Oct6 mRNA of correct size and a functional DNA binding protein, suggesting that the entire Oct6 gene is located on the 6.3kb genomic EcoRI fragment. Since it is unlikely that the 3'UTR of the gene is separated in more exons it is highly likely that the Oct6 gene is a single exon gene. Recently the rat homologue of the Oct6 gene, SCIP, has been described and it was shown to be a single exon gene [20].

The promoter region of the Oct6 gene contains DNAseI hypersensitive sites in expressing cells.

The chromatin stucture of promoter and enhancer regions of actively transcribed genes is often hypersensitive for nucleases, while the entire transcribed region and beyond shows an increased generalized sensitivity when compared with non-transcribed genes ([18, 19] for review). This hypersensitivity reflects the accessibility of the DNA *in vivo* for nucleases and other molecules like transcription factors. Thus the presence of hypersensitive sites within a region of interest is taken as a first indication of the presence of important cis-acting elements. To further delineate the elements involved in the regulation of the Oct6 gene, we examined the presence of DNAseI hypersensitive sites within the 6.3kb genomic EcoRI

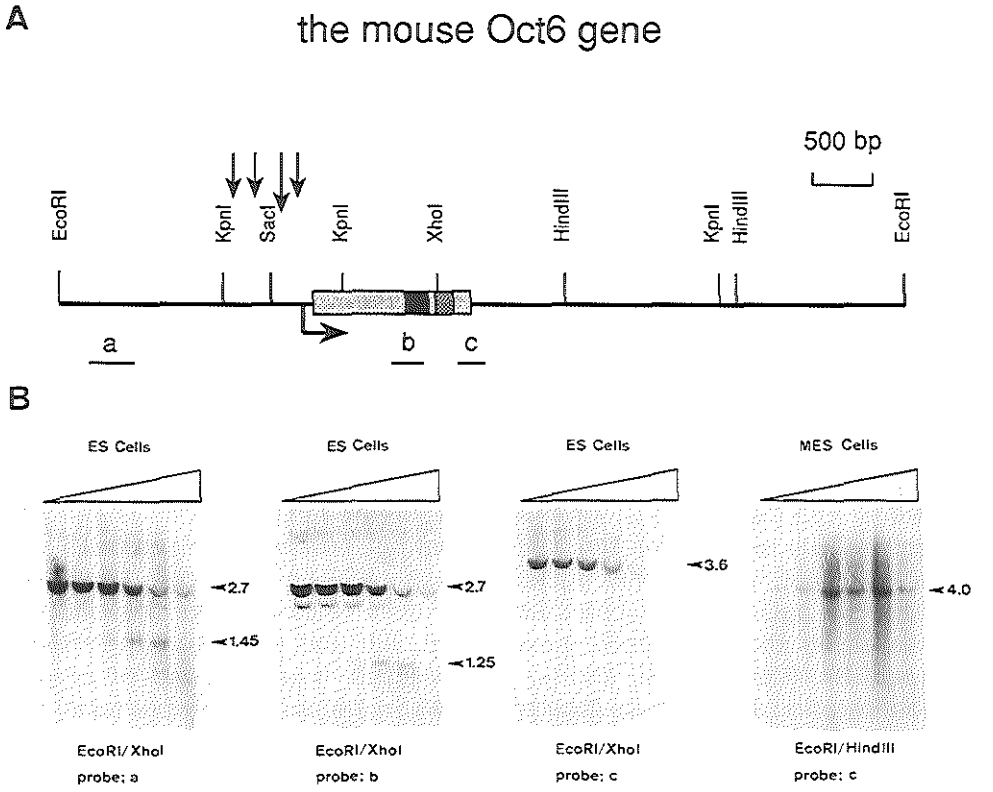


Figure 4. Mapping DNAaseI hypersensitive sites within the mouse Oct6 gene. Panel A shows the genomic region covered by the probes used (a, b and c). The DNAaseI hypersensitive sites are indicated with vertical arrows. The longest arrow indicates the position of the major hypersensitive site (see panel B). DNAaseI fade outs. Southern blots of XhoI/EcoRI restricted DNA isolated from DNAaseI treated ES cell nuclei were probed with the probes indicated in panel A. MES cell DNA was digested with EcoRI and HindIII and probed with probe c.

fragment in Oct6 expressing cells (ES cells) and non-expressing cells (MES cells). Nuclei of ES or MES cells were incubated with various amounts of DNAseI for a short period of time (see M&M). DNA was extracted from the nuclei, digested with restriction enzymes and separated on agarose gels. Southern blots of these gels were probed with various DNA fragments covering the 6.3 EcoRI fragment (Figure 4). The position of the probe fragments on the Oct6 map is indicated in panel A. Probe a detects an EcoRI/XhoI DNA fragment of 2.7kb that disappears with increasing DNAseI concentration. One major band and two smaller bands of lower intensity appear and disappear again. The major band has an estimated size of 1.45kb. Fragment b detects the same 2.7kb EcoRI/XhoI fragment but probes it from the 3' side. A major band of 1.25kb and a minor band of 1.15kb appear with increasing DNAseI concentration. As the sum of the length of the DNAseI generated fragments is the length of the EcoRI/XhoI fragment these fragments result from one major DNAseI hypersensitive site that is located approximately 1.25kb 5' of the XhoI site and 150bp 5'of the transcription initiation site. The position of this site and that of the lower intensity bands are indicated in Figure 4. The position of these sites is only approximate (+ or - 25bp) because of the limitations in determining the length of restriction fragments in agarose gels. Fragment c detects the 3'3.5kb EcoRI/XhoI band. This band gradually disappears with increasing DNAseI concentration. No additional bands appear indicating that there are no DNAseI hypersensitive sites in this part of the Oct6 gene. The presence of the DNAseI hypersensitive site detected by the a and b fragments correlates with the active state of the Oct6 gene as this sites are not present in the non-expressing cell line MES. Thus DNAseI hypersensitive sites are only associated with the 5' region of the Oct6 gene suggesting that important regulatory elements are clustered in this region. However it should be stressed that the analysis presented here only covers the 6.3kb EcoRI fragment. It is possible that additional important regulatory elements are located further away from the gene.

Sequence of the Oct6 gene promoter region.

From the results presented thus far it is clear that promoter and putative regulatory elements are located within the 900bp KpnI fragment (see Figure 4). The nucleotide sequence of this fragment was determined on both strands. Part of this sequence is presented in Figure 5. The promoter sequence of the rat Oct6 homologue, SCIP, has recently been published and is highly homologous to the mouse Oct6 promoter sequence [20]. Nucleotide differences are indicated above the Oct6 promoter sequenc in Figure 5. Because of the high degree of conservation between the mouse and rat sequence the alignment is not very informative with respect to pinpointing sequences that might be important for transcription regulation of the Oct6 gene. The position of the transcription initiation site as determined by primer

extension is indicated with an arrow (Figure 5). The sequence of the Oct6 promoter region is extremely rich in the dinucleotide CpG, as is the entire 5′ region of the Oct6 coding sequence. Thus the gene is located in a CpG island of approximately 1kb. Indeed this region is undermethylated in expressing and non-expressing cells (data not shown). This is a remarkable finding as most CpG islands are associated with the promoter region of housekeeping genes [21]. Notable other exceptions are the α globin and the Thy1 gene [22, 23]. Interestingly the Oct6 promoter has all the hallmarks of a tissue specific promoter. At around -30bp with respect to the transcriptional initiation site there is a TATA-box like sequence and at -90bp there is a typical GCCAAT box. This latter sequence element is highly homologous to the CTF/NF1 binding site in the α globin promoter. The TATA box sequence, TTTAA, resembles the TATA element of the Schwann cell specific P_0 promoter [24]. A computer search identified further a number of additional putative transcription factor binding sites (several sites are indicated in Figure

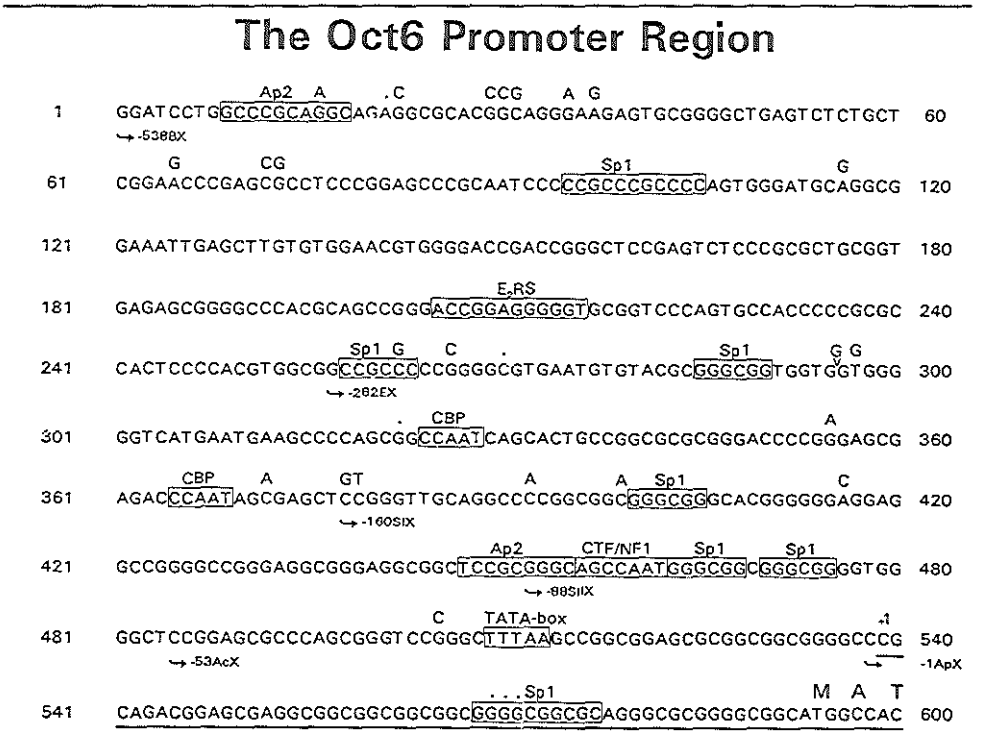


Figure 5. Nucleotide sequence of the Oct6 gene promoter.
The position of the transcriptional start site is indicated (+1). The transcribed region is underlined. Several putative cis acting element are boxed. The 5′ deletion endpoints of the promoter constructs shown in figure 7 are indicated. Nucleotides that are different in the rat SCIP/Oct6 promoter are indicated above the sequence, while dots indicate nucleotides not present in the SCIP/Oct6 sequence [20]

