The ubiquitous octamer-binding protein Oct-1 contains a POU domain with a homeo box subdomain

Richard A. Sturm, Gokul Das, and Winship Herr

Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724 USA

The octamer motif ATGCAAAT is recognized indistinguishably by two mammalian transcription factors: one that is expressed ubiquitously and referred to here as Oct-1, and another, Oct-2, that is expressed in lymphoid cells. We report the cDNA cloning of the human *oct*-1 gene, which encodes Oct-1, by screening λgt11 recombinant phage in situ for octamer motif-specific DNA binding. One λgt11 recombinant expressed a β-galactosidase–octamer-binding fusion protein with a DNA-binding specificity indistinguishable from human HeLa cell Oct-1 protein. As expected for a ubiquitously expressed protein, Oct-1 mRNA is expressed in all five human and two mouse cell lines tested. Polyclonal rabbit antiserum raised against the β-galactosidase fusion protein shows that the DNA-binding domains of Oct-1 and Oct-2 proteins are related antigenically. Deletion analysis of the 743-amino-acid-long *oct*-1 open reading frame shows that the DNA-binding activity lies within a central highly charged domain of 160 amino acids. Comparison of the Oct-1 and Oct-2 sequences reveals that this domain is nearly identical between the two proteins. Highly similar domains are also present in the pituitary-specific transcription factor Pit-1 and the *Caenorhabditis elegans unc-86* cell lineage gene product (see Herr et al. 1988). Within this shared POU (Pit-1, Oct-1 and Oct-2, *unc-86*) domain (pronounced 'pow') lie two subdomains: a POU-related homeo box and a POU-specific box. The Oct-1 protein is unique among the POU-related proteins and other homeo box proteins because it is expressed ubiquitously.

[*Key Words*: Octamer; *oct*-1 gene; *oct*-2 gene; DNA-binding proteins; enhancer/promoter; replication]

Received October 1, 1988; revised version accepted October 17, 1988.

Transcriptional regulation depends largely on the sequence-specific interaction of *trans*-activator proteins with *cis*-acting promoter elements. Sequence-specific *trans*-activators contain two essential domains for function: a DNA-binding domain and a *trans*-activation domain. As first shown by fusion of the bacterial *lex*A DNA-binding domain to the yeast GAL4 *trans*-activation domain (Brent and Ptashne 1985), these domains can be interchanged. The specificity of transcriptional activation is conferred by the DNA-binding domain, which targets the *trans*-activator to the promoter carrying a corresponding DNA-binding site(s) (for review, see Ptashne 1988). To date, *trans*-activation domains have displayed little promoter specificity. For example, the yeast GAL4 *trans*-activator can stimulate a variety of promoters in mammalian cells provided that GAL4 DNA-binding sites are located in the vicinity of the promoter (Kakidani and Ptashne 1988; Ptashne 1988; Webster et al. 1988). Nevertheless, more than one class of *trans*-activation domain may exist because GAL4 fails to activate a U2 small nuclear RNA (snRNA) promoter containing GAL4 DNA-binding sites in mammalian cells (Tanaka et al. 1988).

The lymphoid-specific and ubiquitous octamer/decamer (ATGCAAATNA)-motif-binding proteins provide a model system to study conservation of DNA-binding specificity, as well as to elucidate *trans*-activation specificity. The octamer motif (Falkner and Zachau 1984; Parslow et al. 1984), or closely related sequences, is a functional element in the promoters of the ubiquitously expressed snRNA (Ares et al. 1985; Mattaj et al. 1985; Mangin et al. 1986; Carbon et al. 1987) and histone H2B genes (LaBella et al. 1988; Sturm et al. 1988), a lymphoid-specific element in immunoglobulin upstream and enhancer regions (Falkner and Zachau 1984; Mason et al. 1985; Gerster et al. 1987; Lenardo et al. 1987; Wirth et al. 1987), and a DNA replication element in the adenovirus origin of replication (Pruijn et al. 1986; Rosenfeld et al. 1987). Two factors that bind indistinguishably to the octamer motif have been purified. One, first referred to as NF-A1 (Staudt et al. 1986), is ubiquitously expressed, and the other, first called NF-A2 (Staudt et al. 1986), is restricted to lymphoid cells (Landolfi et al. 1986; Singh et al. 1986; Staudt et al. 1986; Rosales et al. 1987; Scheidereit et al. 1987). A 90- to 100-kD protein responsible for the ubiquitous octamer-binding activity has been purified from HeLa cells in several laboratories and variously referred to as OTF-1 (Fletcher et al. 1987),

OBP100 (Sturm et al. 1987), or NFIII (O'Neill and Kelly 1988). OTF-1 and NFIII, which were initially shown to stimulate histone H2B transcription and adenovirus replication, respectively, are identical proteins (O'Neill et al. 1988); OBP100 is probably also identical to OTF-1/NFIII. The lymphoid-specific octamer-binding protein OTF-2 has also been purified (Scheidereit et al. 1987; Hanke et al. 1988) and stimulates transcription of an immunoglobulin light-chain κ promoter in vitro (Scheidereit et al. 1987).

The different expression patterns of the ubiquitous octamer-binding protein, which we refer to here as Oct-1, and the lymphoid-specific octamer-binding protein, Oct-2 (Clerc et al. 1988, this issue), parallel the activity of the octamer motif in different contexts. In the context of a variety of RNA polymerase II promoters that direct expression of polyadenylated mRNAs (e.g., immunoglobulin, β-globin, SV40 early), the octamer motif displays B-cell-specific enhancer function (Davidson et al. 1986; Gerster et al. 1987; Wirth et al. 1987; Tanaka et al. 1988). But in the context of the U2 snRNA promoter, which is also transcribed by RNA polymerase II (Frederiksen et al. 1978; Gram Jensen et al. 1979), the octamer motif itself displays enhancer function in non-B cells (Mattaj et al. 1985; Ares et al. 1987; Tanaka et al. 1988). The different patterns of octamer motif function in a snRNA promoter, compared with the β-globin promoter, suggest that the ubiquitous octamer-binding protein belongs to a different class of *trans*-activators. Thus, the ubiquitous Oct-1 and lymphoid-specific Oct-2 proteins are an excellent pair of *trans*-activators with which to study both conservation of DNA sequence recognition and divergence of *trans*-activation function.

We report here the isolation of human cDNA clones encoding ubiquitously expressed octamer-binding protein, referred to here as the *oct*-1 gene. First isolated as a λgt11 β-galactosidase fusion protein by the in situ DNA filter detection method (Singh et al. 1988; Staudt et al. 1988; Vinson et al. 1988), this protein meets several criteria (i.e., DNA-binding specificity, ubiquitous expression, antigenic relationship, and size) that indicate it is the ubiquitously expressed octamer-binding protein variously called OTF-1, NFIII, and OBP100. Comparison of the Oct-1 and lymphoid Oct-2 (Clerc et al. 1988) amino acid sequences reveals a highly conserved region, 160 amino acids long, that contains the DNA-binding domain and is also present in the pituitary transcription factor Pit-1 (Ingraham et al. 1988) and the *Caenorhabditis elegans* homeotic *unc-86* gene product (Finney et al. 1988; see Herr et al. 1988, this issue).

Results

Isolation of λC5: a λgt11 human cDNA recombinant encoding a β-galactosidase–octamer-binding fusion protein

To isolate a cDNA clone of the ubiquitously expressed octamer-binding protein, we probed a human λgt11 cDNA expression library for octamer-specific DNA-binding proteins by the in situ filter detection method (Singh et al. 1988; Staudt et al. 1988) using denaturation of filter-bound proteins to enhance the signal (Vinson et al. 1988). Because of the widespread expression pattern of the ubiquitous octamer-binding protein, we chose to screen a readily available λgt11 cDNA library derived from the human teratocarcinoma cell line NTera-2D1 (constructed by J. Skowronski; SenGupta et al. 1986). This cell line displays elevated levels of ubiquitous octamer-binding activity (data not shown) and mRNA expression (see below). We screened 4×10^5 plaques with a radiolabeled multimerized SV40 octamer motif and identified one plaque (termed λC5) that bound the probe effectively. Figure 1A shows that this binding is specific. Replica filter lifts from plates containing either partially purified λC5 phage or L20 phage, which express a β-galactosidase–C/EBP fusion protein that binds specifically to C/EBP-binding sites (Landschulz et al. 1988b; Vinson et al. 1988), were probed with either the octamer motif or C/EBP-binding sites. Each recombinant fusion protein bound only to its corresponding probe (Fig. 1A).

To examine the β-galactosidase–octamer-binding fusion protein expressed by λC5, bacterial extracts were prepared from λC5-, λgt11-, and L20-infected cells and fractionated on a polyacrylamide gel (Fig. 1B). Compared with the uninfected cell extract (lane 5), the λC5-infected cell extract contains a 190- to 200-kD protein and a series of smaller proteins of ~150–190 kD. The λC5-, λgt11-, and L20-encoded proteins were also tested for binding to the octamer motif after transfer to a filter. The larger of the λC5-specific proteins bound to the octamer motif probe (Fig. 1C, lane 2), but β-galactosidase (λgt11) and the C/EBP fusion protein (L20) did not bind to the probe. The low level of binding by the major 150- to 190-kD products present in Figure 1B suggests that many of these smaller λC5 proteins are missing the octamer motif DNA-binding domain. Some of the smaller proteins do, however, bind the octamer motif (Fig. 1C, lane 2), which may explain the heterogeneous pattern of complex formation observed in the gel retardation assays described below.

