Mouse Oct-1 contains a composite homeodomain of human Oct-1 and Oct-2

Noriaki Suzuki[§], Werner Peter⁺, Thomas Ciesiolka[†], Peter Gruss and Hans R. Schöler^{*,†}

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

Received October 19, 1992; Revised and Accepted December 8, 1992 EMBL accession nos X68362–X68364 (incl.)

ABSTRACT

Members of the Oct family of transcription factors specifically interact with the octamer motif, ATGCAAAT, a regulatory element important for tissue- and cell-specific transcription as well as for the expression of housekeeping genes. Except for Oct-1, all Oct factors are expressed in a temporally and spatially restricted mode during murine development and their number varies in a given cell type. Despite its ubiquitous expression pattern Oct-1 may play a role in murine development. As a first step towards elucidating the role of Oct-1 we report the complementary DNA cloning of the mouse *Oct-1* gene. Two large transcripts of 5 and 14 kb are derived from a single gene. The expression patterns of three splicing products of Oct-1 are similar in a number of cells and tissues. In the POU region murine Oct-1 differs in four amino acids from the human homologue and these differences are restricted to helices 1 and 2. Interestingly, two of the four variant amino acids are identical to those in human and mouse Oct-2 and thus the murine Oct-1 homeodomain is intermediary in sequence between human Oct-1 and Oct-2. These two amino acids together with a third one have been shown to be relevant for the interaction between human Oct-1 and herpes simplex virus transactivator VP16. Nevertheless, VP16 interacts albeit weakly with murine Oct-1. We speculate that the differences in the human and mouse Oct-1 homeodomains reflect host-specific differences in protein–protein interactions.

INTRODUCTION

Transcriptional regulation depends primarily on the sequence-specific interplay of *trans*-acting factors with *cis*-acting regulatory elements. The octamer or decamer motif (ATGCAAATNA) or its complement (TNATTTGCAT) and closely related sequences provide a paradigm of *cis*-acting elements involved in both ubiquitous and cell-type specific gene expression (reviewed in 1, 2). The octamer is necessary and sufficient for B-lymphocyte-

specific gene expression of the immunoglobulin genes (3, 4, 5). On the other hand, the octamer motif is a functional element involved in the ubiquitous expression of genes, such as the thymidine kinase gene (6), the snRNA genes (7–11) or histone H2B (12, 13). We have shown that a family of murine proteins interact with the octamer motif (14), raising the possibility that a large set of genes is regulated by distinct octamer-binding proteins or Oct factors in a tissue-specific manner (reviewed in 2). The use of the octamer motif in promoting both ubiquitous and tissue-specific gene expression may be a consequence of selective binding of ubiquitously expressed Oct-1 or by specifically expressed Oct factors.

All cloned Oct factors contain two highly conserved domains which are separated by 14–26 variable amino acids, a POU homeodomain and a POU-specific domain (2, 15). Both are required for DNA binding and are involved in protein–protein interactions. One open issue is the selectivity and specificity of the various Oct factors. It is possible that although Oct factors can activate transcription via the octamer motif, their optimal recognition sequence may encompass more than the octamer or may even be a different motif. Several parameters could influence the affinity of a certain Oct factor for a given recognition sequence. These include the presence and position of other DNA binding proteins, or factors that do not directly interact with DNA but help to stabilize Oct factor binding by generating a multiprotein complex. The combination of promoter elements might explain why the octamer motif can mediate transcriptional activation or repression in the same cell type, as shown for the embryonal carcinoma cell line F9 (16, 17).

A well documented case where an Oct factor interacts with a sequence element that fits only poorly the octamer sequence is so-called TAATGARAT motif. Together with at least one other cellular factor, the Oct-1 protein is recruited by the Herpes simplex Virus (HSV) protein VP16 (also termed Vmw65 or α -TIF) into a multiprotein complex on the TAATGARAT motif which stimulates the immediate early promoters (18–22). The interaction between Oct-1 and VP16 is specific for Oct-1, as the closely related Oct-2 fails to form a detectable complex with VP16. Sequences critical for the interaction with VP16 were

* To whom correspondence should be addressed

Present addresses: ⁺EMBL, Gene expression programme, Meyerhofstr.1 6900 Heidelberg, Germany, [§]Department of Internal Medicine, Saga Medical School, Nabeshima 5-1-1, Saga 849, Japan and [†]Boehringer Mannheim, 8122 Penzberg, Nonnenwald, Germany

identified within the homeodomain of Oct-1 by analyzing Oct-1/Oct-2 chimeras (18). In the absence of VP16, Oct-1 and Oct-2 also differ in their ability to activate transcription, even with respect to the same recognition sequence (23, 24).

