

MOD 00042

Research Papers

Structure, expression and chromosomal location of the *Oct-4* gene

Young Il Yeom¹, Hae-Sook Ha^{1,*}, Rudi Balling², Hans R. Schöler² and Karen Artzt¹

¹ Department of Zoology, The University of Texas at Austin, Austin, Texas, U.S.A.;

² Department of Molecular Cell Biology, Max Planck Institute of Biophysical Chemistry, Göttingen, F.R.G.

(Received 11 March 1991; revision received 29 April 1991; accepted 30 April 1991)

The map position of *Oct-4* on mouse chromosome 17 is between *Q* and *T* regions in the Major Histocompatibility Complex (MHC), and it is physically located within 35 kb of a class I gene. Several *Oct-4*-related genes are present in the murine genome; one of them maps to chromosome 9. The genomic structure and sequence of *Oct-4* determined in *t*-haplotypes reveals five exons, and shows no significant changes in the *t*¹² mutant haplotype making it unlikely that *Oct-4* and the *t*¹² early embryonic lethal are the same gene. By in situ hybridization, detectable onset of zygotic *Oct-4* expression does not occur until compaction begins at 8-cells, suggesting that there might be other regulatory factors responsible for initiating *Oct-4* expression.

Oct-4; *t*-complex; Mouse embryology

Introduction

The POU family has recently been defined as a group of related transcription factors containing a bipartite DNA-binding domain of 81 and 60 conserved amino acids which are separated by a poorly conserved linker sequence (reviewed in Schöler, 1991). With the exception of these two regions, the proteins of this family are highly divergent. The POU family is defined by the sequence homology of three mammalian transcription factors (*Pit-1*, *Oct-1*, and *Oct-2*) and one nematode regulatory protein (*unc-86*) (Herr et al., 1988). The POU homeodomain appears to be sufficient for low affinity binding to AT-rich DNA sequences; however, the addition of the POU-specific domain increases both DNA-binding affinity and specificity (Ingraham et al., 1990). Several POU proteins bind to the octamer sequence ATGCAAAT (Schöler et al., 1989b), a regulatory element required for both ubiquitous and tissue-specific gene expression (reviewed in

Kemler and Schaffner, 1990). Both, the POU-specific domain and the POU homeodomain are also involved in protein-protein interactions. Members of the POU family show distinct expression patterns during embryonic development (He et al., 1989; Schöler et al., 1989a). Octamer-binding transcription factor 4 (*Oct-4*) is a recently described POU transcription factor that is expressed early in the preimplantation embryo, and thus may regulate initial events of murine development (Schöler et al., 1989a; Okamoto et al., 1990; Rosner et al., 1990). The identification of *Oct-4* generated considerable interest because its message is expressed almost exclusively in germ cells and the uncommitted cells of the early embryo.

Oct-4 maps to the *t*-complex on mouse chromosome 17 and was shown to be inseparable from the Major Histocompatibility Complex (MHC) using BXD recombinant inbred strains (Schöler et al., 1990a). Its genetic localization to the MHC and its expression pattern made *Oct-4* a candidate for one of two embryonic *t*-lethal mutations, *tcl-t*¹² or *tcl-t*^{w5} (referred to below as *t*¹² or *t*^{w5}). Like *Oct-4*, both of these genes function in the early embryo; *t*¹² homozygotes fail to compact and form blastocysts at day 3 of development and *t*^{w5} mutants die shortly after 5.5 days when the embryonic ectoderm of the egg cylinder degenerates. These two

Correspondence to: K. Artzt, Dept. of Zoology, The University of Texas at Austin, Austin, Texas, U.S.A.

* Present address: Pohang Institute of Science and Technology, Department of Life Science, Pohang, Korea.

t-lethal mutations also map in the MHC; *t*^{w5} is genetically inseparable from *H-2K* (Artzt et al., 1988), and *t*¹² resides in the *T* region or the uncloned gap between the *Q* and *T* regions (Shin et al., 1984; and unpublished).

Here we show by physical mapping and detailed structural analysis of *Oct-4* from the *t*¹² haplotype that it is unlikely to be either *t*-lethal. *Oct-4* maps between *Q* and *T* in the MHC, is located within 35 kb of a class I gene, and represents the first cloned gene in the uncloned "gap" between *Q* and *T*. In situ analysis of early preimplantation embryos reveals that *Oct-4* is expressed both maternally and zygotically. Surprisingly, zygotic expression is not detectable until the 8-cell stage when compaction begins. In addition to *Oct-4*, several *Oct-4*-related genes are present in the mouse genome, one of them maps to chromosome 9 tightly linked to the *Apoa-1* locus.

Results and Discussion

Oct-4 is located in the MHC between *Q* and *T*

Oct-4 was mapped using a 350 bp fragment spanning the 5' region of the cDNA. The probe covers the distal half of exon 1 and about 20 bp of exon 2 of the *Oct-4* gene (see below). It includes 30 bp of the POU-

specific box, but has only a low sequence homology to other members of the POU family. After digestion with *Pst*I, a restriction fragment length polymorphism (RFLP) was identified in C57BL/6 (B6) versus AKR and A/J. This allowed use of the congenic chromosomes listed in Fig. 1A (Flaherty et al., 1990). Whereas, the B6.K1 and B6.K3 recombinants place *Oct-4* distal to *H-2D*; B6-*Tla*^a, B6.K2, and B6.K3 position it proximal to *T*. In addition, B10.M which is deleted for the *Q* region (O'Neill et al., 1986) has normal genomic *Oct-4* fragments resembling the AKR and A/J type (Fig. 1B). These data place *Oct-4* between *Q* and *T*, excluding its allelism with *t*^{w5}, but leaving the possibility of allelism with *t*¹². When it became clear that *Oct-4* resided between *Q* and *T*, a region of unknown physical size, the distance from the nearest class I gene was approximated using pulsed field gel electrophoresis (PFGE). Since three recombinants occurred between *Oct-4* and *T*, and none between *Oct-4* and *Q* (Fig. 1A), it seems reasonable to presume *Oct-4* is closer to the *Q* region. C3H/DiSn (C3H), and congenic +/*t* and *t/t* spleen DNA plugs were digested with four restriction enzymes and analyzed by PFGE (Fig. 2). In three cases *Oct-4* hybridized with a single band in each haplotype: *Nru*I gave 330 kb, *Not*I gave 460 kb, and *Mlu*I revealed a polymorphism of 335 kb for wild-type and 360 kb for the *t*-type indicating that the bands must represent *Oct-4* on chromosome 17. In