5 with the name of the factors that could bind to these sequences). Not unexpectedly there are at least eight binding sites for the factor Sp1 and several sequences that resemble recognition site for GCF, ETF, Krox 20 and Krox24 [25, 26, 27, 28]. Interestingly around -330 there is a sequence that perfectly matches the concensus binding site for the papilloma virus protein E2 (this region is strongly protected against limited DNAasel digestion by nuclear proteins from ES ad MES68 cells but not kidney). The DNAasel hypersensitive site as mapped in ES cell nuclei is located between -160 and -90. This region is extremely G + C rich and contains a sequence that is very similar to a region found in the NGF-R promoter, as noted earlier for the SCIP

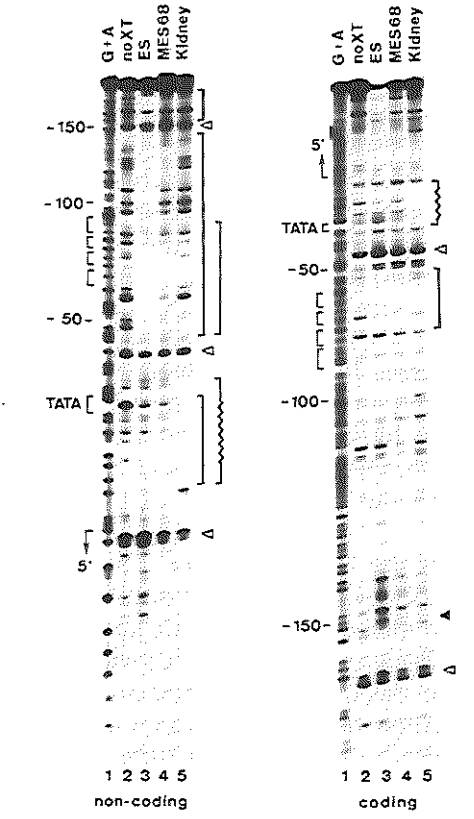


Figure 6. DNAasel footprinting of the Oct6 promoter region. Endlabelled probes were incubated with nuclear extracts as indicated and briefly exposed to DNAasel. Protected regions are indicated with brackets and the transcription initiation site with an arrow (5'). Pronounced DNAasel sensitive phosphodiester bonds are indicated with an open arrowhead.

promoter [20, 29]. The NGF-R gene is expressed in Schwann cells as is the SCIP/Oct6 gene. This sequence could therefore represent a common regulatory element within these promoters.

Footprint analysis

To get an impression of the DNA elements within the Oct6 promoter that interact with nuclear factors and to see whether we could detect differences between expressing and non-expressing cells, we performed DNAseI footprinting assays using probes covering the proximal promoter region up to -150nt. This region contains the major DNAseI hypersensitive site in ES cells. Three different crude nuclear extracts were used , ES cells, MES68 cells and kidney. ES and MES68 cells both express the Oct6 gene, while kidney cells do not express the gene. Probes were either labeled at the coding or noncoding strand. The results of these experiments are presented in Figure 6. The whole of the non-coding strand of the Oct6 promoter from the CAP site to >-150 is protected against limited degradation by DNAseI (lane 3, left panel, the protected regions are highlighted by brackets on the right hand side). Three DNAseI sensitive sites are not protected (indicated with an open triangle, left panel). Strong protection is seen between nt -40 and -150 using ES cell extracts and to a lesser extend in MES68 cells. The region from -90 to -150 seems not to be protected by kidney nuclear extracts. However a strong protected area on both strands, overlapping the TATA box, is observed with kidney extracts (lane 5 in left and right panel indicated with a curved line). Since kidney cells do not express the Oct6 gene, it is possible that this factor is involved in the repression of the gene. Thus the clearest differences in protection patterns between expressing and non expressing cells are found in the region upstream of the GCCAAT box at -90. Possibly these sequences are involved in the regulation of the Oct6 gene in ES and MES68 cells (see below).

Functional dissection of the Oct6 promoter

Expression of the Oct6 gene is restricted to undifferentiated embryonic stem cells, neuronal cells and glial cells of both the peripheral and central nervous system. Oct6 gene expression in ES cells is downregulated upon RA induced differentiation of these cells. To see whether the Oct6 promoter region contains all the information for cell type specific expression and regulation we tested several promoter deletion constructs by transient transfection of a limited number of cell lines. The cell lines we used are ES and MES68 cells both of which express Oct6 and MES cells which don't express the gene. The promoter deletion mutants, driving a CAT reporter gene, used in this set of experiments are depicted in Figure 7A. The activity of the different promoter deletion mutants was measured in a CAT assay using cell extracts of transiently transfected cells (Figure 7B). The longest promoter construct tested contains all of the DNAseI hypersensitive sites (-538BX in Figure 4

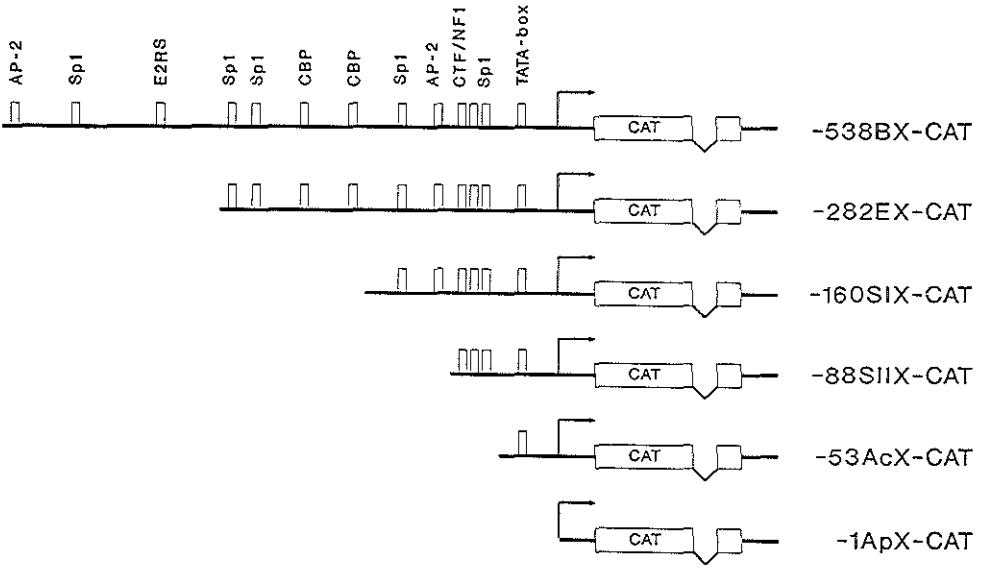
and 7). This promoter construct is active in ES cells. Activity is retained in the deletion mutants -282 and -160 but drops dramatically (5 fold) when the promoter is deleted to -88 and is completely lost when further deleted to -53. The region between -160 and -88 thus contributes significantly to the activity of the Oct6 promoter. This region also contains the major DNAseI hypersensitive site and is strongly protected against limited DNAseI digestion by ES cell nuclear proteins (Figure 4 and 6). The same set of mutants shows a similar activity in MES68 cells (Figure 7, panel D). However the same is observed for MES and differentiated ES cells. (panel B and C in Figure 7). Thus the Oct6 promoter constructs do not show tissue specificity and are not downregulated upon RA induced differentiation of ES cells suggesting that important regulatory elements are missing or that transient assays are not suitable for analyzing the Oct6 promoter (see discussion).

DISCUSSION

The data presented here strongly suggest that the mouse Oct6 gene is a single exon gene located on a 6.4 kb EcoRI fragment. Mapping of the Oct6 cDNA to this genomic EcoRI fragment indicates that the cDNA is colinear with the genomic DNA. The 5'UTR is 55nt long and is also not interrupted by introns. Thus the whole of the ORF and the 5'UTR are encoded by one exon. The sequence of the 1.5kb 3'UTR is not known as yet. However functional data presented here suggest that the 3'UTR maps to the 6.3 kb EcoRI fragment (Figure 3).This rather sketchy picture is in line with the organization of the rat SCIP gene (the rat homologue of Oct6) that has recently been described [20]. Mapping and sequencing of cDNA molecules covering the 3'UTR of the SCIP gene revealed that indeed the entire gene is encoded by one exon. The 5' part of both the rat and mouse genes are extremely rich in CpG dinucleotides. This region of high CpG content extends well beyond the transcriptional initiation site of the gene. However the Oct6 promoter has all the hallmarks of a typical cell type specific promoter in that it is characterized by the presence of a CCAAT and TATA box at canonical distances from the transcription initiation site. Thus the Oct6 gene presents an exceptional situation in which a cell type specific promoter is interwoven with a CpG island. As expected this CpG island is undermethylated in the different cell lines we looked at (MES, ES, ES+RA and MES68 cells). From this we infer that the methylation state of the CpG island does not correlate with the activity of the Oct6 promoter. It has been suggested by Kuhn *et al.* that the intronless SCIP/Oct6 gene arose through gene duplication involving a processed mRNA intermediate [20]. Possibly this duplicated gene intergrated in a CpG island that constitutes the regulatory region of a constitutively expressed gene. As expression on the Oct6 sense strand is cell type specific, ubiquitous expression from this CpG island then is expected to be on the opposite strand and divergent from the Oct6 gene. Whether this is the case remains to be determined.