We are interested in investigating whether Oct-1 activates developmentally regulated genes or whether its sole function is to stimulate the expression of housekeeping genes as suggested by its ubiquitous expression pattern. To assess the role of Oct-1 during murine development, we cloned different mouse Oct-1 cDNAs and examined the expression patterns of three splicing products. Sequence comparison and Southern blotting indicate that the splicing products are derived from a single *Oct-1* gene. Two large transcripts of 5 and 14 kb are detectable by Northern blot analysis. Each of the proteins when expressed from the encoding cDNAs by *in vitro* transcription and translation binds to the octamer motif. In the POU region the murine Oct-1 protein differs in four amino acids from human Oct-1 and these differences are restricted to helices 1 and 2. Two of the four different amino acids between murine and human Oct-1 have been shown to stabilize the interaction between human Oct-1 and VP16. We have assessed whether murine Oct-1 can interact with

the program Clustal V (27). All other methods were described previously: the preparation of the F9 cDNA library in (28), the cell lines, extract preparation and gel shift assay in (17, 29), the *in vitro* translation in reticulocyte lysates in (28). Before transcription the *Oct-1* cDNA clones were digested with XhoI to obtain sense RNA with T3 RNA polymerase. Northern blot analysis and RNase protection assays were performed as in (28, 30), Southern blot analysis is described in (29). For RNase protection assays and the Northern blot analysis the EcoRI-StuI and the StuI-SacI fragments of Oct-1A were cloned into EcoRI-SmaI and SmaI-SacI of Bluescript SK, respectively. To generate probes both clones were linearized with XhoI. This resulted in probes with a heterologous sequence containing 95 additional nucleotides in case of the 590 nt EcoRI-StuI fragment, and 53 additional nucleotides in case of the 556 nt StuI-SacI fragment.

RESULTS

Identification of cDNA clones encoding Oct-1

A cDNA library prepared from F9 cells was screened with a probe spanning the mouse *Oct-2* POU domain (31). Eighteen positive clones were isolated that fell into three groups according to their physical maps. One group turned out to encode Oct-4 and Oct-5 (28), another group Oct-6 (30). The third group, composed of five clones (Figure 1A), was further analyzed by sequential *in vitro* transcription, translation and finally tested for DNA-binding in the gel shift assay (Figure 1B). The products of the three longest cDNAs generated a band with a slightly higher

Figure 1. Schematic representation of different murine Oct-1 cDNA clones. **A:** Structure of three Oct-1 cDNAs. The physical maps of five cDNAs revealed that they can be subdivided into three groups, Oct-1A, Oct-1B and Oct-1C. The black boxes represent exons in the larger transcripts and the numbers stand for their sizes. The POU specific and the POU homeo domains, as well as several restriction enzymes used for the physical map, are indicated in the figure. The clones in group 1A (3, 6) do differ only in their 5' ends (clone 3 is about 100 bp longer than 6); group 1B (8, 11) differ from group 1A by a 72 bp deletion. Whereas clone 8 has almost the same 5' terminus as clone 3, clone 11 has a large truncation at the 5' end. Group 1C has about the same 5' end as clones 3 and 8, but has a 386 bp deletion at the 3' end. **B:** Comparison of the products of the five F9 cDNA clones with F9 and BJA-B octamer binding proteins in the gel shift assay. Radiolabeled octamer probe (1W) was incubated before electrophoresis with a BJA-B (lane 1) or F9 whole cell extract (lane 2), or with *in vitro* translated products of clones 11, 4, 8, 6, or 3 (lanes 3 to 7). Due to a truncation of the 5' end, clone 11 results in a product of higher mobility compared to the products of the other cDNA clones (lane 3). BJA-B is a human B-cell line; F9 is a mouse embryonic teratocarcinoma cell line. **C:** Comparison of the 5' ends and the respective N-termini of clones 4, 3 and 8 with the 5' end of human Oct-1. The N-terminus of a recently published mouse Oct-1* (from the cell line NS/O) is included and differences in protein sequence (Oct-1*) are indicated with italics. The accession numbers to the sequences of the three cDNAs are: Oct-1A: X 68362, Oct-1B: X 68363, Oct-1C: X 68364.

mobility than murine and human Oct-1 (compare lanes 5 to 7 with lanes 1 and 2), whereas the products of the smaller cDNAs had approximately the mobility of Oct-2A and Oct-2B, respectively (compare lanes 3 and 4 with lane 1).

The five cDNAs were subcloned into M13 vectors in both orientations and single-stranded DNA was used to determine the sequences. Nucleotide sequence comparison revealed three different kinds of cDNAs: Oct-1A, Oct-1B and Oct-1C (Figure 1A). In addition to varying 5' ends due to the cloning procedure, the cDNAs differed with respect to their exon composition at their 3' ends: both Oct-1B and Oct-1C lack a 72 bp exon found in Oct-1A, and Oct-1C lacks in addition a 386 bp exon.

Expression pattern of three Oct-1 splicing products

To confirm that the different cDNAs were derived from one gene, a restriction fragment common to all cDNAs was used in a Southern blot analysis (Figure 2A). One prominent band was obtained with either BamHI or PstI digested mouse DNA, indicating that the different cDNA clones are indeed derived from one gene and thus represent alternative splicing products.