DNA binding specificity of the λC5 fusion protein

To establish the authenticity of the octamer-specific DNA-binding activity of the λC5 fusion protein, we tested its affinity for a number of known binding and nonbinding sites for the HeLa cell octamer-binding protein OBP100 in a gel retardation assay. OBP100 displays remarkably flexible DNA-binding specificity (Sturm et al. 1987; Baumruker et al. 1988); in addition to binding to perfect octamer motifs [e.g., the immunoglobulin heavy-chain (IgH) enhancer octamer motif, ATGCAAAT], OBP100 binds to two degenerate SV40 octamer motifs and to the herpes simplex virus (HSV) TAATGARAT consensus sequence, which represents a very degenerate octamer motif (Baumruker et al. 1988). Figure 2 shows a comparison of the binding activities of *Escherichia coli*-derived λC5 fusion protein and the partially purified OBP100 protein from HeLa cells. These

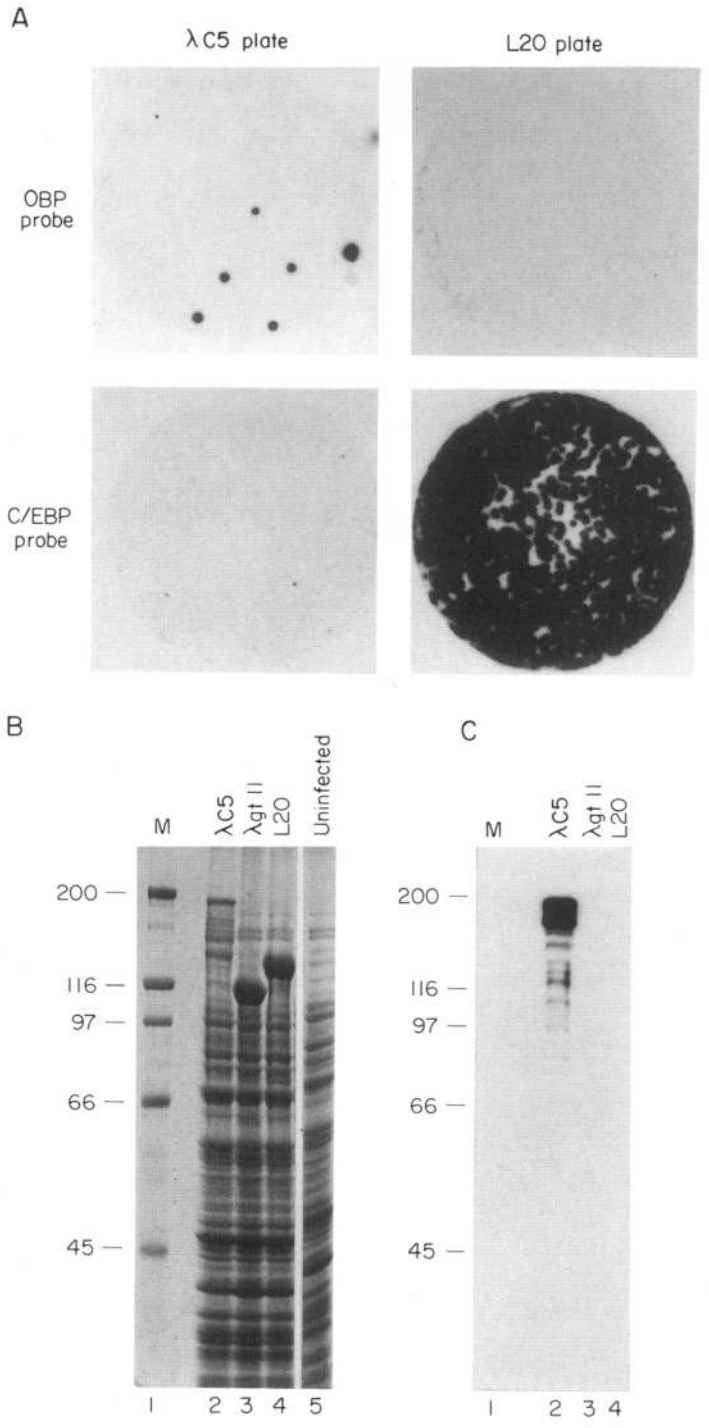

Figure 1. Octamer motif binding of the λC5 β-galactosidase fusion protein. (*A*) Two replica filter lifts from plates containing partially purified λC5 or the C/EBP-binding bacteriophage L20 were each screened reciprocally with ^{32}P-labeled 14xB17 restriction fragment (OBP probe) or a ligated oligomer containing a C/EBP-binding site (C/EBP probe; Vinson et al. 1988). Each probe was radiolabeled by nick-translation and added to the respective filters at 1×10^5 cpm/ml after processing the filters by the guanidine hydrochloride denaturation/renaturation procedure (Vinson et al. 1988). (*B*) One milliliter of log-phase *E. coli* strain Y1089 uninfected or infected with λC5, λgt11, or L20 at 5 pfu/cell was heat-shocked and subsequently treated with 10 mM IPTG for 1 hr. The cells were collected by centrifugation, and 20% of the total cell protein was fractionated on an 8% polyacrylamide–SDS gel after boiling in SDS-loading buffer. The fractionated proteins were visualized by Coomassie blue staining. (Lane *1*) Bio-Rad high-molecular-weight markers; (lane *2*) λC5-infected cells; (lane *3*) λgt11-infected cells; (lane *4*) L20-infected extracts; (lane *5*) uninfected cell extract. (*C*) Southwestern blot analysis was performed on four tracks equivalent to lanes *1–4* in *B*. Following electrotransfer to nitrocellulose, the filter was probed with nick-translated 14xB17 restriction fragment, as in *A*, without denaturation/renaturation treatment.

two proteins display the same relative affinity for each of these sites. The best binding site is the wild-type IgH enhancer octamer motif (lanes 1 and 2). The λC5-infected cell extract forms a major complex with a reduced mobility, compared with the bona fide OBP100 complex. This is consistent with the binding of the larger 190- to 200-kD λC5 fusion protein. The more rapidly migrating complexes with the *E. coli* extract may reflect interactions with smaller OBP fusion peptides present in the infected cell extract (see Fig. 1C). Whatever the case, all of the λC5-specific complexes are also octamer specific because two point mutations within the nearly perfect SV40 Octa1 octamer motif (AaGgAAAG; *dpm*8), which abolish OBP100 binding (Sturm et al. 1987; cf. lanes 3 and 4, Fig. 2), also prevent binding of all the λC5 fusion proteins (cf. lanes 5 and 6).

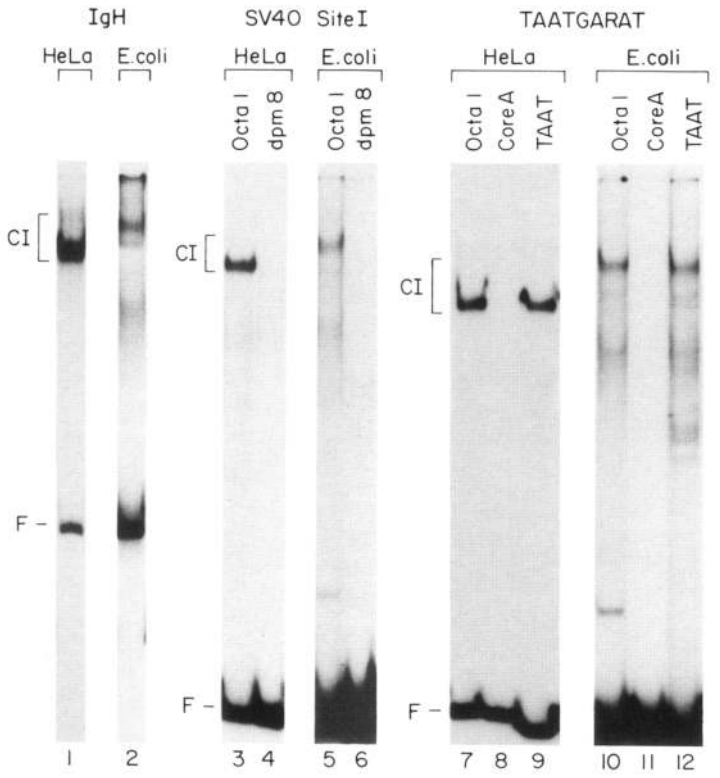

Figure 2. The λC5 β-galactosidase fusion protein displays the same relative affinity for multiple octamer protein-binding sites as HeLa-cell-derived OBP100. An *E. coli* extract was prepared from λC5-infected cells (see Materials and methods) and DNA-binding activity assayed and compared to a partially purified OBP100 fraction from HeLa cells (Sturm et al. 1987) on three well-characterized OBP100-binding sites and two nonfunctional binding sites. Longer exposures of autoradiograms of the gel retardations produced by the bacterial extract are shown to compensate for the lower binding activity of the λC5-infected *E. coli* extract. The major octamer-specific complexes are indicated by the bracket and labeled CI. Restriction fragment probes end labeled with ^{32}P were assayed with either the HeLa or *E. coli* extracts, as described in Materials and methods, and indicated here. (Lanes *1, 2*) A *Pst*I–*Hin*dIII fragment containing 226 bp of the murine IgH heavy-chain enhancer with a perfect consensus octamer motif ATGCAAAT (described in Sturm et al. 1987) and *Eco*RI–*Pst*I restriction fragments from pUC119 containing binding sites cloned into the *Sma*I site; (lanes *3/7* and *5/10*) SV40 Octa1-binding site; (lanes *4, 6*) SV40 Octa1 with *dpm*8 mutations; (lanes *8, 11*) the unrelated SV40 enhancer sequence core A; (lanes *9, 12*) the HSV TAATGARAT sequence found to bind OBP100 (Baumruker et al. 1988).

The most stringent test of λC5 fusion protein-binding specificity is the assay of TAATGARAT motif binding shown in lanes 7–12 of Figure 2. The SV40 Octa1 site (lanes 7 and 10) and TAATGARAT motif used here (lanes 9 and 12) are matched at only 4 out of 14 positions within the OBP100-binding site (Baumruker et al. 1988). OBP100 and the λC5 fusion protein display the same relative affinity for these two binding sites (cf. lanes 7 and 9, and 10 and 12) but do not bind to an unrelated fragment (coreA, lanes 8 and 11). The relative affinity of λC5 fusion protein for a series of mutant and wild-type SV40 OBP100 site II sequences (Baumruker et al. 1988) was also similar to OBP100 (data not shown). Thus, the λC5-expressed protein displays the same binding specificity as authentic ubiquitous HeLa cell octamer-binding protein. In accordance with the nomenclature of Staudt et al. (1988) and Clerc et al. (1988), we refer to the gene encoding this octamer-binding protein as *oct*-1 and to the protein as Oct-1.