The Oct6 gene is expressed at low levels in isolated Schwann cells and B103 neuroblastoma cells. The level of expression is drastically increased upon stimulation of these cells with cAMP. However no concensus cAMP response elements (CRE) are found in the promoter region although they could be present further away from the transcription initiation site. Instead two AP-2 like binding sites are found at position -90 and -540. The AP-2 element has been shown to mediate cAMP induced transcription [30]. In accordance with this, prelimenary data indicate that a -160 promoter/CAT construct transiently transfected in B103 cells is still cAMP inducible, suggesting that the most proximal AP-2 element is involved in mediating the cAMP effect (Ronald Zwart personal communication).

Transient transfection of deletion mutants of the Oct6 promoter suggest the presence of important enhancer elements in a region of approximately 80bp, upstream of the CCAAT box. This region coincides with the major DNAseI hypersensitive site in ES cells and shows homology with a region in the NGF-R promoter. As yet the importance of these findings is still obscure because the promoter constructs used here did not show cell type restricted activity. As the Oct6 promoter contains binding sites for ubiquitous transcription factors it is not surprising that it has a high basal activity in all cell types. Unfortenately, even the levels of activity are comparable in all cell lines tested (compare activity with the RSV promoter control). This lack of cell type specificity could be either a consequence of the transient assay itself or it might indicate that additional sequences are needed for proper cell type specific regulation, despite the inclusion of all the sequences that are DNAseI hypersensitive in ES cells. The latter possibility seems likely for the



following reason. Transient transfection of the pgOct6/Py construct into COS cells did not result in the production of Oct6 protein as assayed by bandshift. In this construct the strong polyoma enhancer is cloned 5' of the Oct6 promoter at position -160 (Figure 2B, compare with construct -160CAT in Figure 7) driving the rest of the Oct6 gene. This result could indicate that sequences within the Oct6 gene from the start codon to the 3' EcoRI site are involved in the regulation of the promoter, or alternatively, that the Py enhancer downregulates the Oct6 promoter in the pgOct6/py construct. The inclusion of this part of the gene in the CAT constructs described in Figure 7 and transfecting them to different cell types would give an answer to this question. If the failure to obtain cell type specific expression from the Oct6

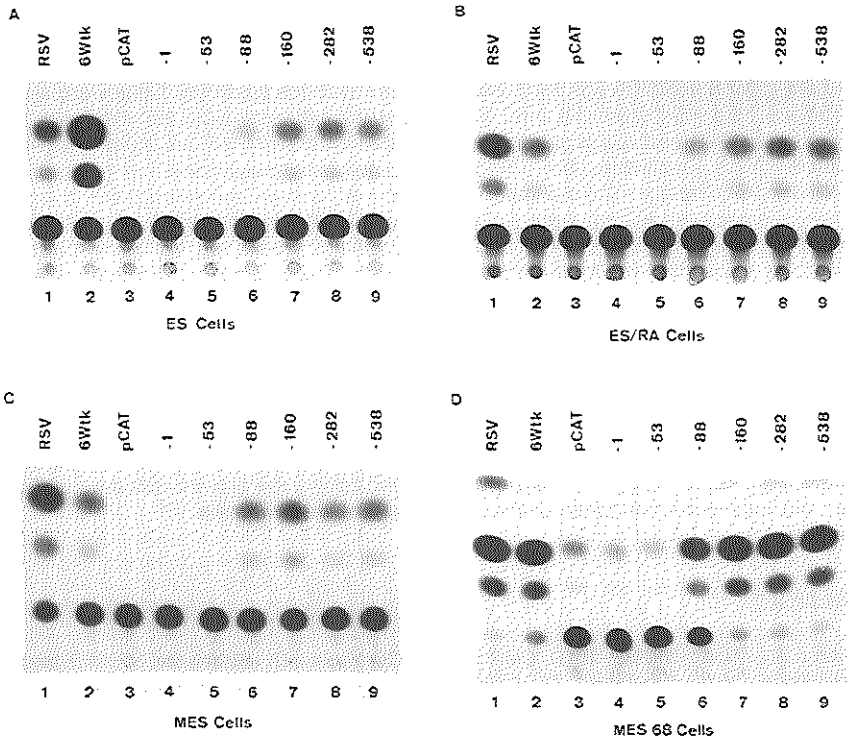


Figure 7. Functional dissection of the Oct6 promoter. A. CAT constructs.A XhoI site was introduced at the ATG codon of the Oct6 ORF (+55) by PCR and used to clone the different promoter constructs in the right orientation upstream of the CAT gene in pBLCAT3. The numbers indicate the number of nucleotides upstream of the CAP site present in the individual deletion constructs (see Figure 5).
B. CAT assays. RSV is the Rous Sarcoma Virus promoter and 6Wtk is the minimal HSV TK promoter as present in pBLCAT2 with a multimerized IgH enhancer upatream of it [7, 32, 33]. Note that the 6Wtk promoter is strongly expressed in undifferentiated ES cells but is downregulated upon RA induced differentiation of these cells.

promoter constructs is due to the nature of the transient assay the constructs can be tested in stable transfected B103 and MES cell lines. The advantage of this approach would be that the promoterconstructs are analysed while integrated in the cells chromosomes and thus in a normal chromatin configuration. Any integration position effects could be circumvented by analysing large pools of transfectants instead of clonal lines. The obvious and ideal alternative would be to generate transgenic mice using an easy identifiable reporter gene like the E.coli lacZ gene. This would enable the analysis of Oct6 expression during all stages of mouse development. Clearly more experiments need to be done to link individual cis-acting sequences with extra and intra cellular stimuli that modulate Oct6 gene expression.

MATERIAL AND METHODS

Mapping of DNAasel hypersensitive sites.

Cultured cells (MES and CCE ES) were washed twice with ice cold PBS. Cells were suspended in 3 ml of ice cold HS-buffer (15 mM Tris-HCl pH 7.4, 60 mM KCl, 15 mM NaCl, 0.2 mM EDTA, 0.2 mM EGTA and 5% glycerol, supplemented with 1mM DTT, 0.15 mM spermine and 0.5 mM spermidine, just before use). The cells were disrupted by passing 5 to 10 times through a 0.5x16 mm (25G) needle. The disruption was followed by examining a drop of the suspension under the microscope. Nuclei were collected by a brief centrifugation at 2500 rpm and resuspended to a final concentration of approximately $5x10^7$ nuclei per mililitre of HS-buffer (the nuclei do not pack thightly after this centrifugation step). Limited DNAasel digestion of chromatin was carried out in a final volume of 500 μl of HS-buffer containing $5x10^6$ nuclei, 5mM $MgCl_2$ and DNAasel, varying in concentration from 0 to 400 units. The reaction was incubated for 15 or 30 minutes on ice, stopped by adding 10 μl of 0.5 M EDTA, 12.5 μl of 20% SDS and 50 μl of a 10mg/ml Proteinase K solution and the mixture was further incubated overnight at 37 C. After phenol/chloroform extraction the DNA was collected by isopropanol precipetation. The DNA was dissolved in 175 μl TE and restricted with EcoRI and XhoI or with EcoRI and HindIII. Restriction fragments were resolved on a 0.8% agarose gel and blotted onto nylon membrane (Zeta Probe). The blots were hybridized to the probes indicated in Figure 4. All other methods used have been described in the previous chapters.