To determine the total size of the different mRNA species a Northern blot analysis was performed by probing F9 poly A+ RNA with a 3' probe common to the cDNAs. Two transcripts were identified: a major species of 5 kb and a minor one of 14 kb (Figure 2B). To test the quality of the RNA the filter was reprobbed with Oct-4 which yielded a very strong signal after eight hours. The tremendous difference in the RNA signal of both genes is in striking contrast to the approximately equal numbers of clones obtained from the F9 cDNA library which was made from the same RNA preparation. We cannot explain this difference although it is possible that splicing accounts for some reduction of the signal. We would also like to note that although a number of groups, including ours, have published Oct-2

Figure 2. Southern and Northern blot analyses of the Oct-1 gene. **A:** DNA of three different murine strains (Dbal2J: lanes 1,4; C57: lanes 2, 5; Ak: lanes 3, 6) cut with either BamHI (lanes 1 to 3) or PstI (lanes 4 to 6) and hybridized with a radioactive Oct-1C probe containing the 5' cDNA region up to the first PstI site (EcoRI/PstI). The EcoRI site is from the linker which was used for the generation of the cDNA library. **B:** Oct-1 transcripts were detected after hybridizing a radioactive Oct-1A Stul/SacI probe to a blot with 5 μ g F9 poly(A+) RNA. This probe contained sequences complementary to all three splicing products. The lengths of the fragments and the transcripts are indicated in the figure.

Northern blots showing several mRNAs, so far no Oct-1 Northern analysis has been reported.

Due to the large transcript sizes, the Northern blot analysis was too insensitive to distinguish mRNAs that differ only in the small 72 or 386 nt exons. For this reason the expression pattern of the *Oct-1* gene was further examined by RNase protection

Figure 3. Expression pattern of different Oct-1 splicing products in various cells and tissues of the mouse. **A:** autoradiograph of an RNase protection experiment with both probes (590 and 556 nt) and RNA from undifferentiated (F9 UD) and differentiated F9 (F9 D) cells is shown. A weak band detectable in the original at the position of the 458 nt is indicated by an arrow. 590 and 556 nt indicate the homology of the probes to the Oct-1A cDNA. Due to subcloning of the cDNA into Bluescript the probes actually contain heterologous sequences and thus unspliced mRNA protecting the complete Oct-1A fragment could be distinguished from incompletely digested probe (see Materials and Methods). **B:** autoradiograph of an RNase protection experiment with the 590 nt probe and RNA from F9 cells and various tissues is presented. The lengths of the probes and the products of the assay are indicated at the right of the panels and are outlined in **C**. **C:** The top line represents that the part of the Oct-1A transcript which differs from Oct-1B and Oct-1C. In the middle, the probes and their lengths are presented, the black box standing for the 72 nt and the stippled box for part of the 386 nt exon. The expected products of the RNase protection assay are summarized at the bottom. The numbers of the probes indicate the lengths corresponding to Oct-1, actually they are longer as a result of the *in vitro* transcription procedure. RNase protection experiments were performed using total RNA and radioactive antisense RNA probes. The RNA source is indicated above each lane. 30 μ g total RNA was used in the experiment, except for lanes F9 (UD) and F9 (D) where only 5 μ g F9 RNA was used.