A

Strain	K	I	S	D	Q	<i>Oct-4</i>	I
B6	B	B	B	B	B	B	B
AKR	K	K	K	K	K	K	K
A/J	A	A	A	A	A	A	A
B6.AK1	B	B	B *	K	K	K	K
B6- <i>Tla</i> ^a	B	B	B	B	B	B *	A
B6.K1	B	B	B	B *	K	K	K
B6.K2	B	B	B	B	B	B *	K
B6.K3	K	K	K	K *	B	B *	A
B10.M	M	M	M	M	-	M	M

B

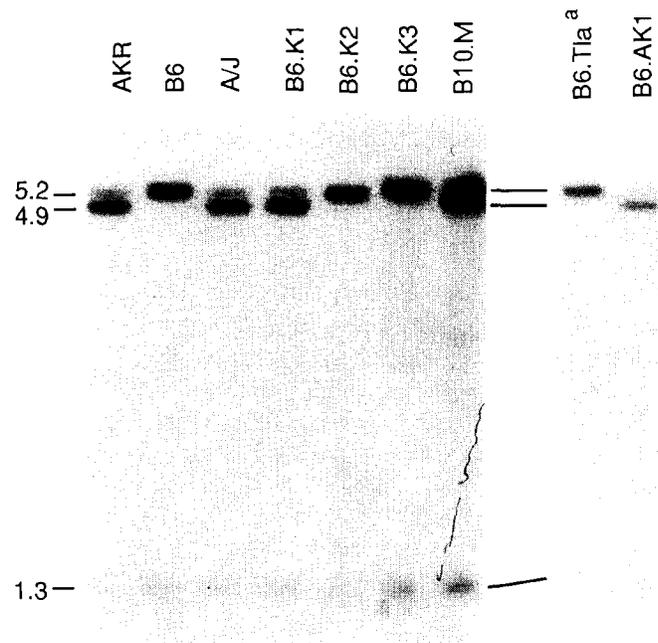


Fig. 1. A, Haplotypes used for mapping *Oct-4*. Asterisks indicate the recombinant breakpoints. B, Southern blots of their DNAs digested with *Pst*I and probed with *Oct-4*.

contrast, *Bss*HII revealed two fragments of 340 and 35 kb. It is unclear what the larger fragment represents because there is no *Bss*HII site in the gene. However, there is a *Bss*HII site in a cosmid containing *Oct-4* (see below). The same pulsed field gel was then probed with pH2IIa, a generic probe for class I genes (Steinmetz et al., 1981). Although predictably this probe hybridizes to many bands, the common fragments were: the *Mlu*I polymorphism, the *Not*I fragment, and the smaller *Bss*HII fragment. This smaller fragment was confirmed to hybridize with *Oct-4* and pH2IIa in standard Southern blots of *Bss*HII-digested genomic DNA (data not shown). Thus, the 35 kb *Bss*HII fragment defines the maximum distance between *Oct-4* and the nearest class I gene. Since the cosmid containing *Oct-4*

does not hybridize with pH2IIa, the minimum distance is the 18 kb flanking the *Bss*HII site on the cosmid.

Oct-4 is one of the first genes mapped to the region between *Q* and *T* and will be an extremely useful marker for this uncloned region. It is noteworthy that there are at least three interesting, very early embryonic genes in this general region: *t*¹², *Oct-4*, and Preimplantation-embryo-development (*Ped*), a gene which influences the rate of cleavage of preimplantation mouse embryos (Warner et al., 1987).

A family of Oct-4-related genes

Although the probe spanning the 5' region of the *Oct-4* cDNA is clearly derived from chromosome 17, it