REFERENCES

1. **Herr, W., R.A. Sturm, R.G. Clerc, L.M. Corcoran, D. Baltimore, P.A. Sharp, H.A. Ingraham, M.G. Rosenfeld, M. Finney, G. Ruvkun and H.R. Horvitz.** 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**: 1513-6
2. **Rosenfeld, M.G.** 1991. POU-domain transcription factors: pou-er-ful developmental regulators. *Genes & Dev.* **5**: 897-907
3. **Ruvkun, G. and M. Finney.** 1991. Regulation of transcription and cell identity by POU domain proteins. *Cell* **64**: 475-8
4. **Schöler, H.R.** 1991. Octamania: The POU factors in murine development. *Trends Genet* **7**: 323-9
5. **Li, S., E.B. Crenshaw III, E.J. Rawson, D.M. Simmons, L.W. Swanson and M.G. Rosenfeld.** 1990. Dwarf locus mutants lacking three pituitary cell types result from mutations in the POU-domain gene pit-1. *Nature* **347**: 528-33
6. **Kemler, I. and W. Schaffner.** 1990. Octamer transcription factors and the cell

type-specificity of immunoglobulin gene expression. *Faseb J* **4**: 1444-9
7. **Schöler, H.R., R. Balling, A.K. Hatzopoulos, N. Suzuki and P. Gruss.** 1989. Octamer binding proteins confer transcriptional activity in early mouse embryogenesis. *Embo J* **8**: 2551-7
8. **He, X., M.N. Treacy, D.M. Simmons, H.A. Ingraham, L.W. Swanson and M.G. Rosenfeld.** 1989. Expression of a large family of POU-domain regulatory genes in mammalian brain development. *Nature* **340**: 35-41
9. **Meijer, D., A. Graus, R. Kraay, A. Langeveld, M.P. Mulder and G. Grosveld.** 1990. The octamer binding factor Oct6: cDNA cloning and expression in early embryonic cells. *Nucleic Acids Res* **18**: 7357-65
10. **Monuki, E.S., G. Weinmaster, R. Kuhn and G. Lemke.** 1989. SCIP: a glial POU domain gene regulated by cyclic AMP. *Neuron* **3**: 783-93
11. **Schöler, H.R., A.K. Hatzopoulos, R. Balling, N. Suzuki and P. Gruss.** 1989. A family of octamer-specific proteins present during mouse embryogenesis: evidence for germline-specific expression of an Oct factor. *Embo J* **8**: 2543-50
12. **Suzuki, N., H. Rohdewohld, T. Neuman, P. Gruss and H.R. Schöler.** 1990. Oct-6: a POU transcription factor expressed in embryonal stem cells and in the developing brain. *Embo J* **9**: 3723-32
13. **Monuki, E.S., R. Kuhn, G. Weinmaster, B.D. Trapp and G. Lemke.** 1990. Expression and activity of the POU transcription factor SCIP. *Science* **249**: 1300-3
14. **He, X., R. Gerrero, D.M. Simmons, R.E. Park, C.J. Lin, L.W. Swanson and M.G. Rosenfeld.** 1991. Tst-1, a member of the POU domain gene family, binds the promoter of the gene encoding the cell surface adhesion molecule Po. *Mol Cell Biol* **11**: 1739-44
15. **Meijer, D., A. Graus and G. Grosveld.** 1992. Mapping the transactivation domain of the Oct6 POU transcription factor. submitted for publication
16. **Okayama, H. and P. Berg.** 1983. A cDNA cloning vector that permits expression of cDNA inserts in mammalian cells. *Mol Cell Biol* **3**: 280-9
17. **Yamaguchi, M. and A. Matsukage.** 1990. DNA replication can overcome the silencer function on transcription. *New Biol.* **2**: 343-350
18. **Gross, D.S. and W.T. Garrard.** 1988. Nuclease hypersensitive sites in chromatin. *Annu Rev Biochem*
19. **Svaren, J. and R. Chalkley.** 1990. The structure and assembly of active chromatin. *Trends Genet* **6**: 52-6
20. **Kuhn, R., E.S. Monuki and G. Lemke.** 1991. The gene encoding the transcription factor SCIP has features of an expressed retroposon. *Mol Cell Biol* **11**: 4642-50
21. **Bird, A.** 1987. CpG islands as gene markers in the vertebrate nucleus. *Trends Genet.* **3**: 342-347
22. **Bird, A.P., M.H. Taggart, R.D. Nicholls and D.R. Higgs.** 1987. Non-methylated CpG-rich islands at the human alpha-globin locus: implications for evolution of the alpha-globin pseudogene. *Embo J* **6**: 999-1004
23. **Spanopoulou, E., V. Giguere and F. Grosveld.** 1988. Transcriptional unit of the murine Thy-1 gene: different distribution of transcription initiation sites in brain. *Mol Cell Biol* **8**: 3847-56
24. **Lemke, G., E. Lamar and J. Patterson.** 1988. Isolation and analysis of the gene encoding peripheral myelin protein zero. *Neuron* **1**: 73-83
25. **Kadonaga, J.T., K.R. Carner, F.R. Masiarz and R. Tjian.** 1987. Isolation of cDNA encoding transcription factor Sp1 and functional analysis of the DNA binding domain. *Cell* **51**: 1079-90
26. **Kageyama, R. and I. Pastan.** 1989. Molecular cloning and characterization of a human DNA binding factor that represses transcription. *Cell* **59**: 815-25
27. **Lemaire, P., C. Vesque, J. Schmitt, H. Stunnenberg, R. Frank and P. Charnay.** 1990. The serum-inducible mouse gene Krox-24 encodes a sequence-specific transcriptional activator. *Mol Cell Biol* **10**: 3456-67
28. **Chavrier, P., C. Vesque, B. Galliot, M. Vigneron, P. Dolle, D. Duboule and P. Charnay.**

1990. The segment-specific gene Krox-20 encodes a transcription factor with binding sites in the promoter region of the Hox-1.4 gene. *Embo J* **9**: 1209-18
29. **Sehgal, A., N. Patil and M. Chao.** 1988. A constitutive promoter directs expression of the nerve growth factor receptor gene. *Mol Cell Biol* **8**: 2543-50
30. **Imagawa, M., R. Chiu and M. Karin.** 1987. Transcription factor AP-2 mediates induction by two different signal-transduction pathways: protein kinase C and cAMP. *Cell* **51**: 251-60
31. **Thomas, K.R. and M.R. Capecchi.** 1987. Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. *Cell* **51**: 503-12
32. **Gorman, C.** 1985. High efficiency gene transfer into mammalian cells. in "DNA Cloning", D. M. Glover IRL Press, Oxford
33. **Luckow, B. and G. Schutz.** 1987. CAT constructions with multiple unique restriction sites for the functional analysis of eukaryotic promoters and regulatory elements. *Nucleic Acids Res* **15**:

Chapter IX

CONCLUDING REMARKS

The work described in this thesis concerns the identification and characterization of octamer binding transcription factors that could play a role in several aspects of cellular differentiation processes. Two mouse genes encoding octamer binding factors have been characterized and studied. The Oct2 gene is highly expressed in B cells, in brain and in meiotic male germ cells, while the Oct6 gene is expressed in early embryonic cells, the developing and adult nervous system and in differentiating glia of the peripheral and central nervous system (Chapter V and [1-4]). In the absence of genetic data, a role for these factors in differentiation is mainly inferred from their tissue and stage specific expression patterns and the fact that they can function as transcription factors. As the testis specific expression of the mouse Oct2 gene and its putative role in germ cell development has been discussed extensively in Chapter V, I will restrict myself here to a discussion of a number of aspects of the Oct6 gene and its regulated expression.

We used P19 EC cells to study the regulated expression of a number of octamer binding factors, focusing on the Oct6 protein (Chapter VI). The direction of differentiation *in vitro* of this cell line can be manipulated using different culture conditions in combination with differentiation inducing agents such as retinoic acid (RA) or dimethylsulfoxide (DMSO) [5, 6, 7]. Aggregation of P19 cells followed by administration of RA leads to differentiation in a neurectodermal direction. The Oct6 gene is transiently induced during the first period of differentiation and reappears at a later stage, suggesting that the gene plays a role at an early commitment stage and at a later stage in the establishment of differentiated cell types [4]. This assumption is in line with what is known about the expression of the Oct6 gene in the developing embryo and during postnatal development of the nervous system. Using *in situ* hybridization histochemistry, Oct6 mRNA could be detected in day 8 rat embryos at the egg cylinder stage. At later stages of mouse and rat development, the expression of the gene is confined to discrete regions of the developing central nervous system, while in new born animals high levels of expression are observed in differentiating glia of the peripheral as well as the central nervous system. Furthermore, Oct6 expression could also be detected in discrete regions of the adult brain [1, 3, 8]. Despite its obvious limitations, the P19 EC cell system might be useful as a model system to study the possible role(s) of the Oct6 gene during some aspects of neuronal differentiation. In particular the early expression of the gene, preceding morphological differentiation of the P19 cells, is intriguing (Chapter VI). It is possible that the expression of the gene is involved in an early step that commits the cells to follow a neurectodermal

differentiation pathway. At present it is not known whether all differentiating cells express Oct6 at this early stage, or that expression is limited to a subset of cells. The use of antibodies directed against Oct6 in conjunction with antibodies against different differentiation markers should enable us to address this question. Studying the neuronal differentiation capacity of P19 cells in which the expression pattern of the Oct6 gene is altered might give insight into the role this gene plays in the differentiation process. Manipulation of the Oct6 expression pattern in P19 cells could be done either by constitutive expression of the gene, or by inhibiting the expression of the gene through the use of anti-sense expression vectors or anti-sense oligonucleotides. An alternative approach would be the creation of Oct6 mutant mice through the targeted disruption of the Oct6 gene in ES cells followed by reintroduction of these cells into blastocyst embryos. Breeding of germ line chimeric animals to homozygosity will hopefully give insight into the role of the Oct6 gene in normal development.

The rat homologue of the Oct6 gene, called Tst-1 or SCIP, has been shown to be expressed in rat testis. However we could not detect expression of the Oct6 gene in mouse testis by conventional Northern blot experiments, or by PCR amplification of POU domain protein encoding mRNAs using degenerate oligonucleotides. The last approach was also taken by Goldsborough *et al.*[9]. They obtained cDNAs for Oct1, Oct2, Brn3 and a gene which they called Oct11, but not Oct6. In contrast with these results, Suzuki *et al.* detected a very low expression of the Oct6 gene in mouse testis [8]. As described in Chapter V, the Oct2 gene is highly expressed in mouse testis, whereas this gene does not seem to be expressed in the rat testis. So there are clear differences between rat and mouse testis with respect to the POU domain genes they express. The reason for these differences is unclear but might indicate that the different POU proteins are functionally redundant in testis.