m Oct-1A	MNNPSETNKSMESEDASTGTQTNGLDFOKQVPVVGGAISTAQAQAFGLHQLVQLAGTSLQAAAQSLNVQSKSS	75
m Oct-1B	.....	75
m Oct-1C	.....	75
m Oct-1*	.....L.....	57
h Oct-1	.....S.P.....G.GN.....N	75
m Oct-1A	EESGDSQSSSQPSSQPPSVQSAIPQTQLMLAGGQITGLTLTPAQQLLLQQAQAQLLAAAVQQHSASQHSAA	150
m Oct-1B	.....	150
m Oct-1C	.....	150
m Oct-1*	.....	132
h Oct-1	.....P.....Q.....A.....	149
m Oct-1A	GATISASAATPMTQIPLSQPIQIAQDLQQLQQLQQNLNLQQFVLVHPTTNLQPAQFIISQTPQGGQGLLQAQNL	225
m Oct-1B	.....	225
m Oct-1C	.....	225
m Oct-1*	.....	207
h Oct-1	.....	224
m Oct-1A	LTQLPQQSQANLLQPPSITLTSQPTTPTRTIAAASVQTLQSQSTPKRIDTPSLEEPSDLLEELEQFAKTFKQRR	300
m Oct-1B	.....	300
m Oct-1C	.....	300
m Oct-1*	.....	282
h Oct-1	Q.....S.....A.....TPI.....	299
POU specific domain		
m Oct-1A	IKLGFQGDVGLAMGKLYGNDFSQTTISRFEALNLSFKNMCKLKPILLEKWLDAENLSSDSTASSPSALNSPGLG	375
m Oct-1B	.....	375
m Oct-1C	.....	375
m Oct-1*	.....	357
h Oct-1	.....S.....S.....SL.....I-	373
r Oct-1	.....S.....S.....L.D.....	
POU homeo domain		
m Oct-1A	AEGLNRRRKRTSIETNIRVALEKSFMEQKPTSEDLTIAEQLNMEKEVIRVWFCNRRQKEKRINPPSSGGTSS	450
m Oct-1B	.....	450
m Oct-1C	.....	450
m Oct-1*	.....	432
h Oct-1	.....S.....L.....E..M..D.....	447
r Oct-1	.....WS.....	
m Oct-1A	SPIKAIFPSPASLVATPSPSLVTSSTATTLTVNPVPLTSAAVTNLSLTDQDLRRGCSWEVLRLSLPDRVTTAGTT	525
m Oct-1B	.....	501
m Oct-1C	.....	501
m Oct-1*	.....	483
h Oct-1	.....TS.....A.....S.....V.....S	498
m Oct-1A	DSTNNNTATVISTAPPASSAVTSPSLSPSPSASASTSEASSASETNTTQTSTPLSPGLASQVMVTTGLQTA	600
m Oct-1B	.....	576
m Oct-1C	.....	576
m Oct-1*	.....	558
h Oct-1	.....T.....S.....S.....T.....AS.....	572
m Oct-1A	AAA-LQGAGQLPANASLAAMAAAAGLSPGLMAPSQFAAGGALLSLSPGTLSALSALMSNSTLATIQALASSGS	674
m Oct-1B	.....	650
m Oct-1C	.....	650
m Oct-1*	.....*	596
h Oct-1	.....A.....A.....N.S.....N.....SG.....G.....	647
m Oct-1A	LPITSLDATGNLVPANAGGAPNIVTAPLFLNPQLSLTSPVSLVSAAAASTGNSAPTASLHASSTSTESIQQS	749
m Oct-1B	.....	725
m Oct-1C	.....	692
h Oct-1	.....A.....V.....T.....A.....N	722
m Oct-1A	LFTVASASGPASTTTAASKAQ*	770
m Oct-1B	.....*	746
m Oct-1C	.....DCFMDWRTF*	701
h Oct-1	.....A.....T.....*	743

Figure 4. Comparison of the predicted amino acid sequences of the three murine Oct-1 cDNA groups with Oct-1 from human origin. The POU region of rat Oct-1 has been partially determined and is included in the comparison (34). The numbering system shown at the right of the amino acid sequences (in single-letter code) begins at the same methionine as published for the human Oct-1 cDNA clone (13). The POU specific domain, the POU homeo domain and the lengths of the proteins are indicated in the figure. Points stand for residues equal with respect to murine Oct-1A, dashes for deletions, and asterisks for stop codons. A previously published murine Oct-1 sequence is included in the comparison (Oct-1*; 34). The human cDNA sequence was published in (13).

assay. The probes were tested first with RNA from undifferentiated F9 cells to confirm the identity of the reactive RNAs as authentic transcripts of the *Oct-1* gene. Subsequently, RNAs of several tissues of adult mice and day 13 embryos were examined.

Uniformly labelled RNA was derived from two fragments of Oct-1A with which the products of the suggested different splicing could be separately analyzed (Figure 3C). The first probe spans the 72 nt insertion and only in the case of Oct-1A related transcript(s) should give rise to a 590 nt band. Whenever the 72 nt had been deleted, as with Oct-1B and C, the probe should be degraded to 389 and 129 nt fragments. Indeed, all three protected fragments are easily detectable with RNA from F9 stem

cells (Figure 3A lanes 3 and 4). The 389 and 129 nt bands are stronger than the 590 nt band indicating that Oct-1 lacking the 72 nt insertion is the more abundant mRNA.

We have previously shown by gel retardation analyses that upon differentiation only the Oct-1 complex remains constant, whereas the Oct-4, Oct-5 and Oct-6 complexes decrease (17). To test whether these splicing products described above could also be detected after differentiation, we compared RNA from differentiated and undifferentiated F9 cells in the RNase protection assay (Figure 3A lanes 3 and 4). The same pattern and the same relative band intensity was obtained indicating that the Oct-1 gene is not downregulated and the 72 nt insert is not differentially spliced upon F9 differentiation. When the RNase

Figure 5: C-termini of the murine Oct-1 proteins. **A:** Sequences of the 3' end of the different Oct-1 cDNAs and the respective C-termini of Oct-1A/B and Oct-1C. The region deleted in the Oct-1C cDNA is indicated with a double-arrow, the splice acceptor and donor sites as well as the lariat consensus sequence are underlined. The stop codons are highlighted with three asterisks each. **B:** Amino acid comparison of the C-termini of Oct-1A and Oct-2B. The last 122 amino acids (649 to 770) of Oct-1A and the last 127 amino acids (457 to 583) of Oct-2B are aligned. Small gaps were introduced into the Oct-1A sequence for optimal alignment. Conservative changes (threonine to serine; leucine to isoleucine) are also considered.

protection assay was done using day 13 embryos and isolated adult tissues, the 389 and 129 nt bands were again the predominant products (Figure 3B). Because the boundaries of the 72 nt insert do not show any homology to known splice acceptor or donor sites the insert is most probably not a non-spliced intron.