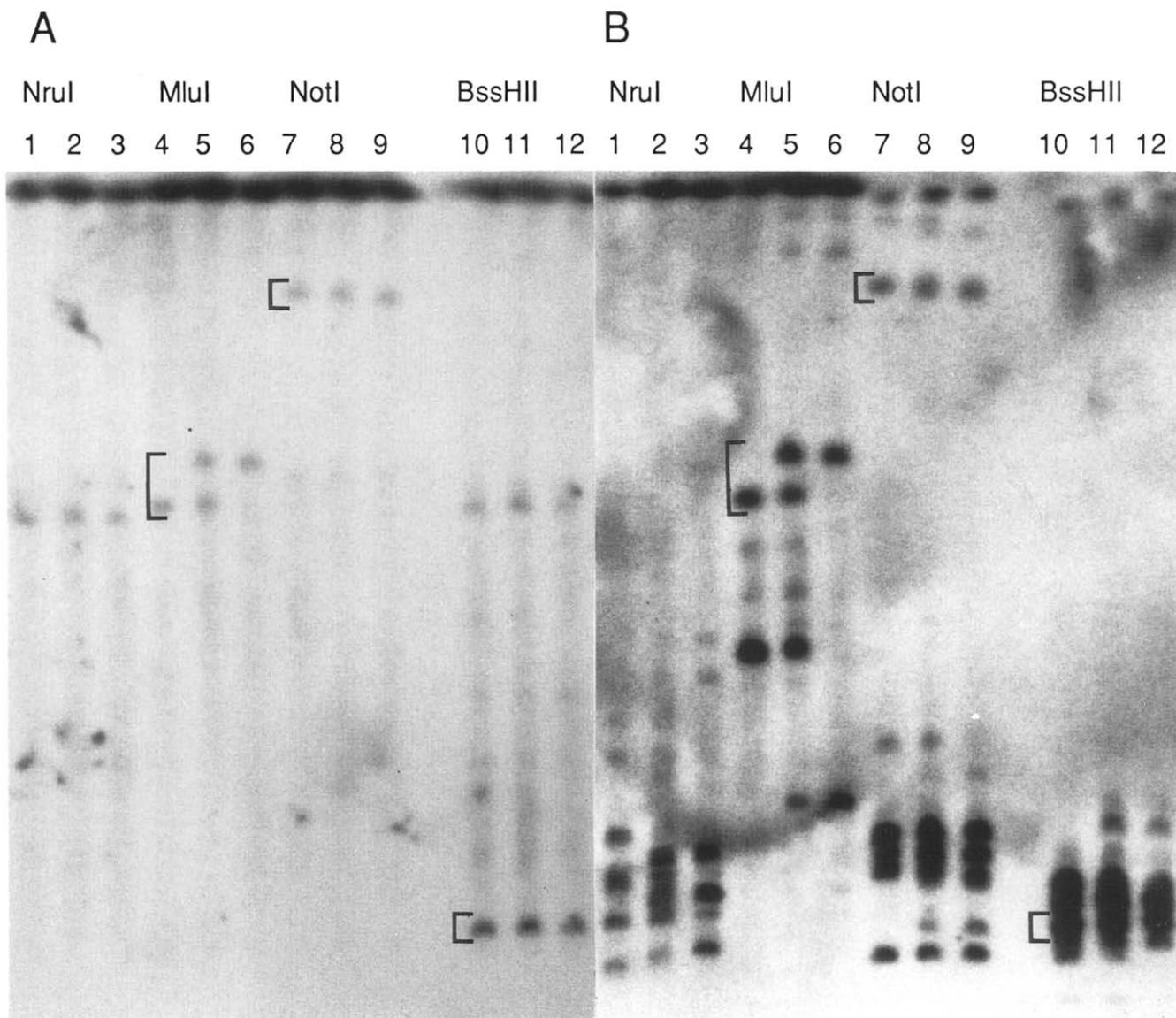


Fig. 2. Physical location of *Oct-4* with respect to class I genes. PFGE of C3H (+/+), lanes 1, 4, 7, and 10; C3H.+/*t*^{w5}, lanes 2, 5, 8, and 11; and C3H.*t*^{w5g}/*t*^{w5g}, lanes 3, 6, 9, and 12. *t*^{w5g} is a viable revertant of *t*^{w5} (unpublished). A is hybridized with *Oct-4*, and B is the same blot hybridized with pH2IIa. The common bands are indicated by brackets.

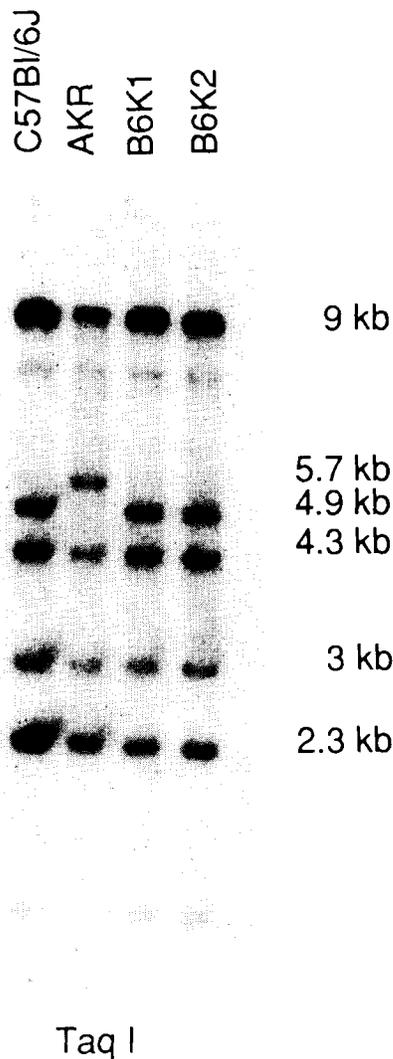


Fig. 3. An example of the *TaqI* polymorphism between C57BL/6 and AKR used to map an *Oct-4*-related gene on chromosome 9 in AKXD RI strains. C57BL/6 gives the same pattern as DBA/2J. Washing was done twice for 45 min in $2\times$ SSC at 65°C.

also detects cross-hybridizing bands. To determine their chromosomal location, DNAs from additional inbred strains were screened for a polymorphism using the 350 bp probe. A *TaqI*-digest revealed a polymorphism between AKR versus C57BL/6J and DBA/2J. A total of five different bands were detected with only one of them polymorphic between these strains (Fig. 3). Whereas, in C57BL/6 and DBA/2J bands of 9.0, 4.9, 4.3, 3.0, and 2.3 kb were seen, in AKR DNA a 5.7 kb

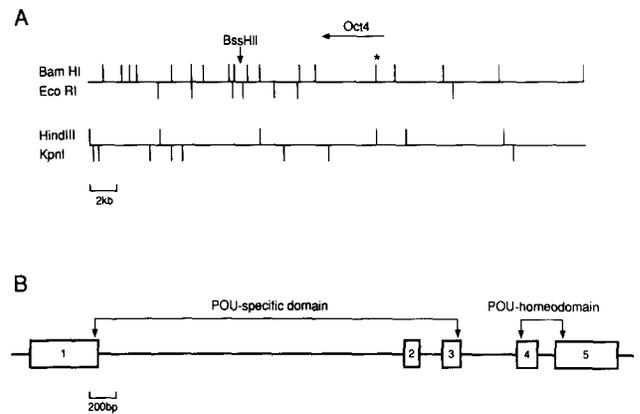


Fig. 4. Genomic structure of *Oct-4*. A, Restriction map of an *Oct-4* cosmid derived from the t^{w5g} complete *t*-haplotype. The position of the transcription unit is indicated with a horizontal arrow running 5' to 3'. The position of the *Bam*HI site missing in wild-type (C3H) and responsible for the RFLP between *t* and wild-type is marked with an asterisk. Except for this difference, the restriction maps are identical. A vertical arrow indicates a *Bss*HII site; *Not*I, *Nru*I and *Mlu*I sites were absent. B, Exon/intron organization of the *Oct-4* transcription unit.