Given the observation that the Oct6 protein is a sequence specific DNA binding protein that is capable of stimulating or repressing transcription, it is likely that the protein is involved in cellular differentiation through regulation of a set of downstream genes. However, the Oct6 protein is one of a family of proteins that bind to the octamer sequence, and many cell types contain multiple octamer binding proteins that could compete for the same binding site. For example, ES cells express three octamer binding factors; Oct1, Oct3/4 and Oct6. This poses the question as to how a specific set of target genes can be regulated by individual members of the octamer binding protein family. Obviously, there are two levels at which specificity can be achieved: differences in DNA binding specificity for different binding sites, and differences in the spectrum of proteins with which the individual oct factors interact via protein-protein contacts. Thus although embryonic stem cells express three different proteins that bind to the octamer sequence *in vitro*, the affinity for this sequence and related sequences differs for the individual

proteins, reflecting differences in their respective POU domains. We have studied the differential affinity of the Oct1 and the Oct6 protein for a number of octamer and octamer related binding sites (Chapter V). It was concluded that these two proteins bind to a different but largely overlapping set of binding sites. A comparison between the seven Oct6 binding sites studied here and the five Oct6 binding sites as present in the P_0 promoter yielded the following rather loose consensus; A/TATG/TNA/TAATNA/C (see figure).
It was shown for the Pit-1 protein that phosphorylation of a serine residue in the amino terminal cluster of the homeodomain results in altered DNA binding affinities for different Pit-1 DNA binding sites [11]. Similarly, phosphorylation of the Oct1 protein as a function of the cell cycle modulates its DNA binding affinity for the octamer element [12]. Thus phosphorylation of POU proteins might be a general mechanism by which the specificity/affinity of DNA binding is regulated. Whether this is also true for the Oct6 protein remains to be established
Given the fact that the DNA binding specificity of Oct6 overlaps greatly with that of Oct1 and other Oct factors, it is not likely that DNA binding alone determines the specificity for the genes that are regulated by the individual

Oct6 DNA binding sites

		Relative affinity
AblOct	GGAAATGCAAATTCCT	+++
IgHOct	AGAAATGCAAATTACC	+++
ad4	TAATATGCAAATAAGG	+++
ad2	CAATATGATAATGAGG	+++
ICP4	TCGGGCGGTAATGAGA	+++
EN26	CATACTTAAAATTCAA	+
Cf1a	CATAAATCAAATTGTA	+
P_0 I	CCTGCTTAAAATCCCC	+
P_0 II	AGCAGTTTAAATGATC	++
P_0 III	TACCATTCTAATACGA	+++
P_0 IIIas	TAGAATGGTAATTCAG	+++
P_0 IVas	TGCACTAGGAATACCA	++
P_0 Vas	GATAAGGGAAATAAGG	+

Consensus: $\begin{matrix}A\\T\end{matrix}AT\begin{matrix}G\\T\end{matrix}N\begin{matrix}A\\T\end{matrix}AATN\begin{matrix}A\\C\end{matrix}$

Consensus DNA binding site for the Oct6 protein. It should be noted that the relative affinity of Oct6 for the first seven DNA binding sites was based roughly on the amount of complexed DNA in a bandshift experiment using equal amounts of crude nuclear extracts. The relative affinities of the P_0 DNA binding sites were taken from experiments by He *et al.* [10] and can therefore not be directly compared to our experiments.

proteins. From a number of examples it is clear that further specificity is achieved through interaction with other transcription factors, depending on determinants in the POU domain as well as outside the POU domain (see Introduction). Thus the Oct2 protein is capable of activating a simple promoter consisting of a minimal β-globin promoter and flanking octamer motif, whereas Oct1 is not [13, 14]. These proteins do not differ in their DNA binding affinity for the octamer element ([15, 16], and references therein). Both Oct1 and Oct2 contain an amino terminal transcriptional activation domain, while Oct2 has an additional activation domain carboxyl terminal of the POU domain [13, 17, 18]. Through the use of chimeric molecules it was shown that transcriptional activation by Oct2 depends on the presence of both the amino terminal and the carboxyl terminal activation domain [13]. Activation of the IE genes of HSV depends on the productive interaction between the viral protein VP16, Oct1 and at least one other cellular factor on the IE response element [19-22]. The interaction between Oct1 and VP16 depends on critical amino acid residues within the Oct1 homeodomain. The Oct2 protein is not capable of interacting with VP16 [23]. Thus intrinsic differences between Oct1 and Oct2 determine whether a promoter can be activated or not.

A synthetic promoter consisting of a minimal tk promoter flanked by a multimerized IgH octamer/μE4 enhancer (6W) was shown to be highly active in undifferentiated EC and ES cells. However, ectopic expression of the Oct4 protein in Hela cells failed to activate this promoter, suggesting that additional cell type specific factors are missing or that the promoter is repressed in Hela cells [24, 25]. Activation of this promoter in Hela cells could be achieved when the Oct4 protein was coexpressed with the adenovirus E1A gene. Overexpression of either Oct4 or E1A resulted in a reduced activity of the 6Wtk promoter. Furthermore, bandshift experiments suggested that the Oct4 protein and E1A interact directly. Activation of the 6Wtk promoter is dependent on an intact transcriptional activation domain of the Oct4 protein. The interaction between Oct4 and E1A is specific, as Oct2 failed to stimulate the 6Wtk promoter in the presence of E1A. Thus the E1A protein functions as a bridging factor that mediates the transcriptional activation effect of the Oct4 protein to the promoter. It was suggested that the viral protein mimics the action of a stem cell specific factor with which Oct4 interacts in these cells, and that this factor is the elusive E1A like activity [26]. Thus different sets of gene promoters can be regulated by different Oct factors through a combination of DNA binding specificity and selective interaction with other factors.

Although most Oct factors are clearly involved in transcriptional regulation, a number of observations implicate these factors also in cellular proliferation, reflecting a direct or indirect role in DNA replication. For example the Oct1 and Oct2 proteins stimulate adenovirus DNA replication *in vitro*. The stimulation of viral DNA replication does not depend on the

transcriptional activation domains of these proteins, as an intact POU domain is sufficient [27]. Furthermore, the specific inhibition of Oct3/4 mRNA translation in fertilized mouse eggs resulted in a failure of these cells to undergo cell division. Coinjection of *in vitro* transcribed Oct3/4 mRNA could revers the effect. This inhibitory effect on the first cell division could also be obtained by injection of double-stranded oligonucleotides containing an octamer binding site. A direct role of the Oct3/4 protein in DNA replication is suggested by the following observations. Treatment of one-cell mouse embryos with high concentrations of α-aminitin, which inhibits RNA polymerase II and III mediated transcription, does not block the first cell division but results in a developmental arrest at the two cell stage. Furthermore, the injection of Oct3/4 anti-sense oligonucleotides resulted not only in a block of cell division but also in a 87% reduction of DNA replication [28]. The fact that the block in the first cell division can be relieved by coinjection of *in vitro* transcribed Oct3/4 mRNA opens the possibility to test whether this function solely depends on the Oct3/4 DNA binding domain or that the transcriptional activation domain is essential as well. A role for the Oct6 protein in cellular proliferation has been suggested on the basis of its high expression in proliferating Schwann cells, and is down regulated when these cells differentiate and stop to divide [3]. It is possible that the high transient expression of the gene during early neuronal differentiation of P19 cells is linked in a similar way to the proliferation of precursor cells. Again the P19 cell system could prove to be a valuable system to study the proposed function of the Oct6 protein in cellular proliferation. If it is true that the Oct6 protein is involved in the proliferation of precursor cells, and that down-regulation of the gene is required for further differentiation, it is anticipated that a sustained expression of the gene will block terminal differentiation of these cells. It would be very interesting to study the effects of sustained expression of Oct6 in Schwann cells and oligodendrocyte precursor cells in mice. This would require the generation of transgenic animals carrying the Oct6 gene, preferably under control of a Schwann cell specific promoter. The Oct6 promoter itself would be the ideal promoter to use, provided that it is possible to identify and selectively mutate the cis-acting element(s) that govern the down regulation of this promoter during terminal cell differentiation. Clearly a detailed study of the Oct6 promoter is mandatory.

One remarkable piece of data from the preliminary characterization of the Oct6 gene as presented in Chapter VIII, is that the Oct6 promoter is a classical cell type specific promoter although it is embedded in a CpG island. CpG islands are a characteristic feature of promoter regions of ubiquitously expressed genes [29]. Most of these promoters lack a TATA box, and some generate transcripts from both DNA strands. Transcription is mostly promiscuous with respect to the transcriptional initiation site used. It is possible that this CpG islands represents the promoter of a housekeeping

gene that is located on the bottom strand and is thus transcribed in the opposite direction with respect to the Oct6 gene. If so, transcription on the Oct6 sense strand from this CpG promoter must be actively prevented. In this respect it is interesting to note that we observed a strong footprint overlapping the TATA box when using nuclear extracts from kidney in DNAase I protection assays. Since kidney cells do not express the Oct6 gene, this factor could be involved in the repression of the Oct6 promoter in these cells. Thus, it is possible that the Oct6 gene is under positive as well as negative control. A first dissection of the Oct6 gene promoter might be undertaken using the B103 cell line and promoter deletion constructs. This cell line expresses the Oct6 gene at a very low level, but expression is increased upon stimulation of these cells with cAMP. As a consequence of cAMP stimulation, the cells differentiate morphologically and start to express the Schwann cell specific marker gene PLP. Thus this cell line will be useful to study one aspect of Oct6 gene regulation that might be relevant for its regulation in developing Schwann cells. Ultimately, the promoter deletion constructs must be tested in transgenic animals to asses the relative importance of the individual promoter elements. Using then these cis-acting DNA elements as probes we will be able to study the factors that are involved in the tissue specific modulation of the Oct6 gene itself.