The second probe should give rise to a 556 nt band in the case of Oct-1A and B related transcript(s) and a 458 nt band in the case of Oct-1C when the very 3' end has been altered due to splicing. With this probe a strong 556 nt band is obtained, whereas the 458 nt band is only weakly detectable and may be due to degradation of the probe (Figure 3A lanes 5 and 6). Although the probe was totally degraded when mixed with tRNA (not shown), the RNase protection assay does not prove that this splicing product actually exists. When the same tissues examined above were tested, only the undegraded 556 nt band was obtained, indicating that the Oct-1C mRNA represents a minor Oct-1 transcript (data not shown). The boundary sequences of the deleted 386 nt region and the lariat consensus sequence indicate that the faint band in the RNase protection assay is a splicing product and the deletion is not due to a cloning artifact (see Figure 5A). The 3' splice site (TCTTTTCAG G) matches perfectly with the consensus ((Py)_nNPYAG G), as well as the CTAAT sequence with the lariat consensus (reviewed in 32, 33). The 5' consensus (A/C)AG GTAAGT and the Oct-1 donor site CCA GTAAGC match well on the intron side, but the exon sequence is CCA (in human Oct-1: CCT) instead of CAG and might explain why splicing is not very efficient at this site. Although the 3' splice site and the putative lariat sequence are in the 3' UTR of Oct-1A/B, these sequences are identical in the human Oct-1 cDNA (13).

Taken together, at least three different Oct-1 transcripts are detectable in the mouse, and are spliced the same in all tissues examined. In all the cells and tissues tested Oct-1B represents by far the most predominant transcript, containing the 386 nt exon and lacks the 72 nt exon.

Structure of the Oct-1 protein

The sequence of the Oct-1A cDNA as derived from the longest open reading frame encodes a protein of 770 amino acids (Figure 4). Due to different lengths of the C-terminus, Oct-1B and Oct-1C are smaller, only contain 746 and 701 amino acids, respectively. Compared to Oct-1B and Oct-1C, Oct-1A contains 24 additional amino acids in the C-terminus proximal to the POU homeodomain (Figures 1A and 4). The 72 bp deletion in Oct-1B and Oct-1C does not change the open reading frame but loss of the 386 bp exon resulted in an altered reading frame at the C-terminus (Figure 5A; see also Figure 1, lane 4). As a consequence, the last 54 amino acids of Oct-1A and B are replaced by a different stretch of 9 amino acids in Oct-1C. The C-terminus of Oct-1A/B is rich in serine, alanine and threonine residues, representing 65% of this region. The different C-termini are reminiscent of differential splicing of the Oct-2 gene which generates different C-termini in Oct-2A and Oct-2B (31). Interestingly, the C-terminus of Oct-1A/B shows a high degree of homology with that of Oct-2B (Figure 5B).

In Figure 4 the protein sequences of Oct-1A, B and C as derived from the respective cDNAs are compared with the human Oct-1 sequence (13) and a murine Oct-1 sequence published recently (35). The DNA sequences of murine Oct-1B and human Oct-1 are 93% identical, indicating that the respective mRNAs are from genes that are homologous in both organisms. This is also indicated by the Southern blot analysis (Figure 2A). In addition, comparison of the proteolytic intermediates caused by digestion with different amounts of the non-specific protease dispase, reveals that the pattern of the cloned proteins is very similar to that of HeLa Oct-1 (data not shown).

Interaction of murine Oct-1 and VP-16

Mouse Oct-1 differs in the homeodomain in four amino acids from human Oct-1 and these differences are again restricted to helices 1 and 2 (Figures 4 and 6B). VP16 forms a stable complex with human Oct-1, but not with the related Oct-2 protein. Three of the seven amino acids differing between helix 1 and 2 of the Oct-1 and Oct-2 homeodomains are involved in the interaction with VP16 (ref. 18; indicated with arrows in Figure 6B). Interestingly, two of the four residues of murine Oct-1 that differ from human Oct-1 are identical to residues in Oct-2 and both are divergent at positions critical for VP16 binding. Since the murine Oct-1 homeodomain has an intermediate sequence between human Oct-1 and Oct-2 we tried to determine if murine Oct-1 can form a complex with VP16.