The oct-*1 gene is expressed in all human and mouse cell lines tested*

We expect the gene encoding the ubiquitous octamer-binding protein to be expressed ubiquitously. To detect expression, we performed an RNase protection assay, probing cytoplasmic RNAs from five human and two mouse cell lines with an antisense *oct*-1 RNA probe. This assay is very sensitive and stringent, because mismatches within an RNA–RNA duplex are targets for cleavage by RNase (Winter et al. 1985). Thus, it is a gene-specific expression assay. Figure 3 shows the result of probing cytoplasmic RNA from the human cell lines NTera-2D1, from which the cDNA library was derived; HeLa, either grown in spinner culture or on plates; 293, an adenovirus early-region-transformed embryonic kidney cell line; WI-38, a nontransformed lung fibroblast cell strain; and BJAB, a B-cell line containing a high concentration of the lymphoid-specific *oct*-2 mRNA (Staudt et al. 1988). The two mouse cell lines tested were NIH-3T3, and NS-1, a B-cell line. The human cell line RNAs all protect the complete *oct*-1 complementary sequences (468 nucleotides; see the asterisk in Fig. 3) within the probe (Fig. 3, lanes 2, 4, 6, 8, 9, and 10). Alkali treatment of the human RNA samples (lanes 3, 5, 7, and 11) prior to hybridization abolishes the signal, showing that the probe is being protected by RNA and not contaminating chromosomal DNA. The levels of *oct*-1 RNA are similar in each cell line except that the NTera-2D1 cell line has elevated *oct*-1 RNA levels (about three- to fivefold) and the WI-38 cell strain expresses less *oct*-1 RNA. These differences may reflect the relative growth rates of these cells because the NTera-2D1 cells grow rapidly, whereas the nontransformed WI-38 cells proliferate less rapidly.

Assay of the two mouse cell lines confirms that the protection assay is gene specific because the probe is no longer fully protected (Fig. 3, lanes 12 and 13). Instead, four shorter fragments (identified by the black dots) appear. These fragments probably represent RNase

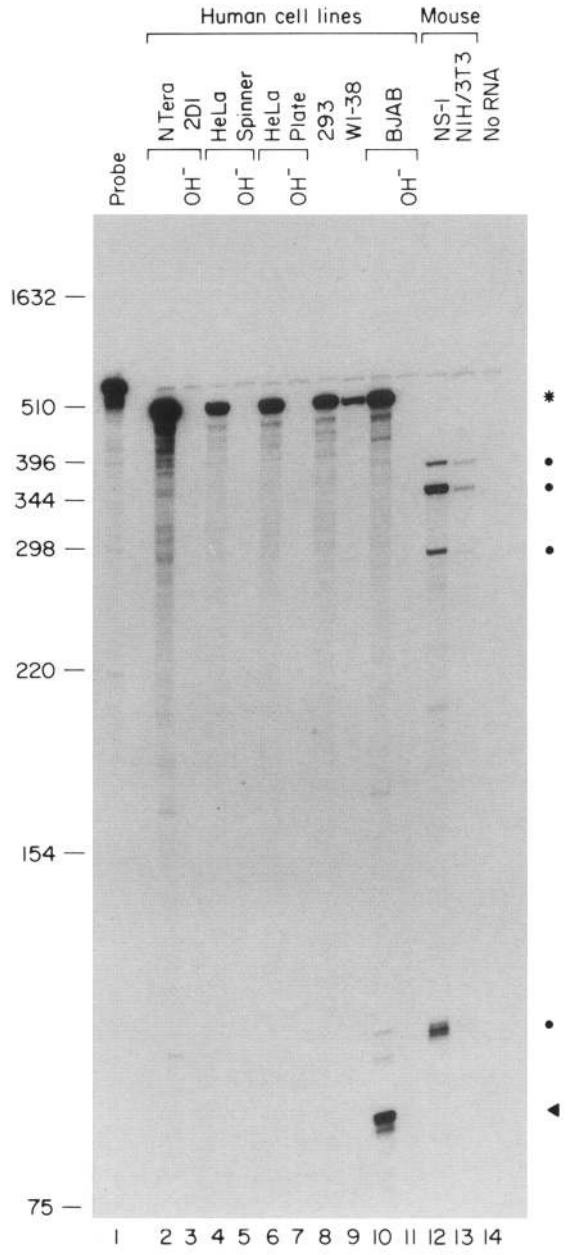

Figure 3. *oct*-1 mRNA expression in human and mouse cell lines. An RNase protection assay was performed using an internally labeled *oct*-1 antisense–RNA probe generated by T3 RNA polymerase from a template prepared by *Hae*III digestion of pBSλC51.1$^+$, which contains the 5′ 1.1-kb *oct-1 Eco*RI fragment from λC5 cloned into pBSM13$^+$. Except for the NIH–3T3 sample, which contained 10 μg of RNA, 20 μg of total cytoplasmic RNAs were hybridized, treated with RNase A and T_1, and fractionated on a 6% denaturing polyacrylamide gel as described (Zinn et al. 1983). (Lane *1*) Undigested probe alone; (lane *14*) mock protection assay done in the absence of complementary RNA. Human cell line RNAs: (lanes *2*, *3*) NTera-2D1; (lanes *4*, *5*) HeLa cells from spinner culture; (lanes *6*, *7*) HeLa cells grown on plates; (lane *8*) 293; (lane *9*) WI-38; (lanes *10*, *11*) BJAB. Lanes labeled OH$^-$ (*3*, *5*, *7*, *11*) contained RNA samples that had been hydrolyzed by pretreatment with 0.6 M sodium hydroxide for 5–10 min at ambient temperature, followed by neutralization with acetic acid and ethanol precipitation. The 468-nucleotide-long fragment (nucleotides 642–1109 in Fig. 5A) protected in all of the human RNA samples is marked by an asterisk (*), and the appearance of a B-cell-specific band in the BJAB sample is indicated by an arrowhead. The mouse cell lines were NS-1 (lane *12*) and NIH–3T3 (lane *13*); the positions of four major protected fragments in these samples are indicated by black dots.

cleavage at mismatches between the human antisense probe and mouse *oct*-1 transcripts. The antisense *oct*-1 probe appears to cross-hybridize to *oct*-2 mRNA because an additional ~100-nucleotide-long *oct*-1 probe fragment is protected in the BJAB-derived RNA (Fig. 3, lane 10, see arrowhead). This protected fragment is consistent with the very high degree of homology between the *oct*-1 and *oct*-2 genes within a portion of the region spanned by this probe (see below).

Ubiquitous and lymphoid-specific octamer-binding proteins are related antigenically

The experiments described above show that the protein encoded by the λC5 cDNA clone has the same DNA-binding specificity as HeLa cell octamer-binding protein and that the gene is expressed ubiquitously. To verify further the relationship between ubiquitous octamer-binding protein and the λC5 fusion protein, we prepared a rabbit antiserum directed against the β-galactosidase–octamer-binding fusion protein (see Materials and methods). To test the effects of this antiserum on Oct-1 DNA binding, different dilutions of the antiserum (into preimmune serum) were added to binding reactions with partially purified OBP100, and the effects on complex formation in a gel retardation assay are shown in Figure 4A. In these experiments, we observe a general and reproducible enhancement of complex formation by addition of preimmune rabbit serum, which is even more marked in the assays shown in Figure 4, B and C. Never-

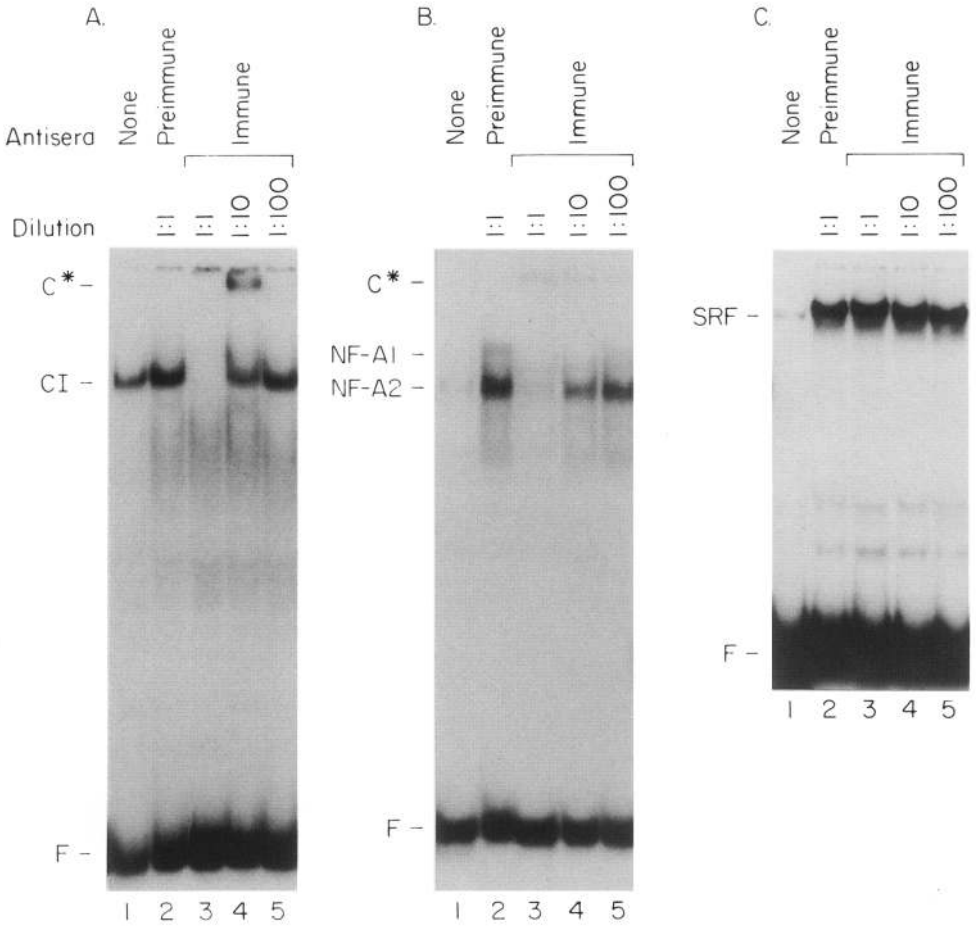

Figure 4. Rabbit anti-Oct-1 antiserum inhibits ubiquitous and lymphoid-specific octamer-binding activities. (*A*) A standard gel retardation assay (see Materials and methods) was performed using the wild-type SV40 site I probe in a 10-μl binding mix containing a partially purified HeLa cell fraction enriched for OBP100. The effect of adding to the prebinding mix, either no rabbit antiserum (lane *1*); 1 μl preimmune antiserum (lane *2*); 1 μl immune serum (lane *3*); 1 μl 1 : 10 dilution immune serum (lane *4*); 1 μl of 1:100 dilution immune serum (lane *5*). All immune serum dilutions were done with preimmune antiserum. Note the appearance of a second retardation complex (C*) upon dilution of the immune serum. (*B*) An *Xho*I end-filled fragment containing the SV40 Octa1 site was incubated with a total nuclear extract (Dignam et al. 1983) prepared from BJAB cells. The ubiquitous and B-cell-specific complexes NF-A1 and NF-A2 (Staudt et al. 1986) are indicated. Antisera were added to the binding mixes for lanes *1–5*, as in *A*. (*C*) An end-labeled *Eco*RI–*Hin*dIII fragment containing the c-*fos* SRE (Gilman 1988) was used in a binding mix with a heparin–agarose fractionated nuclear extract from H9 cells enriched for SRF (a gift from W. Ryan). The antisera added to the samples in lanes *1–5* were as described in *A*. The strong enhancement of sequence-specific complex formation in *B* and *C* is a nonspecific effect of the added sera. The exact explanation for this effect is not known but may reflect a general enhancing effect of the added protein from the serum.