A detailed study of the Oct6 promoter and enhancer sequences and the protein factors that interact with them, will give insight into the regulatory mechanisms that incorporate intra- and extra-cellular cues to modulate the expression of the Oct6 gene, and place the Oct6 gene in the proposed hierarchy of events that lead to the acquisition of a fully differentiated phenotype.

REFERENCES

1. **He, X., M.N. Treacy, D.M. Simmons, H.A. Ingraham, L.W. Swanson and M.G. Rosenfeld.** 1989. Expression of a large family of POU-domain regulatory genes in mammalian brain development. *Nature* 340: 35-41
2. **Monuki, E.S., G. Weinmaster, R. Kuhn and G. Lemke.** 1989. SCIP: a glial POU domain gene regulated by cyclic AMP. *Neuron* 3: 783-93
3. **Monuki, E.S., R. Kuhn, G. Weinmaster, B.D. Trapp and G. Lemke.** 1990. Expression and activity of the POU transcription factor SCIP. *Science* 249: 1300-3
4. **Meijer, D., A. Graus, R. Kraay, A. Langeveld, M.P. Mulder and G. Grosveld.** 1990. The octamer binding factor Oct6: cDNA cloning and expression in early embryonic cells. *Nucleic Acids Res* 18: 7357-65
5. **McBurney, M.W., K.R. Reuhl, A.I. Ally, S. Nasipuri, J.C. Bell and J. Craig.** 1988. Differentiation and maturation of embryonal carcinoma-derived neurons in cell culture. *J Neurosci* 8: 1063-73
6. **Rudnicki, M.A. and M.W. McBurney.** 1987. Cell culture methods and induction of differentiation of embryonal carcinoma cell lines. in "Teratocarcinomas and Embryonic Stem Cells", E. J. Robertson ed. IRL Press, Oxford
7. **Jones Villeneuve, E.M., M.A. Rudnicki, J.F. Harris and M.W. McBurney.** 1983. Retinoic acid-induced neural differentiation of embryonal carcinoma cells. *Mol Cell Biol* 3: 2271-9
8. **Suzuki, N., H. Rohdewohld, T. Neuman, P. Gruss and H.R. Schöler.** 1990. Oct-6: a POU transcription factor expressed in embryonal stem cells and in the developing brain. *Embo*

J **9**: 3723-32
9. **Goldsborough, A., A. Ashworth and K. Willison.** 1990. Cloning and sequencing of POU-boxes expressed in mouse testis. *Nucleic Acids Res* **18**: 1634
10. **He, X., R. Gerrero, D.M. Simmons, R.E. Park, C.J. Lin, L.W. Swanson and M.G. Rosenfeld.** 1991. Tst-1, a member of the POU domain gene family, binds the promoter of the gene encoding the cell surface adhesion molecule Po. *Mol Cell Biol* **11**: 1739-44
11. **Kapiloff, M.S., Y. Farkash, M. Wegner and M.G. Rosenfeld.** 1991. Variable effects of phosphorylation of Pit-1 dictated by the DNA response elements. *Science* **253**: 786-9
12. **Segil, N., S.B. Roberts and N. Heintz.** 1991. Mitotic Phosphorylation of the Oct-1 Homeodomain and Regulation of Oct-1 DNA Binding Activity. *Science* **254**: 1814-1816
13. **Tanaka, M. and W. Herr.** 1990. Differential transcriptional activation by Oct-1 and Oct-2: interdependent activation domains induce Oct-2 phosphorylation. *Cell* **60**: 375-86
14. **Tanaka, M., J.-S. Lai and W. Herr.** 1992. Promoter-selective activation domains in Oct-1 and Oct-2 direct differential activation of an snRNA and mRNA promoter. *Cell* **68**: 755-767
15. **Kemler, I. and W. Schaffner.** 1990. Octamer transcription factors and the cell type-specificity of immunoglobulin gene expression. *Faseb J* **4**: 1444-9
16. **Schaffner, W.** 1989. How do different transcription factors binding the same DNA sequence sort out their jobs? *Trends Genet* **5**: 37-9
17. **Gerster, T., C.G. Balmaceda and R.G. Roeder.** 1990. The cell type-specific octamer transcription factor OTF-2 has two domains required for the activation of transcription. *Embo J* **9**: 1635-43
18. **Müller, I.M., W. Schaffner and P. Matthias.** 1990. Transcription factor Oct-2A contains functionally redundant activating domains and works selectively from a promoter but not from a remote enhancer position in non-lymphoid (HeLa) cells. *Embo J* **9**: 1625-34
19. **Katan, M., A. Haigh, C.P. Verrijzer, P.C. van der Vliet and P. O'Hare.** 1990. Characterization of a cellular factor which interacts functionally with Oct-1 in the assembly of a multicomponent transcription complex. *Nucleic Acids Res* **18**: 6871-80
20. **Kristie, T.M., J.H. LeBowitz and P.A. Sharp.** 1989. The octamer-binding proteins form multi-protein--DNA complexes with the HSV alpha TIF regulatory protein. *Embo J* **8**: 4229-38
21. **O'Hare, P., C.R. Goding and A. Haigh.** 1988. Direct combinatorial interaction between a herpes simplex virus regulatory protein and a cellular octamer-binding factor mediates specific induction of virus immediate-early gene expression. *Embo J* **7**: 4231-8
22. **Preston, C.M., M.C. Frame and M.E. Campbell.** 1988. A complex formed between cell components and an HSV structural polypeptide binds to a viral immediate early gene regulatory DNA sequence. *Cell* **52**: 425-34
23. **Stern, S., M. Tanaka and W. Herr.** 1989. The Oct-1 homoeodomain directs formation of a multiprotein-DNA complex with the HSV transactivator VP16. *Nature* **341**: 624-30
24. **Schöler, H.R., G.R. Dressler, R. Balling, H. Rohdewohld and P. Gruss.** 1990. Oct-4: a germline-specific transcription factor mapping to the mouse t-complex. *Embo J* **9**: 2185-95
25. **Schöler, H.R., R. Balling, A.K. Hatzopoulos, N. Suzuki and P. Gruss.** 1989. Octamer binding proteins confer transcriptional activity in early mouse embryogenesis. *Embo J* **8**: 2551-7
26. **Schöler, H.R., T. Ciesiolka and P. Gruss.** 1991. A nexus between Oct-4 and E1A: implications for gene regulation in embryonic stem cells. *Cell* **66**: 291-304
27. **Verrijzer, C.P., A.J. Kal and P.C. van der Vliet.** 1990. The DNA binding domain (POU domain) of transcription factor oct-1 suffices for stimulation of DNA replication. *Embo J* **9**: 1883-8
28. **Rosner, M.H., S.R. De, H. Arnheiter and L.M. Staudt.** 1991. Oct-3 is a maternal factor required for the first mouse embryonic division. *Cell* **64**: 1103-10
29. **Bird, A.** 1987. CpG islands as gene markers in the vertebrate nucleus. *Trends Genet.* **3**: 342-347

SUMMARY

Molecular genetic studies on *Drosophila* and *C.elegans* have suggested the existence of a hierarchy of regulatory genes that operate during embryogenesis. The sequential activation of these genes accomplish the transformation of genetic information into body form and structure through a spatial and temporal regulation of effector genes, which in turn determine cell identity.

The work described in this thesis aimed at the elucidation of mechanisms that govern cellular differentiation events in male germ cell development, especially during the post-meiotic phase, and in embryonal carcinoma cells. The experimental work focused on the characterization of a class of cell type specific transcription factors, the octamer binding factors, that could play a role in the cellular differentiation events mentioned above. These factors belong to the POU domain protein family. In Chapter I this family of proteins is introduced and their role in cellular differentiation is briefly discussed. The first part of this thesis (Chapters II to IV) is mainly concerned with the testis specific expression of the c-abl gene. In the second part of this thesis (Chapters V to VIII) two genes encoding octamer binding factors (Oct2 and Oct6) are described that are expressed during distinct stages of male germ cell development (Oct2) and neuronal differentiation of P19 EC cells (Oct6). Their supposed role in cellular differentiation is discussed in Chapters V and VIII, respectively.