fragment could be detected instead of the 4.9 kb band. This *TaqI* polymorphism between DBA/2J and AKR could be used to screen the AKXD RI strains to map this band. Unexpectedly, analysis of the strain distribution pattern in the AKXD RI set showed tight linkage to the *Apoa-1* locus on chromosome 9 and not to chromosome 17 (data not shown). No recombinants were found out of 21 AKXD RI strains analyzed, arguing that the polymorphic *TaqI* band reflects an *Oct-4*-related gene located on chromosome 9. Some of the other bands might represent *Oct-4*-related genes, although their chromosomal location remains to be determined. Evidence for the existence of *Oct-4*-related genes on different chromosomes has also been obtained by another laboratory, although none so far have been located on chromosome 9 (Siracusa et al., 1991).

Since there seems to be a family of *Oct-4* genes in the mouse related at their 5' end, we used the same probe to examine the evolutionary conservation of the N-terminal domain. An interspecies hybridization analysis revealed that, unlike the POU domains, the *Oct-4*-specific N-terminal nucleotide sequence is only conserved in mammals. Rat, cat, mink, and human had cross-hybridizing fragments, but other vertebrates such

Fig. 5. Nucleotide sequence of the genomic *Oct-4* transcription unit from the t^{f2} haplotype. Exon sequences are shown in capital letters, introns and 3' and 5' flanking DNA are in lower case. Numbers indicate nucleotide positions in exons. Introns 1 and 3 were only partially sequenced. Shown in bold type and underlined are: the cap site located about 550 bp upstream of the first exon (position 1) which was determined by primer extension analysis, and the start and stop codons at positions 164 and 1220 respectively. The consensus "GT--AG" splice sequences at exon/intron junctions, and the polyadenylation signal are underlined. Arrowed sequences indicate oligonucleotides used for PCR reactions (see Experimental procedures). Comparison of this genomic sequence to the cDNA sequence shows that the poly(A)⁺ tail must be added at position 1441. The nucleotide "T" at position 1308 (bolded and underlined) is "C" in wild-type (F9 cells) and the t^{w5g} haplotype.

Exon1

1 aagggttgct ctgtccagac gtccccaaacc tccgtctgga agacacagge agatagcct CGCCTCAGTT
 13 TCTCCCACC CCACAGCTCT GCTCCTCCAC CACCCAGGGG GCGGGGCCAG AGGTCAAAGC TAGAGGGTGG
 83 GATTGGGAG GGAGAGGTGA AACCGTCCCT AGGTGAGCCG TCTTTCCACC AGGCCCCCGG CTCGGGGTGC
 153 CCACCTTCCC CATGGCTGGA CACCTGGCTT CAGACTTCGC CTTCACACC CCACCAGGTG GGGGTGATGG
 223 GTCAGCAGGG CTGGAGCCGG GCTGGGTGGA TCCTCGAACC TGCTAAGCT TCCAAGGGCC TCCAGGTGGG
 293 CCTGGAATCG GACCAGGCTC AGAGGTATTG GGGATCTCCC CATGTCCGCC CGCATACGAG TCTGCGGAG
 363 GGATGGCATA CTGTGGACCT CAGGTTGGAC TGGGCCTAGT CCCCCAAGTT GGCGTGAGA CTTTGACGCC
 433 TGAGGGCCAG GCAGAGGCAC GACTGGAAG CAACTCAGAG GAACTCTCT CTGAGCCCTG TGCCGACCCG
 503 CCCAATGCCG TGAAGTTGGA GAAGGTGGAA CCAACTCCCG AGGAG gLaagtgaag ggac-----

---(1st intron: -2430bp total)---gttcgtctg tcaattctac agccttaaaa cttcttcaga
 atagggtgac attttgcct cagtggggcg gttttgagta atctgtgagg agataggaac ttgctggggt
 actgcacaga actctgggta gtgtggtact gtagatggct aggttctggg ggggaagag ccatctatgt
 cacctaggaa tagagtgaat aacatttata taatcagacc agcccttgag gaggetgaga tcttttcag

Exon2

548 gggcaccccta gggtcacagt cccagctggt gtgactctca caagtctgcc tttctcacta cag TCCCAG
 554 GACATGAAG CCCTGCAGAA GGAGCTAGAA CAGTTGCCA AGCTGCTGAA GCAGAAGAGG ATCACCTTGG
 624 GGTACACCCA GGCCGACGTG GGGCTCACC TGGCGTTCT CTTTG gAgg gtctccccca gcatgtcttg
 atctcacggc tcttaagtta ggcgcaaggg ggtgggacat ttaggagct gcttctccac aggtaaagga
 ggattagaca cttgtagctt gaactgtcag aggtggggct tgggtccct tcttctgcc tcaactcactg

Exon3

669 gtttgatcgg cctttcag G AAAGGTGTT CAGCCAGACCA CCATCTGTCG CTTTCAGGCC TTGCAGCTCA
 720 GCCTAAGAA CATGTGTAAG CTCCGGCCCC TGCTGGAGAA GTGGGTGGAG GAAGCCGACA ACAATGAGAA
 790 CCTTCAGGAG gAgggagtg gcaggtatgt tgcattgtct gccagccaca gtcccctctc ctgcttccat
 tctctggctg aaactcctcc ctctccaacc---(3rd intron: -450bp total)---ttccagagcc
 cctttcagta acccctggct ctggggccac atccagtcaa tgctccctta gcacaatccc ttagcggttt
 gttcttcagt cccatctcaa ggtggggctg ttgccaaagcc aaatactaaa gttgctcttg tcgccccat