theless, addition of the λC5 immune antiserum to the partially purified HeLa nuclear extract abolishes octamer-specific complex formation (Fig. 4A, lane 3) but not formation of complexes between the unrelated serum response factor (SRF) and the c-*fos* serum response element (SRE) (Fig. 4C, lanes 2–5). Tenfold dilution of the immune serum with the preimmune serum restored much of the octamer-specific complex but also generated a new complex with reduced mobility, labeled C* (lane 4). Further 10-fold dilution of the antiserum eliminated the C* complex and restored complex formation to the same level as with the preimmune serum.

We interpret these effects on octamer motif/Oct-1 complex formation as follows. At high concentrations, antibodies in the immune antiserum that are directed against the Oct-1 DNA-binding domain prevent complex formation. Upon dilution of the immune antiserum, the concentration of DNA-binding domain-specific antibodies is insufficient to inhibit all complex formation. At this concentration, antibodies directed against epitopes of the Oct-1 protein that are not involved in DNA binding can piggyback onto the Oct-1 protein/DNA complexes and create the C* complex.

These results show that the λC5 fusion protein is related antigenically to the ubiquitous octamer-binding protein. Similar experiments with a nonfractionated

HeLa nuclear extract and NTera-2D1 nuclear extract, using wild-type and mutant octamer probes, showed that the CI complex is the only evident octamer-specific complex to form in these cell extracts and that all of the CI complexes are abolished by addition of the λC5 immune antiserum. These results suggest that there is only one major octamer-binding activity in these cells and that the *oct*-1 gene cloned in λC5 encodes this protein.

Because the lymphoid-specific and ubiquitous octamer-binding proteins bind the octamer motif indistinguishably, we tested whether the DNA-binding domains of these two proteins might be antigenically related. Figure 4B shows that this is indeed the case. When immune antiserum is added to a BJAB nuclear extract, both ubiquitous (labeled NF-A1; Staudt et al. 1986) and lymphoid-specific (NF-A2) octamer complex formation are inhibited (Fig. 4B, lane 3); but upon dilution of the immune serum, the octamer-specific complexes are restored. These results indicate that the Oct-1 and Oct-2 proteins contain antigenically related DNA-binding domains.

When the preimmune and immune λC5 antisera were used in a Western blot of a HeLa cell nuclear extract, multiple species reacted, thereby preventing identification of the Oct-1 protein. In an immunoprecipitation of a radiolabeled 293 cell extract, however, among the many species that were immunoprecipitated, the only obvious difference between the preimmune and immune antisera precipitations was a protein of ~95 kD (data not shown). This molecular mass is consistent with the size of OTF-1 (Fletcher et al. 1987), NFIII (O'Neill and Kelly 1988), and OBP100 (Sturm et al. 1987).

Nucleotide sequence of oct-*1*

Figure 5A shows the 2584-nucleotide-long sequence of the *oct*-1 gene plus poly(A) sequence, as deduced from the λC5 insert and from a λgt10 *oct*-1 cDNA clone (λC7) described below. The λC5 insert (2424 nucleotides long, excluding the *Eco*RI linkers) was sequenced in entirety on both strands and over all restriction sites by the chain-termination method (see Materials and methods). The nucleotide sequence of the λC5 insert revealed an open reading frame, 762 amino acids long in-frame with the β-galactosidase-coding sequences. To characterize the 5′ and 3′ ends of the *oct*-1 mRNA, we isolated 16 independent λgt10 clones from a separate NTera-2D1 cDNA library (provided by J. Skowronski; Skowronski et al. 1988), using the λC5 insert as the probe. Analysis of these cDNAs showed that they were either similar in size or shorter than the λC5 insert, suggesting that the λC5 insert is nearly full length. Nucleotide sequence analysis of the 5′ and 3′ termini of the longest λgt10 cDNA clone, λC7, extended the 5′ *oct*-1 nucleotide sequence by 9 bp and revealed the nucleotide sequence of the entire 3′-untranslated region. Not shown in Figure 5A is the presence of an extra sequence in λC7, 102 nucleotides long, located near the 5′ end, which may reflect alternative splicing. An extensive analysis of λC7 and the other cDNAs will be described in a separate communication.

Determination of the exact position of the 5′ end of the *oct*-1 mRNA by primer extension and determination of mRNA size by Northern blotting have been hampered by a low abundance of *oct*-1 mRNA. The inability to extend significantly the 5′ sequence of λC5 by isolation of multiple cDNAs suggests that in λC5, the β-galactosidase fusion is into the 5′-untranslated region of the *oct*-1 mRNA. Therefore, we tentatively assign the first in-frame ATG codon (60 nucleotides downstream of the start of the *oct*-1 sequence shown in Fig. 5A) as the initiation codon for the *oct*-1 encoded protein Oct-1 and number both the nucleotide and amino acid sequence from this point. This assignment predicts that the Oct-1 protein is 743 amino acids long with an estimated molecular mass of 76 kD. However, as described below, the in vitro translation product from this cDNA has a considerably larger apparent molecular mass of between 90 and 94 kD.

The amino acid composition of the predicted Oct-1 protein is enriched for glutamine (11%), serine (14%), threonine (10%), and proline (7%). Whereas the proline residues are rather evenly distributed throughout the Oct-1 protein, the glutamine residues are highly concentrated within the amino-terminal third of the protein (26% from amino acid 22–268; see Fig. 5B). The entire protein is rich in serine and threonine residues, but the region from amino acid 441 to 560 is particularly Ser/Thr rich, as 50% (60/120) of these residues are either serine or threonine. A high concentration and uneven distribution of glutamine, and serine and threonine residues is reminiscent of the ubiquitous mammalian transcription factor Sp1 (Kadonaga et al. 1987). Unlike Sp1, however, the Oct-1 protein does not contain zinc finger motifs consistent with the Zn^{2+}-independent DNA-binding activity of the ubiquitous octamer-binding protein (Westin and Schaffner 1988). The central portion of the *oct*-1 coding sequence is very rich in both basic and acidic amino acids. From residue 272 to 436, 18% (29/165) of the residues are either arginine or lysine and 14% (23/165) are the acidic residues aspartic or glutamic acid. It is this 165-amino-acid-charged domain that contains the DNA-binding domain (see below).

Amino acid sequence comparison between the Oct-1 and Oct-2 proteins reveals a highly conserved domain

One of our primary interests in cloning the ubiquitous octamer-binding protein is to understand how it shares the same DNA-binding specificity as the lymphoid-specific octamer-binding protein but apparently differs in *trans*-activation function. Therefore, we compared the amino acid sequence of Oct-1 with the Oct-2 sequence determined by Clerc et al. (1988). As shown in Figure 6A, the amino-terminal regions of the two proteins display significant but patchy homology, in part reflecting the glutamine-rich composition of these two proteins within this region. The carboxy-terminal regions display only a few short regions of sequence similarity. Curiously, in Oct-2, these short sequence similarities are juxtaposed to create a leucine zipper motif (Landschulz