In Chapter II, the molecular characterization is described of the testis specific c-abl mRNA. The c-abl gene expresses two mRNAs of 5.3 and 6.5 kb in all cell lines and tissues tested, both encoding a non receptor type tyrosine specific kinase that is involved in an as yet uncharacterized signaling pathway. The high expression of a shorter c-abl transcript of 4 kb (TSabl) in post-meiotic male germ cells suggested an involvement of the c-abl protein in the terminal cytodifferentiation of these cells. The TS abl mRNA is transcribed from the proximal of the two c-abl promoters and is alternatively processed, resulting in a removal of most of the 3'UTR, that does not affect the coding capacity of the mRNA. The high levels of this mRNA in round and elongated spermatids could be due to two not mutually exclusive mechanisms, i.e. a higher stability of this mRNA or/and ongoing expression of the mRNA during the later stages of spermiogenesis. Indeed ActinomycinD chase experiments showed that the TSabl mRNA in round spermatids has a longer halflife than the two c-abl RNAs expressed in other cell types. In Chapter III we describe experiments that were designed to test the involvement of the 3'UTR in determining messenger halflife. It was shown that the removal of the 3'UTR of the c-abl mRNA did not have an appreciable effect on the rate at which this RNA is turned over in COS cells. Nor does it alter the translatability of the RNA. Therefore, it remains unclear what function the alternative polyadenylation plays in TSabl RNA maturation and

what factors cause the high stability of the transcript in spermatids.
S1 Nuclease protection assays indicated that the TSabl mRNA is transcribed from the proximal of the two c-abl promoters. A prelimenary analysis of the proximal c-abl promoter is presented in Chapter IV, using DNAseI footprinting and gel retardation assays. Four regions were identified that are protected from limited DNAseI digestion by nuclear protein extracts from testis, spermatids, liver and pre-B cells. There are no obvious differences between the footprints obtained. In a gel retardation experiment, we observed a testis specific complex with a probe that contained the octamer sequence. This specific complex, however, most likely resulted from proteolytic degradation of the ubiquitous Oct1 protein.

The initial observation of a testis specific octamer complex stimulated us to undertake the experiments described in Chapter V. These experiments aimed at the cloning of testis specific cDNAs encoding octamer binding factors. We show that the POU domain gene Oct2 is highly expressed in spermatogenic cells, generating two transcripts through a mechanism of alternative processing and/or promoter usage. The two cDNAs analysed potentially encode octamer binding factors of 44 and 39 kD. At the moment it is still unclear whether these proteins play a role in the post-meiotic expression of the c-abl gene, for two reasons. First, the Oct2 gene is highly expressed during the pachytene stage of spermatogensis whereas the TSabl is expressed after completion of meiosis, and second, it is as yet unclear whether the tsOct2 mRNAs are actually translated in male germ cells as no octamer complexes of the expected size were observed in a gel retardation assay using testis nuclear extracts. This failure to detect the tsOct2 complexes does not seem to be due to an intrinsic inability of the tsOct2 mRNAs to be translated, as the two tsOct2 complexes were observed when using nuclear extracts of tsOct2 transfected COS cells. At the same time, a second POU domain gene, called Tst-1, was reported to be expressed at high levels in rat testis. However, we found that the mouse homologue of this gene is not expressed in mouse testis but is expressed in undifferentiated embryonic stem cells and in differentiating P19 embryonal carcinoma cells. Chapter VI describes the temporally regulated expression of a number of octamer binding factors during the neuronal differentiation of P19 EC cells. One factor, called Oct6, is expressed in a bi-phasic pattern suggesting that it might play a role at different stages of development. This factor is further characterized by cloning of the cognate cDNA, and was found to be the mouse homologue of the rat Tst-1 gene. The cDNA encodes a protein of 45 kD that belongs to the POU domain protein family. The Oct6 gene generates a single mRNA of three kb and is expressed, apart from differentiating P19 EC cells, in undifferentiated ES cells and in brain. In Chapter VII we show that the Oct6 factor is capable of activating different promoters via octamer, and octamer related, sequences. The transcriptional activation potential resides in the amino terminal third of the protein, a

domain that is extremely rich in glycine and alanine residues, and is distinct from the DNA binding domain (the POU domain). In Chapter VIII, the genomic organization of the Oct6 gene is described. The entire protein coding region and the promoter are part of a CpG island. Nevertheless, the Oct6 promoter has all the hallmarks of a classical cell type specific promoter in that it contains a TATA and CCAAT box at canonical distances from the transcription initiation site. The Oct6 promoter is associated with a DNAseI hypersensitive chromatin structure in Oct6 expressing ES cells but not in non-expressing MES cells. A prelimenary functional characterization of the Oct6 promoter suggests that sequences upstream of the CCAAT box are neccessary for full activity. A more detailed analysis of the Oct6 promoter is needed to identify the cis-acting elements that govern the strict tissue specific expression of the gene. Characterization of the protein factors that interact with these cis-acting elements will enable us to study the mechanisms that regulate this transcriptional regulator and place the Oct6 gene in a hierarchy of events that lead to a fully differentiated phenotype.

Samenvatting

Moleculair genetisch onderzoek aan *Drosophila* en *C.elegans* suggereert dat er een hiërarchie van regulerende genen bestaat. De opeenvolgende activering van deze genen gedurende de embryogenese 'vertaalt' de genetische informatie in vorm en structuur van het organisme, door de temporele en spatiële regulatie van andere genen welke op hun beurt de uiteindelijke identiteit van de lichaamscellen bepaalt.

Het in dit proefschrift beschreven werk had tot doel inzicht te verkrijgen in de moleculaire mechanismen welke de differentiatie van cellen reguleren, met name de differentiatie van de mannelijke kiemcellen en embryonale carcinoma cellen. Het experimentele werk concentreerde zich op de karakterisering van een klasse van cel type specifieke transcriptie factoren, de octameer bindende factoren, die een rol zouden kunnen spelen in de hierboven genoemde cellulaire differentiatie processen. Deze factoren behoren tot de POU domein eiwit familie. In Hoofdstuk I wordt deze familie van eiwitten geïntroduceerd en wordt de rol van twee van deze eiwitten in verschillende cellulaire differentiatie processen beschreven. Het eerste gedeelte van dit proefschrift (Hoofdstukken II tot en met IV) behandelt voornamelijk de testis specifieke expressie van het c-abl gen. In het tweede gedeelte (Hoofdstukken V tot en met VIII) worden twee genen beschreven die octameer bindende transcriptie factoren coderen (Oct2 en Oct6). Het Oct2 gen komt tot expressie in een specifiek stadium van de zich ontwikkelende mannelijke zaadcel terwijl het Oct6 gen tot expressie komt gedurende de neurale ontwikkeling van P19 EC cellen. De veronderstelde rol die deze eiwitten spelen in cellulaire differentiatie wordt respectievelijk bediscussieerd in Hoofdstuk V en VII.

Hoofdstuk II beschrijft de moleculaire karakterisering van het testis specifieke c-abl mRNA. Het c-abl gen expresseert twee mRNAs van 5,3 en 6,5 kb in alle celtypen. Beide mRNAs coderen voor een niet-receptor type tyrosine kinase eiwit dat betrokken is bij een tot nu toe niet geïdentificeerd signaal transductie systeem. De hoge expressie van het kortere testis specifieke c-abl (TSabl) mRNA in haploïde spermatiden suggereert dat het een rol speelt bij de terminale differentiatie van deze cellen. Het c-abl gen heeft twee promotors. Het TSabl mRNA wordt afgeschreven van de proximale promotor en wordt op een alternatieve manier gemodificeerd, hetgeen resulteert in de verwijdering van het grootste gedeelte van de 3'UTR. Dit heeft geen consequenties voor het te coderen eiwit. Het hoge expressie niveau van het TSabl mRNA in ronde en elongerende spermatiden kan een gevolg zijn van twee elkaar niet uitsluitende mechanismen: Het TSabl mRNA is zeer stabiel en/of transcriptie van het TSabl mRNA persisteert tot de latere stadia van de spermiogenese. Met behulp van zogenaamde ActinomycineD experimenten konden we aantonen dat het TSabl mRNA in spermatiden inderdaad veel stabieler is dan de twee c-abl mRNAs in andere

cel typen. In Hoofdstuk III worden experimenten beschreven die er op gericht waren om de mogelijke betrokkenheid van de 3'UTR bij de mRNA stabiliteit te onderzoeken. Uit deze experimenten bleek dat de 3'UTR nauwelijks invloed heeft op de stabiliteit van het c-abl mRNA in COS cellen. Truncaties van de 3'UTR hebben ook geen invloed op de transleerbaarheid van het mRNA. Het is daarom niet duidelijk geworden wat de functie is van de alternatieve modificatie van het TSabl mRNA en welke factoren de hoge stabiliteit van het TSabl mRNA in spermatiden veroorzaken.

S1 Nuclease protectie experimenten wezen uit dat het TSabl mRNA getranscribeerd wordt van de proximale c-abl promotor. In Hoofdstuk IV wordt een eerste voorlopige karakterisering beschreven van deze promotor, gebruik makend van zogenaamde DNAseI footprinting en gelretardatie experimenten. Vier regio's werden op deze wijze geïdentificeerd in de c-abl promotor welke beschermd werden tegen een beperkte digestie met het nuclease DNAseI na incubatie met ruwe kernextracten van testis, spermatiden, lever en pre-B cellen. Er werden geen duidelijke verschillen gevonden tussen de footprint patronen die verkregen werden met de verschillende kernextracten. Echter in een gelretardatie experiment vonden wij een testis specifiek complex met een DNA probe die een octameer sequentie bevatte. Dit specifieke complex was hoogst waarschijnlijk het gevolg van proteolytische afbraak van het Oct1 eiwit. De observatie van een testis specifiek octameer complex stimuleerde ons de experimenten te ondernemen zoals beschreven in Hoofdstuk V. Deze experimenten waren er op gericht testis specifieke cDNAs te cloneren die octameer bindende factoren coderen. We vonden dat het POU domein gen Oct2 hoog tot expressie komt in mannelijke germinale cellen in de vorm van twee mRNAs, die gegenereerd worden door een mechanisme van alternatieve processing en/of promotor gebruik. De twee cDNAs coderen mogelijk octameer bindende factoren van 39 en 44 kD. Op het moment is het nog niet duidelijk of deze twee eiwitten een rol spelen in de post-meiotische expressie van het c-abl gen. Er zijn twee redenen om aan te nemen dat dit niet het geval is: Het Oct2 gen komt tot expressie gedurende het pachyteen stadium van de meiose, terwijl het c-abl gen hoog tot expressie komt nadat de kiemcellen de reductie deling voltooid hebben. Ten tweede, het is onduidelijk of de tsOct2 mRNAs wel vertaald worden aangezien geen octameer complexen van de verwachte grootte werden waargenomen in gelretardatie experimenten met testis kernextracten. Het feit dat de tsOct2 complexen niet werden waargenomen is hoogst waarschijnlijk niet het gevolg van een intrinsiek probleem in de transleerbaarheid van de tsOct2 mRNAs. Dit werd geconcludeerd uit het feit dat expressie van de twee tsOct2 cDNAs in COS cellen aanleiding geeft tot specifieke octameer complexen in een gelretardatie experiment. Rond die tijd werd de expressie beschreven van een tweede POU domein gen in rat testis. Dit gen werd Tst-1 genoemd. Echter, wij konden geen expressie aantonen van dit gen in de testis van de muis, maar