Exon4

800 ctccccctgc ccag ATATG CAAATCGGAG ACCCTGGTGC AGGCCCGGAA GAGAAAGCGA ACTAGCATTG
 855 AGAACCGTGT GAGGTGGAGT CTGGAGACCA TGTTTCTGAA GTGCCGGAAG CCCTCCCTAC AGCAGATCAC
 925 TCACATGCC AATCAGCTTG GGCTAGAGAA GGAT gAgg tgcacaagtc ctgccctgtg gtacctggat
 gtttccctgt tccattccc caccccccct cccccccac cgccactgct gactgcagca tcccagagct

Exon5

959 tatgatctga tgtccatctc tgtgccatc ctag GTGGT TCGAGTATGG TTCTGTAACC GCGCCAGAA
 994 GGGCAAAGA TCAAGTATTG AGTATTCCTA ACGAGAAGAG TATGAGGCTA CAGGGACACC TTTCCCAGGG
 1064 GGGGCTGTAT CTTTCTCTCT GCCCCAGGT CCCCACCTTG GCACCCAGG CTATGGAAGC CCCCACTCA
 1134 CCACACTCTA CTCAGTCCCT TTTCTGAGG GCGAGGCCCT TCCTCTGTT CCGTCACTG CTCTGGGCTC
 1204 TCCATGCAT TCAAACTGAC GCACCAGCCC TCCCTGGGGA TGCTGTGAGC CAAGGCAAGG GAGGTAGACA
 1274 AGAGAACCTG GAGCTTTGGG GTTAAATTCT TTTATTGAGG AGGGATTAAG AGCACAACAG GGGTGGGGG
 1344 TGGGATGGG AAAGAAGCTC AGTGATGCTG TTGATCAGGA GCCTGGCCTG TCTGTCACTC ATCATTTTGT
 1414 TCTTAAATAA AGACTGGGAC ACACAGTaga tagetgaatc tcttttctct tcaqtctcta gagagcctgc
 gttggagaaa gccagtaatg gattctcaaa ccccaggtga tcttcaaac aggccccatt gaaaccattg
 gagtccccc aaaaatgccag ggatagttgg ggttggagcc caacctatag aggaagccat tgcattctg
 ccatctcaga ggcgtaagt ctctgctagc tgatggacat cacctcatag ccattgtctg gcag

as chicken, frog, and fish were negative, as were all invertebrates tested (data not shown).

Genomic organization and sequence of the Oct-4 gene from t- and wild-type

We were unable to map the chromosome 17 copy of *Oct-4* using recombinants derived from t^{12} because 29 restriction enzymes failed to identify a RFLP between the relevant parental *t*-haplotypes. This lack of polymorphism was not unexpected because all *t*-haplotypes are generally believed to be descended from a single ancestral chromosome. Another approach to examine the relationship between *Oct-4* and t^{12} is to analyze the genomic structure and sequence of *Oct-4* from the t^{12} haplotype. A t^{w5} homozygous revertant (t^{w5g} unpublished) cosmid library was screened with the same *Oct-4* probe to exclude trivial differences between the *t*- and wild-type sequences. A cosmid that mapped to chromosome 17 was identified using a *Bam*HI RFLP between *t*- (4.73 kb) and wild-type (6.2 kb) and Southern blots of several *t*-haplotypes congenic on C3H (data not shown). The entire cDNA of the Juhachi clone of *Oct-4* (Schöler et al., 1990b) is contained on the 4.73 kb *t*-derived *Bam*HI fragment (Fig. 4A). Based on the published cDNA sequence of the wild-type *Oct-4*, primers were synthesized and used to sequence the t^{w5g} genomic copy of *Oct-4*. The *Oct-4* gene consists of five exons and four introns (Fig. 4B). The size of both short introns, 2 and 4, was accurately assessable by DNA sequencing, whereas the size of introns 1 and 3 was approximated by PCR analysis of the cloned t^{w5g} genomic *Oct-4* DNA. The exon/intron junction sequences are shown in Fig. 5. C3H (wild-type), t^{12} , and t^{w5g} showed no apparent differences in the size and organization of exons and introns. It is noteworthy that both the POU-specific and POU-homeodomains are encoded by two or three split exons. The POU-domains of *unc-86*, *Oct-2*, and *Pit-1* are also scattered over several exons (Finney et al., 1988; Hatzopoulos et al., 1990; Li et al., 1990). The cap site of the *Oct-4* transcript was determined by primer extension using total RNA from t^{w5}/t^{w5} ES cells. It is located 163 bp upstream of the start codon (data not shown).

To obtain a t^{12} genomic copy for sequencing, liver DNA from a C3H. t^{12} mouse was digested with *Eco*RI and a minilibrary of gel purified 11.7 kb *Eco*RI fragments was prepared in EMBL 4 because both the wild-type and *t*-specific *Bam*HI fragments reside inside a nonpolymorphic 11.7 kb *Eco*RI fragment. After screening with *Oct-4*, a plaque containing the t^{12} -derived 4.73 kb *Bam*HI fragment was selected for subcloning and subsequent sequencing.

The sequences of the *t*-derived copies of *Oct-4* are remarkably similar to wild-type considering *t*-haplotypes and wild-type have evolved independently for

approximately 1-2 Myr. In 1280 bp of coding sequence compared, the only difference in wild-type, t^{w5g} , and t^{12} was a C to T transition in t^{12} in the 3' untranslated region 110 bp 5' to the polyadenylation signal. Since the POU-specific and POU-homeodomains of t^{12} were identical to wild-type it is unlikely that t^{12} is *Oct-4*, although a promoter or enhancer defect cannot be excluded. However, using polymorphic dinucleotide repeats and t^{12} recombinant chromosomes from the Austin laboratory, Hiroshi Uehara found t^{12} and *Oct-4* are genetically separable (Uehara, 1991).