A DNA fragment containing the TAATGARAT sequence was incubated with cellular extract of human or murine origin and the complexes formed were analysed in the gel shift assay. With both extracts, Oct-1 complexes of the same intensity were obtained (Figure 6A, lanes 1 and 3). In parallel, both extracts were incubated with bacterially expressed VP16 and yield an additional band with lower mobility indicating that VP16 can form a complex with both proteins, although less efficiently with mouse Oct-1 (lanes 2 and 4).

Figure 6: Formation of VP-16 induced complexes with murine and human Oct-1. **A:** Gel shift assay: either HeLa (lanes 1 and 2) or murine liver extracts (lanes 3 and 4) were incubated with (lanes 2 and 4) or without (lanes 1 and 3) bacterial expressed VP-16 and the ICPO TATGARAT probe. **B:** Comparison of homeodomain protein sequences of murine and human Oct-1 and Oct-2. **C:** Schematic diagram of view along the axis of helix 2 of hOct-1, mOct-1 and Oct-2. The circled residues indicated by arrows in B represent the differences in comparison with hOct-1. The ICPO probe contains a TATGARAT homology from the promoter of the ICPO gene of HSV and is described in (18).

DISCUSSION

In this report we describe the isolation of cDNAs encoding three different murine Oct-1 proteins and their patterns of expression. The analysis of five independent clones indicates that in murine cells a single gene gives rise to at least three different mRNAs. Oct-1B cDNA encodes the equivalent of human Oct-1 and appears to be the predominant Oct-1 product in the mouse. The second, mOct-1A, encodes a larger variant of 770 amino acids and the third, mOct-1C, has a deletion at the 3' end which creates a shorter open reading frame distinct from Oct-1A and B. We have tested RNA from sources representing early and late embryonal stages, and also from various tissues of the adult mouse for the relative abundance of the different splicing products. In this study we could not detect any difference in their expression pattern. It is possible that other cells express a different repertoire of transcripts and that additional splicing products may be found when probes spanning other parts of the Oct-1 gene are used in the RNase protection assay. The Oct-1 clones did not show any alterations at their 5' end. However, splicing might also occur at the 5' region as indicated by a previously published mouse Oct-1 cDNA clone (Oct-1* in Figure 4; 35). In the Oct-1* cDNA, 54 nt are deleted and thus the protein lacks 18 amino acids in the predicted sequence from position 54 to 71 (Figure 4). A further comparison with this Oct-1* clone shows more differences which are mainly restricted to the termini.

What is the physiological relevance of different Oct-1 splicing products? The octamer motif is an eukaryotic regulatory element found in promoters and enhancers, as well as in replication origins. Furthermore, Oct-1 and Oct-2 proteins can stimulate DNA replication *in vitro* and, as shown for Oct-1, the DNA binding domain suffices (36). Thus it is possible that the various transcripts encode translation products which have different functions in transcription and replication. It is worth noting that

different CTF/NF1 polypeptides are also encoded by alternatively spliced mRNAs and have transcription and replication activities (37).

The murine Oct-1 homeodomain has a composite sequence of human Oct-1 and Oct-2

Oct-1 binds to the octamer and to the TAATGARAT motif. However, only binding to the latter motif enables the virally encoded transcription factor VP16 to interact stably with Oct-1 (reviewed in 21, 22). VP16 contains a potent C-terminal acidic activation domain that strongly activates HSV immediate-early (IE) and related promoters (reviewed in 21, 22). In the absence of VP16, Oct-1 does not efficiently activate these promoters and thus VP16 acts as a transcriptional coactivator that alters the promoter selectivity of Oct-1, allowing stimulation of a distinct set of genes.

The interaction between Oct-1 and VP16 is specific for Oct-1, as the closely related Oct-2 does not form a detectable complex with VP16. As indicated in Figure 6C, three residues which contribute to the interaction with VP16 can be placed at the same side of α -helix 2. When these three residues are as in the natural Oct-2 (L E L), VP16 binding is undetectable. Binding is residual when in Oct-1 helix 2 is replaced by that of Oct-2. With the reciprocal construct, in which Oct-2 contained the helix 2 of Oct-1, binding was still undetectable, indicating that the three different residues in helix 2 are necessary but not sufficient for complex formation (18).

To get an idea which amino acids of the POU region are important for an interaction with VP-16, human and murine Oct-1 sequences are first compared and then both murine Oct-1 and Oct-2 sequences. Murine Oct-1 differs in the POU homeodomain in four amino acids from human Oct-1 and the differences are restricted to helices 1 and 2. Two of the four different amino

acids are identical to those found in Oct-2 and thus the murine Oct-1 homeodomain represents a sequence intermediary between human Oct-1 and Oct-2. Mouse Oct-1 associates less efficiently with VP-16 than human Oct-1, since the intensity of the murine Oct-1/VP16/DNA complex was significantly lower. The differences in helix 2 (Fig. 6C), or the leucine to methionine change in helix 1, might account for the reduction of VP16-induced complex formation. Of the three differences in helix 2, two represent conservative changes (glutamate to aspartate, and vice versa), and the third a methionine to leucine change. Although the conservative changes might weaken complex formation due to different spacing between VP-16 and Oct-1, the non-conservative change is considered to be the more important one. In this case, the methionine residue in helix 2 of human Oct-1 would be directly involved in VP-16 interaction.