A.

```
                                                      1
GAGGAGCAGCGAGTCAAGATGAGAGTTCAGCCGCGGCGGCAGCAGCAGCAGACTCAAGAATGAACAATCCGTCAGAAACCAGTAAACCATCTATGGAGAGTGGAGATGGCAACACAGGC    60
      ProSerGlnAspGluSerSerAlaAlaAlaAlaAlaAlaAlaAspSerArgMETAsnAsnProSerGluThrSerLysProSerMETGluSerGlyAspGlyAsnThrGly    20
β-gal→|—λC5→

        BstXI
ACACAAACCAATGGTCTGGACTTTCAGAAGCAGCCTGTGCCTGTAGGAGGAGCAATCTCAACAGCCCAGGCGCAGGCTTTCCTTGGACATCTCCATCAGGTCCAACTCGCTGGAACAAGT   180
ThrGlnThrAsnGlyLeuAspPheGlnLysGlnProValProValGlyGlyAlaIleSerThrAlaGlnAlaGlnAlaPheLeuGlyHisLeuHisGlnValGlnLeuAlaGlyThrSer    60
  ├—Gln rich→

TTACAGGCTGCTGCTCAGTCTTTAAATGTACAGTCTAAATCTAATGAAGAATCGGGGGATTCGCAGCAGCCAAGCCAGCCTTCCCAGCAGCCTTCAGTGCAGGCAGCCATTCCCCAGACC   300
LeuGlnAlaAlaAlaGlnSerLeuAsnValGlnSerLysSerAsnGluGluSerGlyAspSerGlnGlnProSerGlnProSerGlnGlnProSerValGlnAlaAlaIleProGlnThr   100

CAGCTTATGCTAGCTGGAGGACAGATAACTGGGCTTACTTTGACGCCTGCCCAGCAACAGTTACTACTCCAGCAGGCACAGGCACAGGCACAGCTGCTGGCTGCTGCAGTGCAGCAGCAC   420
GlnLeuMETLeuAlaGlyGlyGlnIleThrGlyLeuThrLeuThrProAlaGlnGlnGlnLeuLeuLeuGlnGlnAlaGlnAlaGlnAlaGlnLeuLeuAlaAlaAlaValGlnGlnHis   140

TCCGCCAGCCAGCAGCACAGTGCTGCTGGAGCCACCATCTCCGCCTCTGCTGCCACGCCCATGACGCAGATCCCCCTGTCTCAGCCCATACAGATCGCACAGGATCTTCAACAACTGCAA   540
SerAlaSerGlnGlnHisSerAlaAlaGlyAlaThrIleSerAlaSerAlaAlaThrProMETThrGlnIleProLeuSerGlnProIleGlnIleAlaGlnAspLeuGlnGlnLeuGln   180
←Gln rich—┤                                                     ├—Gln rich→

CAGCTTCAACAGCAGAATCTCAACCTGCAACAGTTTGTGTTGGTGCATCCAACCACCAATTTGCAGCCAGCGCAGTTTATCATCTCACAGACGCCCCAGGGCCAGCAGGGTCTCCTGCAA   660
GlnLeuGlnGlnGlnAsnLeuAsnLeuGlnGlnPheValLeuValHisProThrThrAsnLeuGlnProAlaGlnPheIleIleSerGlnThrProGlnGlyGlnGlnGlyLeuLeuGln   220

GCGCAAAATCTTCAAACGCAACTACCTCAGCAAAGCCAAGCCAACCTCCTACAGTCGCAGCCAAGCATCACCCTCACCTCCCAGCCAGCAACCCCAACACGCACAATAGCAGCAACCCCA   780
AlaGlnAsnLeuGlnThrGlnLeuProGlnGlnSerGlnAlaAsnLeuLeuGlnSerGlnProSerIleThrLeuThrSerGlnProAlaThrProThrArgThrIleAlaAlaThrPro   260

                         HincII
ATTCAGACACTTCCACAGAGCCAGTCAACACCAAAGCGAATTGATACTCCCAGCTTGGAGGAGCCCAGTGACCTTGAGGAGCTTGAGCAGTTTGCCAAGACCTTCAAACAAAGACGAATC   900
IleGlnThrLeuProGlnSerGlnSerThrProLysArgIleAspThrProSerLeuGluGluProSerAspLeuGluGluLeuGluGlnPheAlaLysThrPheLysGlnArgArgIle   300
      ←Gln rich—┤      ├—Acidic/Basic→

AAACTTGGATTCACTCAGGGTGATGTTGGGCTCGCTATGGGGAAACTATATGGAAATGACTTCAGCCAAACTACCATCTCTCGATTTGAAGCCTTGAACCTCAGCTTTAAGAACATGTGC  1020
LysLeuGlyPheThrGlnGlyAspValGlyLeuAlaMETGlyLysLeuTyrGlyAsnAspPheSerGlnThrThrIleSerArgPheGluAlaLeuAsnLeuSerPheLysAsnMETCys   340

                                                                                             EcoRI
AAGTTGAAGCCACTTTTAGAGAAGTGGCTAAATGATGCAGAGAACCTCTCATCTGATTCGTCCCTCTCCAGCCCAAGTGCCCTGAATTCTCCAGGAATTGAGGGCTTGAGCCGTAGGAGG  1140
LysLeuLysProLeuLeuGluLysTrpLeuAsnAspAlaGluAsnLeuSerSerAspSerSerLeuSerSerProSerAlaLeuAsnSerProGlyIleGluGlyLeuSerArgArgArg   380

                                                                                                           BclI
AAGAAACGCACCAGCATAGAGACCAACATCCGTGTGGCCTTAGAGAAGAGTTTCTTGGAGAATCAAAAGCCTACCTCGGAAGAGATCACTATGATTGCTGATCAGCTCAATATGGAAAAA  1260
LysLysArgThrSerIleGluThrAsnIleArgValAlaLeuGluLysSerPheLeuGluAsnGlnLysProThrSerGluGluIleThrMETIleAlaAspGlnLeuAsnMETGluLys   420
                                                                                             [        H        ]——T——

                                                                   PflMI
GAGGTGATTCGTGTTTGGTTCTGTAACCGCCGCCAGAAAGAAAAAAGAATCAACCCACCAAGCAGTGGTGGGACCAGCAGCTCACCTATTAAAGCAATTTTCCCCAGCCCAACTTCACTG  1380
GluValIleArgValTrpPheCysAsnArgArgGlnLysGluLysArgIleAsnProProSerSerGlyGlyThrSerSerSerProIleLysAlaIlePheProSerProThrSerLeu   460
[        H        ] ←Acidic/Basic—┤          ├—Ser/Thr rich→

GTGGCGACCACACCAAGCCTTGTGACTAGCAGTGCAGCAACTACCCTCACAGTCAGCCCTGTCCTCCCTCTGACCAGTGCTGCTGTGACGAATCTTTCAGTTACAGGCACTTCAGACACC  1500
ValAlaThrThrProSerLeuValThrSerSerAlaAlaThrThrLeuThrValSerProValLeuProLeuThrSerAlaAlaValThrAsnLeuSerValThrGlyThrSerAspThr   500

                                HaeII
ACCTCCAACAACACAGCAACCGTGATTTCCACAGCGCCTCCAGCTTCCTCAGCAGTCACGTCCCCCTCTCTGAGTCCCTCCCCTTCTGCCTCAGCCTCCACCTCCGAGGCATCCAGTGCC  1620
ThrSerAsnAsnThrAlaThrValIleSerThrAlaProProAlaSerSerAlaValThrSerProSerLeuSerProSerProSerAlaSerAlaSerThrSerGluAlaSerSerAla   540

AGTGAGACCAGCACAACACAGACCACCTCCACTCCTTTGTCCTCCCCTCTTGGGACCAGCCAGGTGATGGTGACAGCATCAGGTTTGCAAACAGCAGCAGCTGCTGCCCTTCAAGGAGCT  1740
SerGluThrSerThrThrGlnThrThrSerThrProLeuSerSerProLeuGlyThrSerGlnValMETValThrAlaSerGlyLeuGlnThrAlaAlaAlaAlaAlaLeuGlnGlyAla   580
                                  ←Ser/Thr rich—┤
```

Figure 5. *(See following page for legend.)*

et al. 1988a; Clerc et al. 1988) that is not present in Oct-1.

The most striking region of similarity between Oct-1 and Oct-2 lies within the centrally located 160-amino-acid-long highly charged domain extending from residue 280 to residue 439 in Oct-1 (Oct-2 has three additional residues). In this region, these two proteins share identical residues at 142/160 positions. This represents 89% sequence similarity. The amino-terminal half of this conserved domain displays the highest degree of similarity; 74 out of 75 residues are identical. The amino-terminal portion of these 75 residues contains the sequence PSDLEELE, which is a consensus casein kinase II (CKII) phosphorylation site (Kuenzel et al. 1987), and there is a second adjacent potential CKII phosphorylation site (PSLEE in Oct-1) just upstream. The central portion of the shared charged domain is less related, displaying ~50% identity (14 out of 24 positions in Oct-1). The remainder of the charged domain (residues 378–439) is 89% identical. This latter region contains a putative helix-turn-helix motif, characteristic of many DNA-binding proteins (Pabo and Sauer 1984), and is distantly but significantly related to the homeo box motif first identified in *Drosophila* homeotic genes (McGinnis et al. 1984; Scott and Weiner 1984).

Figure 6B shows a sequence comparison between the

```
                                    NcoI
GCACAGTTGCCAGCAAATGCCAGTCTTGCTGCCATGGCAGCTGCTGCAGGACTAAACCCAAGCCTGATGGCACCCTCACAGTTTGCGGCTGGAGGTGCCTTACTCAGTCTGAATCCAGGG  1860
AlaGlnLeuProAlaAsnAlaSerLeuAlaAlaMETAlaAlaAlaAlaGlyLeuAsnProSerLeuMETAlaProSerGlnPheAlaAlaGlyGlyAlaLeuLeuSerLeuAsnProGly   620

ACCCTGAGCGGTGCTCTCAGCCCAGCTCTAATGAGCAACAGTACACTGGCAACTATTCAAGCTCTTGCTTCTGGTGGCTCTCTTCCAATAACATCACTTGATGCAACTGGGAACCTGGTA  1980
ThrLeuSerGlyAlaLeuSerProAlaLeuMETSerAsnSerThrLeuAlaThrIleGlnAlaLeuAlaSerGlyGlySerLeuProIleThrSerLeuAspAlaThrGlyAsnLeuVal   660

TTTGCCAATGCGGGAGGAGCCCCCAACATCGTGACTGCCCCTCTGTTCCTGAACCCTCAGAACCTCTCTCTGCTCACCAGCAACCCTGTTAGCTTGGTCTCTGCCGCCGCAGCATCTGCA  2100
PheAlaAsnAlaGlyGlyAlaProAsnIleValThrAlaProLeuPheLeuAsnProGlnAsnLeuSerLeuLeuThrSerAsnProValSerLeuValSerAlaAlaAlaAlaSerAla   700

GGGAACTCTGCACCTGTAGCCAGCCTTCACGCCACCTCCACCTCTGCTGAGTCCATCCAGAACTCTCTCTTCACAGTGGCCTCTGCCAGCGGGGCTGCGTCCACCACCACCACCGCCTCC  2220
GlyAsnSerAlaProValAlaSerLeuHisAlaThrSerThrSerAlaGluSerIleGlnAsnSerLeuPheThrValAlaSerAlaSerGlyAlaAlaSerThrThrThrThrAlaSer   740

                                                                    Poly A
AAGGCACAGTGAGCTGGGCAGAGCTGGGCTGCCAGAAGCCTTTTTCACTCTGCAGTGTGATTGGACTGCCAGCCAGGTTAATAAACTGAAAAATGTGATTGGCTTCCTCTCGCCGTGTTG  2340
LysAlaGln***                                                                                                              743

                                                                                                        Poly A
TGAGGGCAAAGGAGAGAAGGGAGAAAAAAAAAAAAAAAACCACACACACCCATACACACATACCAGAAAAAGAAAGAAAGGATGGAGACGGAACATTTGCCTAATTTTGTAATAAAACACT  2460
                  ←λC5—|

GTCTTTTCAGGATTGCTTCATGGATTGGAGAACTTTCTAACCAAAAATTAAAAAAAAAAAAAAAA                                                        2525
```

B.