Maternal and zygotic expression of the Oct-4 gene

The embryonic expression of *Oct-4* has been studied by in situ hybridization in zygotes and early embryos from the 16- to 32-cell stage onward (Rosner et al., 1990; Schöler et al., 1990a). Since it is expressed in primordial germ cells of both sexes and certainly represents a maternal message in mature oocytes, it seemed pertinent to examine its temporal expression between the stages of zygote and 16-cells to determine if expression is continuous or not. A 462 bp probe covering the POU-homeodomain and part of the POU-specific-domain was used for in situ hybridization to embryos serially sectioned inside the oviduct. Every sixth section was stained without hybridization for morphological staging of the embryos, and sections of the ovary were included on the same slide as a control for signal level. At the 4-cell stage, the *Oct-4* message level is comparable to the background seen in the oviductal epithelium, and negative when compared to ovarian oocytes processed together. At the 8-cell stage, all embryos are expressing *Oct-4* at message levels equivalent to that in growing oocytes, and by the 16-cell stage, the embryos are intensely labeled (Fig. 6). Although the sensitivity of in situ hybridization might not allow detection of low levels of message earlier than 8-cells, zygotic expression does not achieve peak abundance until 16-cells. Thus, the *Oct-4* maternal message diminishes or disappears sometime after fertilization and zygotic expression is detectable at 8-cells rising rapidly thereafter.

Although these data indicate that *Oct-4* is an early embryonic message, it is not among the earliest messages expressed. In the mouse, zygotic expression is detectable at 2-cells, and all classes of RNA are synthesized from the 4-cell stage onwards (Johnson, 1981). Since there is temporal separation or diminution between maternal and zygote expression, it is reasonable to speculate that there are earlier acting factors responsible for the onset of zygotic message. Morphological examination of the embryos revealed that the zygotic expression of *Oct-4* appears to be coincident with compaction. This is precisely the morphological transition at which t^{12} homozygous mutant embryos

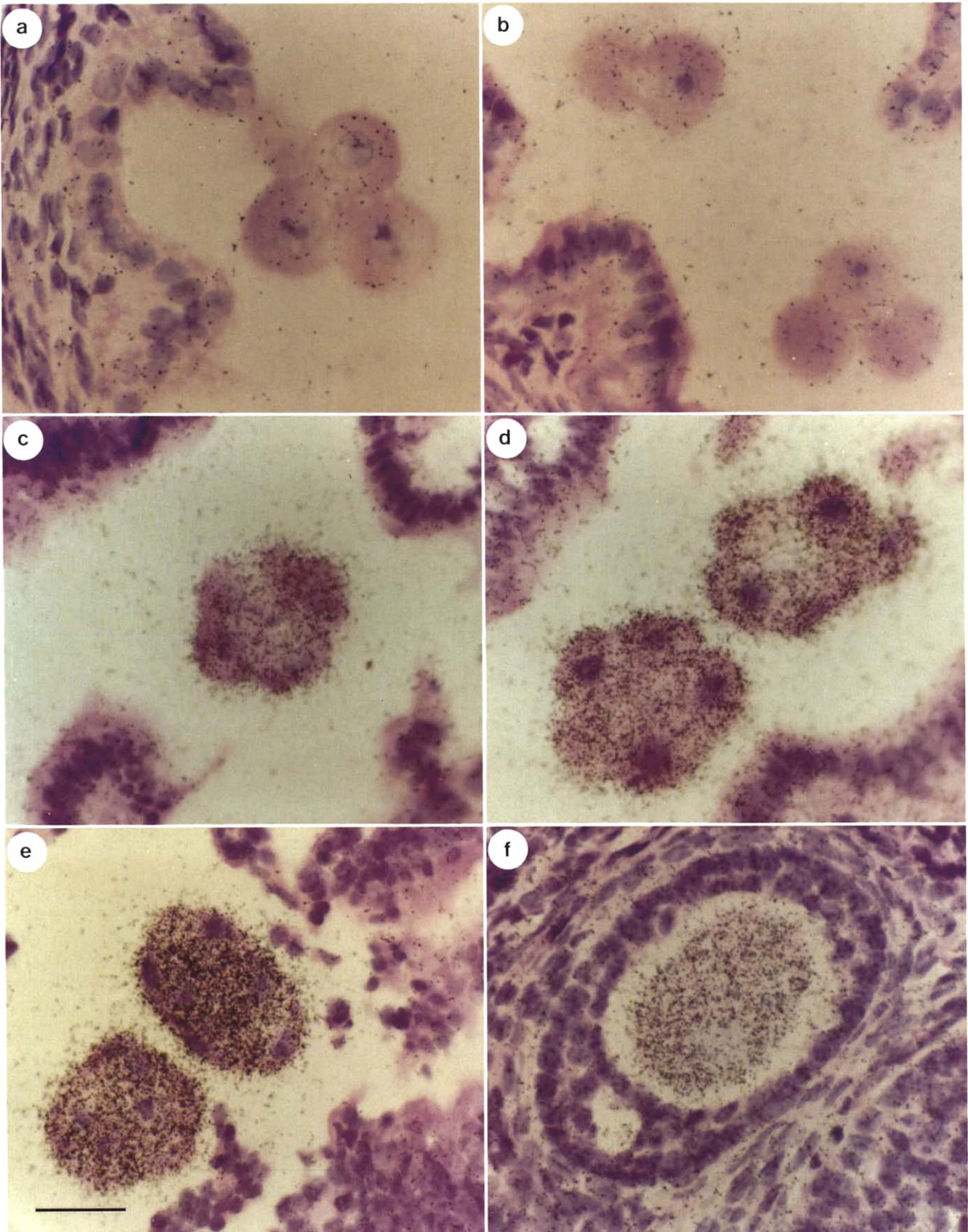


Fig. 6. Maternal and zygotic expression of the *Oct-4* gene. In situ hybridization of the *Oct-4* probe to 4-cell embryos, a and b; 8-cell embryos, c and d; and 16-cell embryos, e. f is an oocyte from the mother of c and d shown for comparison of signal intensity. The bar indicates 30 μ m.

fail. Although there is good evidence that t^{12} and *Oct-4* are not the same gene, in situ hybridization to t^{12} embryos indicates that there might be an interesting biological relationship. A proportion of 8-celled embryos from t^{12} heterozygous matings, consistent with the expected number of homozygotes, appears to have very little or no *Oct-4* message (data not shown). However, although the apparently negative embryos are still morphologically normal, it is difficult to prove they are zygotically transcriptionally active or not physiologically retarded. Further studies might determine if there is a relationship between *Oct-4* and t^{12} . Since *Oct-4* appears to require regulatory factors which are present before *Oct-4* is zygotically expressed, perhaps the as yet unidentified t^{12} product is acting upstream of *Oct-4*.