When murine Oct-1 and Oct-2 are compared, the most striking differences are the valine to phenylalanine and the glutamate to alanine changes in helix 1 and the threonine to leucine change in helix 2. Since the helix 2 of human Oct-1 in Oct-2 was not sufficient for complex formation, the valine and/or the glutamate appears to be critical for an interaction with VP-16. Our experiments do not exclude that poor complex formation is due to a weak interaction between VP-16 and cellular components (variously termed C1, HCF, CFF; for review see 22).

Although man is the natural host for herpes simplex viruses, rodents are susceptible to infection and thus an interaction of VP16 with murine Oct-1 is likely to occur. However, due to the different stabilities of the VP16/Oct-1 complexes, the course of virus infection in mice may differ from that in humans. Nevertheless, it has been found that not only does HSV replication occur in mice, but that alteration of VP-16 severely attenuates the virus, so that such a mutated virus is essentially avirulent when injected into mice (38). Mice transgenic for human Oct-1 might provide an appropriate test system to analyse differences in HSV infection.

Differences in the homeodomains of murine and human Oct-1 might reflect host-specific differences of protein-protein interactions

What conclusions can be drawn from the observed differences between murine and human Oct-1? It might be useful to first consider which parts of the POU region are conserved when different POU factors are compared (2). The most conserved parts are the POU specific domain and helix 3 of the POU homeodomain. This conservation is obvious when the amino acid sequences of all known POU factors are compared and is even more apparent when the comparison is restricted to Oct-1 and Oct-2. Both the POU specific domain and helix 3 are identical in mouse and human Oct-1, as they are between murine and human Oct-2, and only one conservative change is found in these regions when Oct-1 and Oct-2 are compared (Fig. 4; 15, 31). The homology in these two parts of the POU domain might reflect their involvement in conserved functions. Such a function is known for helix 3, which is considered to be the major determinant in DNA binding and most of the known POU proteins can bind to very similar DNA binding sites. However, it is unclear why the POU specific domain has been so highly conserved. A hint may be given by experiments demonstrating that Oct-1 is involved in DNA replication and that the Oct-1 POU region promotes adenovirus DNA replication *in vitro* (35). Thus it is possible that the POU specific domain interacts

with the replication machinery which indeed contains highly conserved components. According to this idea, specificity in transcription and replication could come from the interaction of cell-specific proteins with the sequence containing helix 1 and 2, because it represents the least conserved region of both POU domains (2). The crystal structures of the engrailed and MAT α 2 homeodomains indicate that helices 1 and 2 can be regarded as a helix-loop-helix and helices 2 and 3 as a helix-turn-helix motif (39, 40). The differences between Oct-1 and Oct-2 might then be due to different cellular factors interacting with the respective helix-loop-helix, and VP16 could mimic a cellular factor interacting with only the Oct-1 helix-loop-helix. Differences in the helix-loop-helix of murine and human Oct-1 then would reflect host-specific differences in protein-protein interactions.

ACKNOWLEDGEMENTS

We would like to thank Christiane Peiker for help with the experimental work and Winship Herr for kindly providing VP16 and the plasmid containing the ICPO probe. Susan L. Palmieri, Catherine Ovitt and Henk Stunnenberg for critical reading the manuscript. We also thank an anonymous reviewer for thoughtful comments on the VP-16/Oct-1 interaction. N.S. was a recipient of the Behrens-Weise Foundation, T.C. was a recipient of a fellowship of the Deutsche Forschungsgemeinschaft (DFG). This project was supported by the Max-Planck Society and by the Bundesministerium für Forschung und Technologie (BMFT).