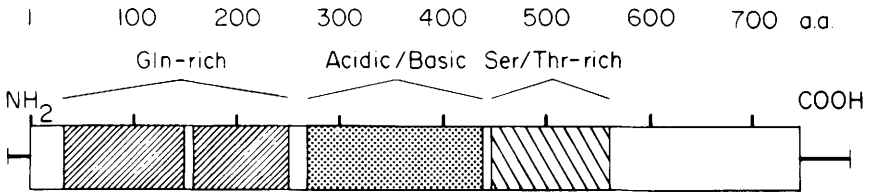

Figure 5. Nucleotide sequence of human *oct*-1 and deduced amino acid sequence of the Oct-1 protein. (*A*) The DNA sequence and deduced amino acid sequence of the *oct*-1 gene derived from the λC5 clone and the overlapping λC7 clone. The numbering system shown at the *right* of the sequence begins for both nucleotide and peptide sequences at the putative initiation codon ATG (labeled 1 over the underlined MET; Kozak 1984). The position of the fusion of *oct*-1 to β-galactosidase in λC5 is indicated, and the extra proline residue encoded by the *Eco*RI linker *oct*-1 junction sequence is underlined; the 3′ end of the λC5 clone is also indicated. Glutamine, acidic/basic, and serine/threonine-rich regions, discussed in the text, are bounded by arrows. The position of restriction enzymes used in the mapping of the *oct*-1 DNA binding-domain are highlighted on the sequence, and the helix-turn-helix motif is represented below the amino acid sequence by the two boxes (helixes) joined by the line (turn). The position of two consensus polyadenylation hexamers, AATAAA, is also indicated. (*B*) A schematic illustration of the structure of the deduced 743-amino-acid reading frame of the Oct-1 protein beginning at the predicted initiator MET (amino acid 1 in *A*). The regions outlined by arrows in *A* are boxed and shaded in *B*.

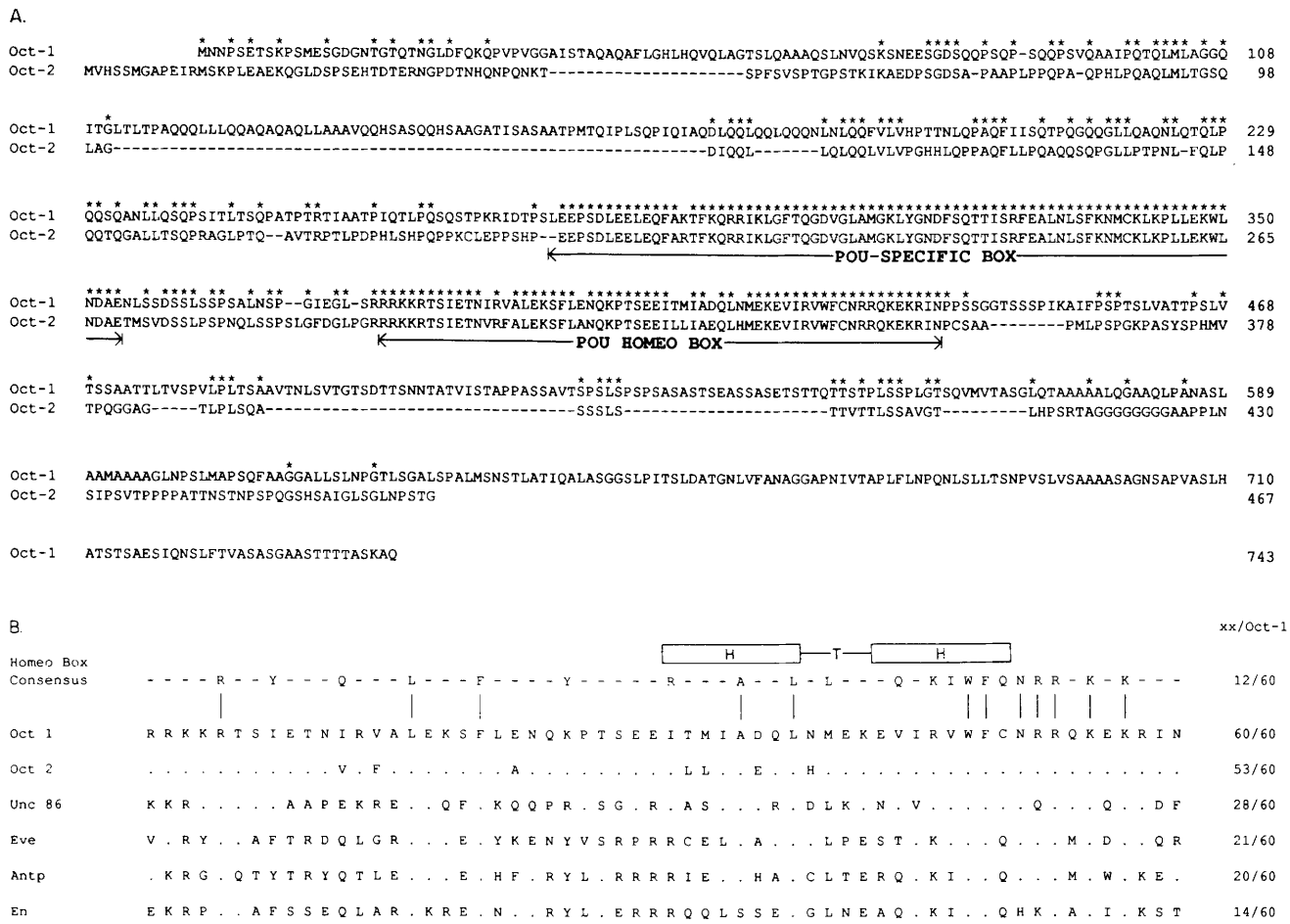

Figure 6. Sequence homology between Oct-1 and the lymphoid-specific octamer-binding protein Oct-2 and comparison between the Oct-1 homeo box domain and both prototypic *Drosophila* homeo box domains and the *C. elegans unc-86* gene product (Finney et al. 1988). (*A*) The Oct-1 and Oct-2 protein sequences are shown in the one-letter code and have been aligned for maximum homology. Positions of identity between the two sequences are indicated by the asterisks (*) above the sequence, and the POU-specific and POU-related homeo box domains are shown below the sequence. (*B*) The amino acid sequence of the Oct-1 homeo box region is compared with Oct-2 (Clerc et al. 1988), *unc-86* (Finney et al. 1988), *eve* (Macdonald et al. 1986), *Antp* (Schneuwly et al. 1986), and *en* (Poole et al. 1985) homeo box domains. The dots below the Oct-1 sequence represent positions of identity. A homeo box consensus sequence (Scott et al. 1989) is shown for comparison above the Oct-1 sequence. The position of variable amino acids in the consensus sequence is shown by dashes, and those conserved residues aligning with the Oct-1 sequence are indicated by the vertical lines. The number of residues shared with the Oct-1 sequence in the 60-amino-acid homeo box region is tabulated to the right of each sequence.

homeo box region of Oct-1 and Oct-2 and the homeo box region of three different prototypic *Drosophila* homeo box proteins: *even-skipped* (*eve*), *Antennapedia* (*Antp*), and *engrailed* (*en*) and the recently described *C. elegans unc-86* cell lineage gene product (Finney et al. 1988). Also included is a homeo box consensus sequence developed by Scott et al. (1989), after comparing 83 different homeo box motifs (including Oct-1). The Oct-1 homeo box domain shares identity with the homeo box domains of Eve, Antp, and En at 21/60, 20/60, and 14/60 positions, respectively. This region of Oct-1 also shares 11 out of the 21 highly conserved residues established in the homeo box consensus sequence. Although the sequence relationship is concentrated in the carboxy-terminal third of the domain, within and adjacent to the putative DNA recognition helix of the helix-turn-helix motif, there are four significant sequence identities outside of this region (see Fig. 6B).

Comparison of the Oct-1 and Oct-2 homeo-box-related domains to the homeo-box-related domain in the *unc-86* gene product (Finney et al. 1988) shows a considerably higher level of homology between one another (28/60 between Oct-1 and *unc-86*) than with the other homeo box proteins. The homeo box domain described in the pituitary-specific transcription factor Pit-1 (Ingraham et al. 1988) is also closely related to the Oct-1 homeo box. These results indicate that Pit-1, Oct-1, Oct-2, and *unc-86* form a related family. Surprisingly, this same family is even more related to one another within the 75-amino-acid region that is nearly identical

between Oct-1 and Oct-2 (see Herr et al. 1988 and discussion below). Because the entire 160-amino-acid-long charged region is conserved among these four proteins, we refer to it as the POU (Pit-1, Oct-1 and Oct-2, and *unc-86*)-related domain (pronounced 'pow').

The Oct-1 DNA-binding domain lies within the highly conserved POU domain

To map the Oct-1 DNA-binding domain, we used in vitro-translated Oct-1 protein in gel retardation assays, as first described by Hope and Struhl (1985). The entire λC5 insert was cloned into the BlueScript vector (pBS M13$^+$; Stratagene), which carries opposing T3 and T7 bacteriophage RNA polymerase promoters. To produce full-length Oct-1 protein and mutant derivatives, sense *oct*-1 RNA was transcribed in vitro with T7 RNA polymerase and the RNA translated in an in vitro rabbit reticulocyte translation system. The expression construct was designed so that the first AUG codon in the in vitro transcribed RNA is the first ATG in the λC5 insert (labeled nucleotide 1 in Fig. 5A). The in vitro-translated proteins were analyzed by polyacrylamide gel electrophoresis and subsequently used in a gel retardation assay with radiolabeled wild-type and mutant octamer-containing DNA fragments. The largest *oct*-1 in vitro translation product has an apparent molecular mass of about 90–94 kD, considerably larger than the predicted 76 kD of the 743-amino-acid-long λC5-encoded protein but, nevertheless, ~2–5 kD smaller than purified HeLa cell octamer-binding protein (data not shown). This difference could reflect usage of the wrong initiation codon or lack of appropriate modifications.