Experimental Procedures

Mice

All C3H/DiSn (C3H) and C3H.*t* mice were bred at the University of Texas at Austin colony. t^{w5g} is a viable revertant that arose in the C3H. t^{w5} stock (unpublished). It is a complete haplotype identical in all respects to t^{w5} except it no longer carries the t^{w5} lethal. t^{w5g} homozygotes are a convenient source of t^{w5}/t^{w5} DNA if the region of the lethal gene is not under consideration. Other inbred or RI strains were obtained from the Jackson Laboratory except the B6.K recombinants and B10.M which were a gift from Lorraine Flaherty. Embryos for in situ hybridization were obtained from random-bred CF1 mice maintained in Austin.

Southern hybridization analysis

Southern blot analysis was done essentially as described (Uehara et al., 1987) using Hybond N membrane (Amersham) except that the final wash was in $0.1 \times \text{SSC}/0.1\% \text{SDS}$ at 65°C for 10 min unless otherwise noted. Probes were radiolabeled by random priming (Feinberg and Vogelstein, 1983) or nick translation (Rigby et al., 1977).

Pulsed field gel electrophoresis (PFGE)

DNA plugs were made as described (Smith et al., 1986) from spleen cells aiming to have $5 \mu\text{g}$ of DNA in a $70 \mu\text{l}$ block of 0.5% low-melting-point agarose. The plugs were digested overnight using three units of restriction enzyme/ μg DNA. DNA was size-fractionated in 1.2% agarose gels ($20 \times 20 \text{ cm}$) in $0.25 \times \text{TBE}$ ($22 \text{ mM Tris base}/22 \text{ mM boric acid}/0.5 \text{ mM EDTA}$) using a hexagonal gel box designed by Chu et

al. (1986). Running conditions were a 45 s pulse time at 3 V/cm for 16 h, followed by a 35 s pulse time at 5 V/cm for 43 h at $4 \sim 6^\circ\text{C}$.

After electrophoresis, gels were irradiated with uv for 2 min/side, treated with $0.5 \text{ N NaOH}/1 \text{ M NaCl}$ for 1.5 h, and neutralized in $0.5 \text{ M Tris-HCl}/3 \text{ M NaCl}$, pH 7.5. DNA was transferred to Hybond N membrane by vacuum blotting for 5 h at a pressure of $50 \sim 60 \text{ cmH}_2\text{O}$. The blot was hybridized at 42°C in $5 \times \text{SSC}/50\% \text{ formamide}/50 \text{ mM sodium phosphate}$, pH 6.5/ $1 \text{ mM EDTA}/0.1\% \text{ SDS}/2.5 \times \text{Denhardt's solution}/7.5\% \text{ dextran sulfate}$, containing $100 \mu\text{g/ml}$ herring sperm DNA and $2 \times 10^6 \text{ cpm/ml}$ of ^{32}P -labeled probes, and washed in $2 \times \text{SSC}/0.1\% \text{ SDS}$ for 50 min at room temperature with one change.

Cloning and sequencing

C3H and t^{12} copies of the 11.7 kb *EcoRI* fragment containing the *Oct-4* transcription unit were cloned as follows: $11 \sim 12.5 \text{ kb EcoRI}$ fragments from C3H. $t^{12}/+$ DNA were collected from a preparative gel and cloned into EMBL 4 (Stratagene) to make a minilibrary which was then screened with *Oct-4* DNA probe. Positively hybridizing plaques were classified into C3H- or t^{12} -type by the *BamHI* restriction pattern of their inserts. Inserts were subcloned into pBluescript II KS (Stratagene). The t^{w5g} copy of the 11.7 kb *EcoRI* fragment was obtained from a shotgun cloning of an *Oct-4* cosmid into pBluescript II KS.

These cloned DNAs were used as templates in double-strand sequencing by the dideoxynucleotide chain termination method (Sanger et al., 1977) using Sequenase (USB).

Primer extension analysis

A 17 mer oligonucleotide complementary to the sequence at position 267 ~ 283 of exon 1 (Fig. 5) was synthesized using an ABI DNA synthesizer. Primer extension was performed following the procedures of Kingston (1990), using $50 \mu\text{g}$ of total RNA obtained from t^{w5}/t^{w5} ES cells (Magnuson et al., 1982). The extension product was analyzed on a 6% polyacrylamide-urea sequencing gel.

Polymerase chain reaction (PCR)

Oligonucleotides were synthesized to the positions overlined in Fig. 5, and used in PCR reactions to determine the size of introns 1 and 3. The template was the C3H, t^{12} or t^{w5g} copy of 11.7 kb *EcoRI* fragment containing the *Oct-4* transcription unit cloned from either C3H. $t^{12}/+$ or C3H. t^{w5g}/t^{w5g} . PCR was carried out in a thermocycler (Coy Laboratory Products, Inc.), using 10 ng of template DNA and 2.5 units

of *Taq* DNA polymerase (Perkin Elmer Cetus) for each reaction. Running conditions: for intron 1, 30 cycles of 94°C (1.8 min), 55°C (1.5 min) and 72°C (5.0 min); for intron 3, 30 cycles of 94°C (1.5 min), 49°C (1.2 min) and 72°C (3.5 min).