REFERENCES

- Schreiber, E., Müller-Immerglück, M.M., Schaffner, W. and Matthias, P. (1988) In Renkawitz, R. (ed.), *Tissue Specific Gene Expression*. Verlag Chemie, Weinheim, pp. 33–55.
- Schöler, H.R. (1991) *Trends Genet.*, **7**, 323–329.
- Gerster, T., Matthias, P., Thali, M., Jiricny, J. and Schaffner, W. (1987) *EMBO J.*, **6**, 1323–1330.
- Lenardo, M., Pierce, J.W., and Baltimore, D. (1987) *Science*, **236**, 1573–1577.
- Wirth, T., Staudt, L. and Baltimore, D. (1987) *Nature*, **329**, 174–178.
- Parslow, T.G., Jones, S.D., Bond, B., and Yamamoto, K.R. (1987) *Science*, **235**, 1498–1501.
- Mattaj, J.W., Lienhard, S., Jiricny, J., and deRobertis, E.M. (1985) *Nature*, **316**, 163–167.
- Mangin, M., Ares, M., Weiner, A.M. (1986) *EMBO J.*, **5**, 987–995.
- Ares, M., Chung, J., Giglio, L., and Weiner, A.M. (1987) *Genes Dev.*, **1**, 808–817.
- Carbon, P., Murgo, S., Ebel, J.P., Krol, A., Tebb, G., and Mattaj, J.W. (1987) *Cell*, **51**, 71–79.
- Tanaka, M., Grossniklaus, U., Herr, W., and Hernandez, N. (1988) *Genes Dev.*, **2**, 1764–1778.
- La Bella, F., Sive, H.L., Roeder, R.G., Heintz, N. (1988) *Genes Dev.*, **2**, 32–39.
- Sturm, R.A., Das, G. and Herr, W. (1988) *Genes Dev.*, **2**, 1582–1599.
- Schöler, H.R., Hatzopoulos, A.K., Balling, R., Suzuki, N. and Gruss, P. (1989) *EMBO J.*, **8**, 2543–2550.
- Herr, W., Sturm, R.A., Clerc, R.G., Corcoran, C.M., Baltimore, D., Sharp, P.A., Ingraham, H.A., Rosenfeld, M.G., Finney, M., Ruvkun, G. and Horvitz, H.R. (1988) *Genes Dev.*, **2**, 1513–1516.
- Lenardo, M.J., Staudt, L., Robbins, P., Kuang, A., Mulligan, R.C., and Baltimore, D. (1989) *Science*, **243**, 544–546.
- Schöler, H.R., Balling, R., Hatzopoulos, A.K., Suzuki, N. and Gruss, P. (1989) *EMBO J.*, **8**, 2551–2557.
- Stern, S., Tanaka, M., and Herr, W. (1989) *Nature*, **341**, 624–630.
- Katan, M., Haigh, A., Verrijzer, C.P., Van der Vliet, P.C., O'Hare, P. (1990) *Nucl. Acids Res.*, **18**, 6871–6880.
- Kristie, T.M. and Sharp, P.A. (1990) *Genes Dev.*, **4**, 2383–2396.
- Goding, C., and O'Hare, P., (1989) *Virology*, **173**, 363–367.
- Thompson, C.C. and McKnight, S.L. (1992) *Trends Genet.*, **8**, 232–236.

23. Tanaka,M., and Herr,W. (1990) *Cell*, **60**, 375–386.
24. Tanaka,M., and Herr,W. (1992) *Cell*, **68**, 755–767.
25. Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Press, Cold Spring Harbor NY.
26. Hogan,B., Constantini,F., and Lacy,E. (1986) *Manipulating The Mouse Embryo: A Laboratory Manual*. Cold Spring Harbor Press, Cold Spring Harbor NY.
27. Higgins,D.G., Bleasby,A.J., and Fuchs,R. (submitted to *CABIOS*)
28. Schöler,H.R., Ruppert,S., Suzuki,N., Chowdhury,K. and Gruss,P. (1990A) *Nature*, **344**, 435–438
29. Schöler,H.R., Dressler,G.R., Balling,R., Rohdewohld,H. and Gruss,P. (1990B) *EMBO J.*, **9**, 2185–2195
30. Suzuki,N., Rohdewohld,H., Neumann,T., Gruss,P. and Schöler,H.R. (1990) *EMBO J.*, **9**, 3723 –3732
31. Hatzopoulos,A.K., Stoykova,A.S, Erselius,J.R., Golding,M., Neuman,T., and Gruss,P. (1990) *Development*, **109**, 349–362.
32. Senepathy,P. (1988) *Proc.Natl.Acad.Sci. U.S.A.*, **85**, 1129–1133.
33. Green,M.R. (1991) *Annu.Rev.Cell.Biol.*, **7**, 559–599.
34. Lillycrop,K.A., and Latchman,D.S. (1991) *Nucl.Acid.Res.*, **19**, 3744.
35. Stepchenko,A.G. (1992) *Nucl.Acid.Res.*, **20**, 1419.
36. Verrijzer,C.P., Kal,A.J., and Van der Vliet,P.C. (1990) *EMBO J.*, **6**, 1883–1888.
37. Santoro,C., Mermod,N., Andrews,P.C., and Tjian, R. (1988) *Nature*, **334**, 218–224.
38. Ace,C.I., McKee,T.A., Ryan,M., Cameron,J.M., and Preston,C.M. (1989). *J.Virol.*, **63**, 2260–2269.
39. Kissinger,C.R., Liu,B., Martin-Blanco,E., Kornberg,T.B., and Pabo,C. (1990) *Cell*, **63**, 579–590.
40. Wolberger,C., Vershon,A.K., Liu,B., Johnson,A.D., and Pabo,C.O. (1991) *Cell*, **67**, 517–528.