vonden dat het tot expressie komt in ongedifferentieerde embryonale stam cellen (ES cellen) en in differentiërende P19 embryo carcinoma cellen (EC cellen). In Hoofdstuk VI beschrijven wij de temporeel gereguleerde expressie van een aantal octameer bindende factoren gedurende de neurale differentiatie van P19 EC cellen. Eén factor, die Oct6 genoemd werd, komt tot expressie in een bi-fasisch patroon wat suggereerde dat de factor een rol zou kunnen spelen tijdens verschillende stadia van de neurale differentiatie van deze cellen. Deze factor werd nader gekarakteriseerd door clonering van het corresponderende cDNA. Dit cDNA bleek het muizen homoloog te zijn van het ratten Tst-1 gen. Het cDNA codeert voor een 45 kD eiwit en behoort tot de POU domein eiwit familie. Het Oct6 gen in de muis genereert een 3 kb mRNA en komt, naast differentiërende P19 cellen, tot expressie in ongedifferentieerde ES cellen en in de hersenen. De experimenten beschreven in Hoofdstuk VII laten zien dat het Oct6 eiwit een transcriptie factor is, dat in staat is verschillende promotors te activeren via octameer en octameer gerelateerde sequenties. Het eiwit domein betrokken bij deze functie bevindt zich in het aminoterminale gedeelte van het eiwit en is rijk aan de aminozuren glycine en alanine. Een mutant eiwit dat uitsluitend bestaat uit het POU domein, bond nog steeds DNA maar was niet in staat transcriptie te activeren van een test promotor. Hoofdstuk VIII beschrijft de genomische organisatie van het Oct6 gen. Het gehele eiwit coderende gedeelte en de promotor maken deel uit van een CpG eiland. Desondanks heeft de Oct6 promotor de karakteristieke eigenschappen van een cel-type specifieke promotor zoals het voorkomen van een TATA en CCAAT box sequentie op de juiste afstanden van de transcriptie initiatie plaats. De Oct6 promotor is geassocieerd met een DNAsel hypersensitieve chromatine structuur in expresserende ES cellen, maar niet in MES1 cellen die geen Oct6 expresseren. Een eerste functionele karakterisering van de Oct6 promotor suggereert dat sequenties 5' van de CCAAT box belangrijk zijn voor de activiteit van de Oct6 promotor. Een meer gedetailleerde analyse van de Oct6 promotor is noodzakelijk om de DNA sequenties te identificeren die de cel-type specifieke expressie van het Oct6 gen dicteren. Karakterisering van de eiwitten die een interactie aangaan met deze DNA sequenties zullen ons in staat stellen de mechanismen te bestuderen die de expressie van het Oct6 gen reguleren. Op deze wijze kan een beeld verkregen worden van de plaats die het Oct6 gen inneemt in een hiërarchie van cellulaire regulatie mechanismen welke uiteindelijk leiden tot een volledig gedifferentieerd fenotype.

CURRICULUM VITÆ

De schrijver van dit proefschrift werd geboren op de twaalfde november van het jaar 1958 in de plaats Almkerk. Na een aantal jaren al moest hij naar de lagere school, in Doorn, waar hij vele nuttige dingen leerde zodat hij in 1971 naar het christelijk lyceum 'Revius' mocht. In 1977 behaalde hij daar het einddiploma Atheneum B. Met dit diploma op zak vertrok hij naar Wageningen om iets te gaan studeren aan de Landbouwhogeschool. Dat werd uiteindelijk Moleculaire Wetenschappen. In 1982 werd het kandidaatsexamen behaald en in 1985 het doctoraal examen. De doctoraalstudie omvatte twee hoofdvakken; Kolloidchemie en Moleculaire Biologie. Voor het hoofdvak Kolloidchemie werd onderzoek gedaan, onder leiding van Dr W. Norde, naar de absorptie van het serum albumine eiwit aan geladen ijzeroxide solen. Het hoofdvak Moleculaire Biologie behelsde de isolatie en karakterisering van de replictie startpunten in het DNA van de tomaat. Dit onderzoek werd uitgevoerd onder de bezielende leiding van Dr P. Zabel en Prof. Dr A. van Kammen. De 'ingenieurs stage' werd uitgevoerd in het Laboratory of Gene Structure and Expression van het National Institute for Medical Research in Mill Hill, London. Onder leiding van Dr F. Grosveld en Dr J.-P. Julien werd daar onderzoek verricht naar de structuur en expressie van de neurofilament genen. Na terugkeer in Nederland werd de vervangende militaire dienstplicht vervuld in het instituut Genetica van de Erasmus Universiteit te Rotterdam. In april 1987 volgde een tijdelijke aanstelling als 'onderzoeker' aan het zelfde instituut. Gedurende de periode in Rotterdam werd het in dit proefschrift beschreven werk uitgevoerd onder leiding van Dr G. Grosveld en Prof. Dr D. Bootsma. Sinds october 1991 is schrijver dezes wederom werkzaam in het Laborotary for Gene Structure and Expression te Mill Hill, London, ditmaal financieel gesteund door de 'Bruno Mendel research fellowship' hem toegekend door de Royal Society.

DANKWOORD

Hoewel op de omslag van dit proefschrift slechts mijn naam prijkt zal het duidelijk zijn dat het hier beschreven werk niet door mij alleen is uitgevoerd. Een woord van dank aan al diegenen die direct of indirect aan dit werk hebben bijgedragen, danwel hun interesse hebben getoond is dan ook op z'n plaats. ....., bedankt voor je bijdrage.

Mijn promotor, Professor Dirk Bootsma ben ik zeer erkentelijk voor het in mij gestelde vertrouwen.

Gerard Grosveld, mijn co-promotor, ben ik uiteraard veel dank verschuldigd voor de vele discussies en goede suggesties. Je hebt je altijd zeer onbaatzuchtig opgesteld, mij in alles mijn gang laten gaan, en mij daarin ook gesteund.

Anneke Graus, 'mijn' analiste, ben ik veel dank verschuldigd voor de voorbeeldige, nauwgezette uitvoering van talloze experimenten. Hoewel je tijd in ons lab niet zonder problemen geweest is, kijk ik terug op een goede samenwerking (Okkie & co). Bedankt voor je inzet en loyaliteit.

Robert Kraay, An Langeveld en Maarten Mulder bedankt voor jullie bijdrage aan het Oct6 werk. Ton de Wit, bedankt voor de Oct2 primer extensie experimenten en voor die keren dat je mij met squash hebt laten winnen.

Dank aan alle (ex-) leden van 'voorheen de CML groep' voor hun interesse en de talloze taartjes, stroopwafels, roomsoezen en andere megajoule eetwaren.

Marieke von Lindern, de freule, goede vriendin en wapenbroeder, bedankt voor je nuchtere kijk op de zaken des labs en des levens. Bedankt ook voor die 'deadline' van 10 juni!

Sjozef van Baal, computerwizard, zonder jouw hulp zou ik nu nog een computer analfabeet zijn. Aan jouw inzet is het te danken dat er op een beschaafd niveau met computers gewerkt kan worden binnen de afdeling Genetica (en ver daarbuiten).

Maarten Fornerod en Martine Jaegle bedankt dat jullie mijn paranimfen wilden zijn. Maarten, met jou deelde ik mijn liefde voor de muziek van Tom Waits, Bob Dylan en andere obscure blues muziek, tot wanhoop van andere 'vhCMLers'. Ik hoop dat je de strijd voortzet.

Sandra d'Azzo, bedankt voor je interesse in het werk en je vriendschap. Hans Morreau, bedankt voor je vriendschap, je humor en je mooie verhalen.

Peter Verrijzer ben ik veel dank verschuldigd voor zijn hulp en zijn interesse in het Oct1 t/m 10 werk.

Mirko Kuit en Tom de Vries Lentsch bedankt voor al het fotografische werk. Rita Boucke en Jeannette Lokker, bedankt voor de gezelligheid en het regelen van vliegtickets.

Frank Grosveld en Ernie de Boer, bedankt voor jullie gastvrijheid en vriendschap. Ik vrees dat ik de bijnaam 'den uitvreter' ondertussen wel verdient heb. Frank, bedankt voor je (niet geheel belangeloze) aansporingen om dat proefschrift maar eens te schrijven, en je geduld. En niet te vergeten je aanmoedigingen, zowel in het lab als op het voetbalveld!