Construction and screening of cosmid library

The cosmid library of C3H.*t*^{w5g} / *t*^{w5g} genomic DNA was made according to the procedure of Dillela and Woo (1985) using the pWE15 (Stratagene) vector. The library was screened as described (Uehara et al., 1987). The restriction map of clones was determined by the terminase-oligomer method (Rackwitz et al., 1985).

In situ hybridization

This was performed as in Ha et al. (1991). Briefly, ³⁵S-labeled RNA was hybridized to paraplast-embedded tissue sectioned at 6 μm. Slides were washed at high stringency with RNase treatment. They were exposed for 5 days, and then stained with hematoxylin and eosin.

Acknowledgements

We thank Lori Flaherty for helpful discussions and reading of the manuscript. This work was supported by NIH grants HD10668 and CA21651 to K.A.; the Max Planck Society, and the Bundesministerium für Forschung und Technologie (R.B. and H.S.).

References

- Artzt, K. (1984) *Cell*, 39, 565–572.
- Artzt, K., Abe, K., Uehara, H. and Bennett, D. (1988) *Immunogenetics*, 28, 30–37.
- Chu, G., Vollrath, D. and Davis, R.W. (1986) *Science*, 234, 1582–1585.
- Dillela, A.G. and Woo, S.L.C. (1985) *Focus* (technical report from BRL), 7, 1–5.
- Feinberg, A.P. and Vogelstein, B. (1983) *Anal. Biochem.*, 132, 6–13.
- Finney, M., Ruvkun, G. and Horvitz, H.R. (1988) *Cell*, 55, 757–769.
- Flaherty, L., Elliott, E., Tine, J.A., Walsh, A.C. and Waters, J.B. (1990) *Critical Rev. Immunol.*, 10, 131–175.
- Ha, H., Howard, C.A., Yeom, Y.I., Abe, K., Uehara, H., Artzt, K. and Bennett, D. (1991) *Dev. Gen.*, in press.
- Hatzopoulos, A.K., Stoykova, A.S., Erselius, J.R., Goulding, M., Neuman, T. and Gruss, P. (1990) *Development*, 109, 349–362.
- He, X., Treacy, M.N., Simmons, D.M., Ingraham, H.A., Swanson, L.W. and Rosenfeld, M.G. (1989) *Nature*, 340, 35–42.
- Herr, W., Strum, R.A., Clerc, R.G., Corcoran, L.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.
- Ingraham, H.A., Flynn, S.E., Voss, J.W., Albert, V.R., Kapiloff, M.S., Wilson, L. and Rosenfeld, M.G. (1990) *Cell*, 61, 1021–1033.
- Johnson, M.H. (1981) *Biol. Rev.*, 56, 463–398.
- Kemler, I. and Schaffner, W. (1990) *FASEB J.*, 4, 1444–1449.
- Kingston, R.E. (1990) In Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A. and Struhl, K. (eds.), *Current protocols in molecular biology*, Greene Publishing Associates and Wiley-Interscience, New York, Vol. 1, pp. 4.8.1–4.8.3.
- Li, S., Crenshaw III, E.B., Rawson, E.J., Simmons, D.M., Swanson, L.W. and Rosenfeld, M.G. (1990) *Nature*, 347, 528–533.
- Magnuson, T., Epstein, C.J., Silver, L.M. and Martin, G.R. (1982) *Nature*, 298, 750–753.
- Okamoto, K., Okazawa, H., Okuda, A., Sakai, M., Muramatsu, M. and Hamada, H. (1990) *Cell*, 60, 461–472.
- O'Neill, A.E., Reid, K., Garberi, J.C., Karl, M. and Flaherty, L. (1986) *Immunogenetics*, 24, 368–373.
- Rackwitz, H.R., Zehetner, G., Frischauf, A.-M. and Lehrach, H. (1985) *Gene*, 40, 259–266.
- Rigby, P.W.I., Dieckmann, M., Rhodes, C. and Berg, P. (1977) *J. Mol. Biol.*, 113, 237–251.
- Rosner, M.H., Vigano, M.A., Ozato, K., Timmons, P.M., Poirier, F., Rigby, P.W.J. and Staudt, L.M. (1990) *Nature*, 345, 686–692.
- Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA*, 74, 5463–5467.
- Schöler, H.R. (1991) *Trends Genet.*, in press.
- Schöler, H.R., Balling, R., Hatzopoulos, A.K., Suzuki, N. and Gruss, P. (1989a) *EMBO J.*, 8, 2551–2557.
- Schöler, H.R., Hatzopoulos, A.K., Balling, R., Suzuki, N. and Gruss, P. (1989b) *EMBO J.*, 8, 2543–2550.
- Schöler, H.R., Dressler, G.R., Balling, R., Rohdewohld, H. and Gruss, P. (1990a) *EMBO J.*, 9, 2185–2195.
- Schöler, H.R., Ruppert, S., Suzuki, N., Chowdhury, K. and Gruss, P. (1990b) *Nature*, 344, 435–439.
- Shin, H.-S., Bennett, D. and Artzt, K. (1984) *Cell*, 39, 573–578.
- Siracusa, L., Rosner, M.H., Vigano, M.A., Gilbert, D.J., Staudt, L.M., Copeland, N.G. and Jenkins, N.A. (1991) *Genomics*, in press.
- Smith, C.L., Warburton, P.E., Gaal, A. and Cantor, C.R. (1986) In Setlow, J. and Hollaender, A. (eds.), *Genetic engineering*. Plenum Press, New York, Vol. 8, pp. 45–70.
- Steinmetz, M., Frelinger, J.G., Fisher, D., Hunkapiller, T., Pereira, D., Weissman, S.M., Uehara, H., Nathenson, S.G. and Hood, L. (1981) *Cell*, 24, 125–134.
- Uehara, H. (1991) *Immunogenetics*, in press.
- Uehara, H., Abe, K., Park, C.-H.T., Shin, H.-S., Bennett, D. and Artzt, K. (1987) *EMBO J.*, 6, 83–90.
- Warner, C.M., Gollnick, S.O. and Goldbard, S.B. (1987) *Biol. Reprod.*, 36, 606–610.