Figure 7A shows the different deletions that were tested for DNA-binding activity in a gel retardation assay. Figure 7B shows a gel retardation assay with the Oct-1 carboxy-terminal deletions. In each case, the proteins were mixed with labeled DNA fragments containing the wild-type SV40 octamer motif (odd-numbered lanes) or the double point mutant *dpm*8 (even-numbered lanes). Authentic HeLa cell octamer-binding protein was used as a marker (lanes 1 and 2). Addition of the rabbit reticulocyte extract itself (lanes 3 and 4) produced a weak endogenous octamer-binding activity that is not visible on the exposure shown. The full-length in vitro translation product produced a gel retardation complex with a slightly greater mobility than the authentic HeLa cell octamer-binding protein (cf. lanes 1 and 5), consistent with the somewhat smaller apparent molecular weight of the in vitro translated product. The largest carboxy-terminal truncation that still retained octamer-specific DNA-binding activity mapped to the carboxyl terminus of the homeo box domain (*Pfl*MI site at amino acid 440; see Fig. 5A) (cf. lanes 11 and 12). Truncation of an additional 26 residues removed the 'recognition' helix of the helix-turn-helix motif and abolished DNA-binding activity (*Bcl*I, lane 13), as did removal of the entire POU domain (*Hin*cII, lane 15). This result shows that the homeo box motif is required for DNA binding, whereas the sequences carboxy terminal to the homeo box are dispensable.

To determine whether the entire conserved POU domain is sufficient for DNA binding, we deleted the amino-terminal domain from residue 23 to 270, just 19 residues upstream of the POU domain (see Figs. 5A and 7A). By retaining the same 5′-untranslated region plus 70 nucleotides of coding sequence, we decreased the likelihood of a change in initiation codon usage during the in vitro translation. Figure 7C shows the effect of this amino-terminal deletion. Either with (*Hin*dIII; lanes 3 and 4) or without (*Pfl*MI; lanes 5 and 6) the sequences downstream of the homeo box domain, this amino-terminal deletion still binds specifically to the octamer motif. These results indicate that the POU-conserved domain is sufficient for DNA sequence recognition, although we cannot preclude that the amino-terminal 23 residues are not also required.

Discussion

We have described the isolation and characterization of human *oct*-1 cDNA clones encoding the ubiquitously expressed octamer-motif-binding protein Oct-1. This gene probably encodes the HeLa cell octamer-binding protein that has been isolated and named differently by several groups, i.e., OTF-1 (Fletcher et al. 1987), NFIII

Figure 7. Deletion mapping of the Oct-1 octamer-motif-binding domain. (*A*) A schematic illustration of the structure of the *oct*-1 protein product produced by in vitro translation of mRNA templates produced from pBS*oct*-1$^+$. The 743-amino-acid reading frame is boxed, and the glutamine, acidic/basic, and serine/threonine-rich regions described in Fig. 5, are shaded. The truncations and internal deletion used to map the DNA-binding region of the protein are outlined below the schematic figure, and the corresponding restriction enzymes and DNA-binding activity are listed on the *right*. The region sufficient for binding activity is indicated by the arrow. The positions of the POU domain and the POU-specific and homeo-related boxes are shown at the bottom. (*B*) The gel retardation assay was performed with the B20 end-labeled *Eco*RI–*Pst*I restriction fragments from pUC119 containing the wild-type SV40 OBP100-binding site 1 (odd-numbered lanes) or with the *dpm*8 mutations (even-numbered lanes). (Lanes *1*, *2*) Partially fractionated HeLa cell OBP100 as a marker; (lanes *3*, *4*) no RNA added to the rabbit reticulocyte in vitro translation system as a negative control; (lanes *5*, *6*) full-length Oct-1 in vitro translation product generated from *Hin*dIII-digested pBS*oct*-1$^+$. Truncations were with restriction enzymes as follows: (lanes *7*, *8*) *Nco*I; (lanes *9*, *10*) *Hae*II; (lanes *11*, *12*) *Pfl*MI; (lanes *13*, *14*) *Bcl*I; (lanes *15*, *16*) *Hin*cII. (*C*) The gel retardation assay was performed as in *B*: (lanes *1*, *2*) full-length Oct-1 in vitro translation product; (lanes *3*, *4*) the *Bst*XI–*Hin*cII amino-terminal deletion of pBS*oct*-1$^+$; (lanes *5*, *6*) the *Pfl*MI truncation of the *Bst*XI–*Hin*cII amino-terminal deletion.

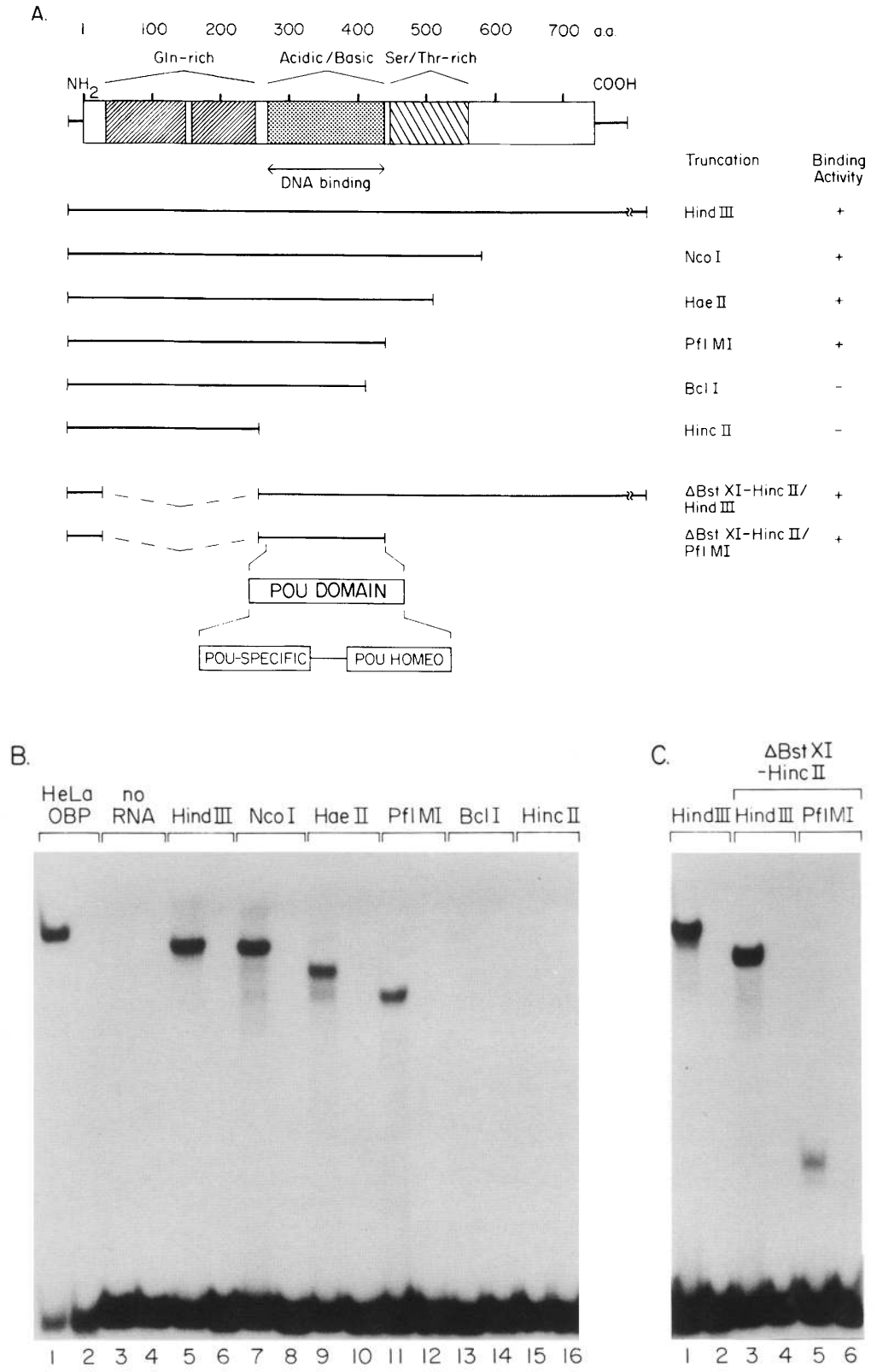

Figure 7. *(See facing page for legend.)*

(Pruijn et al. 1986; O'Neill and Kelly 1988), and OBP100 (Sturm et al. 1987). The arguments in favor of this interpretation are severalfold. First, the cloned octamer-binding protein has the same flexible DNA-binding specificity as HeLa cell Oct-1 protein. Second, the *oct*-1 gene is expressed in all of the five human and two mouse cell lines we examined and is thus expressed ubiquitously. Third, polyclonal rabbit antiserum raised against the cloned λC5 gene product shows an antigenic relationship between the ubiquitous octamer-binding protein and the ubiquitously expressed *oct*-1 gene. Fourth, the *oct*-1 in vitro translation product is nearly the same size as the authentic HeLa cell octamer-binding protein (see the similarly sized complexes formed in Fig. 7). Thus, *oct*-1 apparently encodes both a transcription factor (OTF-1) and an adenovirus replication factor (NFIII) (O'Neill et al. 1988), in a manner similar to CTF/NFI (Santoro et al. 1988).

We do not know, however, whether the actual polypeptide encoded by the λC5 clone can activate either transcription or replication, or both. Multiple Oct-1 species may exist that are of similar size but are generated by alternative splicing of the primary *oct*-1 transcript, and these may have different functions (e.g., activation of different promoters). Different CTF/NFI polypeptides encoded by alternatively spliced mRNAs have already been described, and these polypeptides have different transcription and replication activities (Santoro et al. 1988). In vivo expression studies of the Oct-1 protein from the cloned gene have been hampered by the ubiquitous endogenous expression of this protein. Nevertheless, if the Oct-1 protein encoded by the λC5 cDNA is coexpressed with a U2 snRNA promoter containing multimerized octamer motifs (see Tanaka et al. 1988), U2 transcription is weakly activated (M. Tanaka, N. Hernandez, and W. Herr, unpubl.). We are presently analyzing more *oct*-1 cDNA clones to determine whether the *oct*-1 transcript is alternatively spliced and to perform a functional analysis of the Oct-1 protein.

Oct-1 and Oct-2 share highly related DNA-binding domains

One of the major interests in determining the structure of the ubiquitous and lymphoid-specific octamer-binding proteins is to compare two proteins that bind to DNA with indistinguishable sequence specificity, but display different patterns of transcriptional activation. The amino acid sequence comparison shown in Figure 6A and the mapping of the DNA-binding domain of these two proteins (Fig. 7; Clerc et al. 1988) show that the regions responsible for octamer-specific DNA binding are very closely related. We do not know which regions are responsible for the *trans*-activation activities of these two proteins, but the different activities of the octamer motif in cell-specific, compared with ubiquitously expressed, promoters suggest that the regions of nonsimilarity between Oct-1 and Oct-2 are responsible for *trans*-activation. For example, the highly acidic domain (EEPSDLEELE; residues 280–289 in Oct-1) shared between Oct-1 and Oct-2 is probably not sufficient for activation of mRNA promoters because the octamer motif does not activate such promoters in the presence of Oct-1 (e.g., HeLa cells). This acidic domain is also probably not directly involved in activation of snRNA promoters by Oct-1 because an acidic GAL4 transactivation domain does not activate a U2 snRNA promoter containing GAL4 DNA-binding sites (Tanaka et al. 1988). The region of highest divergence between Oct-1 and Oct-2 is within the carboxy-terminal domain, in which the Oct-2 protein contains a leucine zipper motif (Landschulz et al. 1988a; Clerc et al. 1988). Perhaps this motif permits specific protein–protein interactions that confer activation of mRNA-encoding promoters.

The high degree of amino acid sequence similarity within the DNA-binding domains of Oct-1 and Oct-2 suggests that the *oct*-1 and *oct*-2 genes arose by duplication of a common ancestral gene. This high degree of amino acid sequence conservation appears to have been selected for strongly during evolution because comparison of the *oct*-1 and *oct*-2 (Clerc et al. 1988) nucleotide sequences encoding the 75-amino-acid-long region of highest identity (the POU-specific box) shows that 52% of the silent coding positions differ at the nucleotide level. This represents very little sequence conservation at these positions, strongly arguing that the similar amino acid sequences have been selected for at the level of the protein. What selective pressure may have maintained this remarkable degree of amino acid sequence conservation within the POU domain? Because these proteins bind indistinguishably to the same DNA sequences (Staudt et al. 1986), a clear possibility is that the entire region is conserved to maintain the same DNA-binding specificity. Pit-1 represents the other member of the POU family for which the DNA-binding site has been well characterized (Nelson et al. 1988). Curiously, this protein, in which the POU domain is less closely related to Oct-1 and Oct-2 than Oct-1 and Oct-2 are to themselves (see Herr et al. 1988), binds to a different, yet related, sequence motif (ATGNATA$^{A}/_{T}$A$^{A}/_{T}$T, compared with ATGCAAAT; Nelson et al. 1988). These comparisons suggest that the POU domain as a whole may bind to a similar class of DNA-binding sites and, conversely, that at least some of the differences between the Oct and Pit-1 POU domains are responsible for the differences in DNA sequence recognition.

One of the surprising results in the analysis of the Pit-1 (Nelson et al. 1988) and Oct-1 (OBP100; Baumruker et al. 1988) DNA-binding specificities is that the individual proteins can bind to very dissimilar DNA sequences. We have shown that Oct-1 protein can bind to sequences that are identical at only 4 out of 14 positions (AGTATGCAAAGCAT and GGCATCTCATTACC); nevertheless, simple point mutations within Oct-1 DNA-binding sites can abolish Oct-1 DNA binding. Therefore, Oct-1 exhibits a flexible, yet specific, pattern of DNA-binding site selection. This pattern of diverse sequence recognition is also exhibited by the *Drosophila eve*-encoded homeo box protein (Hoey and Levine 1988). In the case of *eve*, Hoey et al. (1988) have shown that the protein context of the homeo box is important in defining the relaxed binding specificity of the

Eve protein. In our studies, we have shown that the Oct-1 homeo box domain is essential for DNA-binding activity (Fig. 7). By analogy to Eve, the POU-specific box that flanks the homeo box domain in Oct-1 may reflect a highly conserved protein context that allows a POU class of homeo box domains to bind to a specific yet degenerate set of DNA-binding sites. Alternatively, the POU domain may not be involved in DNA binding but, rather, may be responsible for interaction with other proteins involved in transcription. At present, we favor the involvement of the POU-specific box in DNA binding because of the correlation between the degree of similarity in DNA-binding specificity and the degree of similarity in POU domain sequence among the Oct-1, Oct-2, and Pit-1 proteins.

The POU family of proteins represents a new class of homeo box proteins with ubiquitous and cell-specific members

The majority of vertebrate homeo box genes have been cloned by cross-hybridization to either *Drosophila* or other vertebrate homeo box domains (for review, see Dressler and Gruss 1988), whereas the POU homeo box proteins were identified by either biochemical (Oct-1, Oct-2, and Pit-1) or genetic (*unc-86*) assays. Now that this new family of homeo box proteins has been identified, hybridization studies with sequences from the POU domain may uncover more members of this class of homeo box proteins. The identification of both cell-specific and ubiquitously expressed transcription factors, along with a *C. elegans* gene product involved in regulating cell fate, affords a new perspective on homeo box proteins and the POU domain. The presence of well-characterized transcription factors in this class shows that homeo box proteins can be positive activators of gene expression. The cell-specific members show that these *trans*-activators can determine cellular differentiation and cell lineage (Clerc et al. 1988; Finney et al. 1988; Ingraham et al. 1988). The Oct-1 protein is unique in that it is a transcription (and DNA replication) factor with a homeo box domain that is ubiquitously expressed in different cell types. This result suggests that the homeo box is one of several possible DNA-binding motifs (like zinc fingers) that is used by both general and cell-specific transcription factors to regulate gene expression by sequence-specific DNA binding.

Materials and methods

Isolation of λC5 from an NTera-2D1 λgt11 cDNA library

An amplified λgt11 library made from cytoplasmic polyadenylated RNA of NTera-2D1 human teratocarcinoma cells, using oligo(dT) primer (SenGupta et al. 1986), was screened by the in situ filter detection method of Vinson et al. (1988). Recombinant phage at 4×10^5 were screened at a density of 2×10^4 per filter after lytic infection of *E. coli* strain Y1090 (Young and Davis 1983). The probe used in screening the filters was prepared by nick-translation (Maniatis et al. 1982) of purified 14×B17 *Hin*dIII–*Pst*I restriction fragment (Ondek et al. 1987), which contains 14 reiterated copies of SV40 Octa1 motif (Sturm et al. 1987). One positive plaque, named λC5, bound the 14×B17 probe in duplicate with high affinity and was subsequently purified by three rounds of plaque purification (Maniatis et al. 1982) using the same in situ binding procedure. Bacteriophage L20 encoding the DNA-binding domain of C/EBP linked to β-galactosidase (Landshulz et al. 1988b) was used as a negative control for nonspecific binding of the radiolabeled 14×B17 probe to λgt11 recombinant plaques, and a C/EBP-binding site probe (Vinson et al. 1988) was used to test the sequence-specificity of λC5 DNA binding.

Production of β-galactosidase fusion protein, Southwestern blotting, and gel retardation analysis

Phage stocks of λgt11, L20, and λC5 were prepared from plate lysates of Y1090 (Maniatis et al. 1982) and used to infect a culture of Y1089 (Young and Davis 1983) in log-phase growth. An overnight culture of Y1089 was diluted 1 : 50 in phage broth (L-broth, 10 mM Tris-HCl at pH 7.5, 10 mM $MgSO_4$) and grown for 1 hr; the cells were pelleted by centrifugation and resuspended in 0.2 volumes of fresh broth. Infections were done at an estimated moi of 5 for 10 min at room temperature, after which two volumes of phage broth prewarmed to 45°C was added and the culture grown for 10 min at 42°C. Synthesis of the β-galactosidase fusion protein was induced subsequently by addition of IPTG to 10 mM and incubation at 37°C for 1–2 hr before the cells were collected by centrifugation.

When total protein synthesis was examined, the cells were resuspended in SDS-loading buffer and heated to 90°C for 10 min before sample loading onto SDS–polyacrylamide gels. For Southwestern blot analysis, the proteins resolved on SDS–polyacrylamide gels were transferred electrophoretically from the gel to nitrocellulose at 100 mA for 12 hr in 25 mM Tris/190 mM glycine. The filters were then probed according to the protocol of Miskimins et al. (1985), using the same probe and buffer system used to screen the λgt11 recombinant phage filters (see above; Vinson et al. 1988).

For gel retardation analysis, λC5-infected extract was prepared from the cell pellet by first resuspending and repelleting the cells in 1 ml of ice-cold 50 mM Tris-HCl (pH 7.4), and 5 mM EDTA, followed by resuspension in 1 ml of 50 mM Tris-HCl (pH 7.4), 200 mM NaCl, 1 mM EDTA, 1 mM dithiothreitol (DTT), 0.1% NP-40, 10% glycerol, 5 M urea, and cell lysis by sonication. Insoluble material was removed by a 10-min Eppendorf centrifugation, and the supernatant was dialyzed against 50 mM Tris-HCl, 150 mM NaCl, 1 mM EDTA, 1 mM DTT, 0.1% NP-40, and 10% glycerol to allow renaturation of DNA-binding activity. The HeLa cell heparin/agarose-fractionated extract was the same as that described previously (Sturm et al. 1987). A nuclear extract (Dignam et al. 1983) prepared from the BJAB lymphoma cell line was the source of NF-A2 gel shifting activity (Staudt et al. 1986). The H9 T-lymphoma cell heparin–agarose fraction enriched for SRF was a gift from Dr. W. Ryan. Gel retardation analysis was performed as described previously (Sturm et al. 1987), with the modification of preincubating the extract with the binding mix for 10 min before addition of radiolabeled probe.

The probes for gel retardation were 3′ ^{32}P-end-labeled restriction fragments obtained from the following plasmids: pUC119CoreA/CoreA$^-$ (CoreA probe), pUC119B20wt$^+$ (SV40 site I Octa1 probe), and pUC119B20*dpm*8$^+$ (SV40 site I *dpm*8 probe). These plasmids were constructed by ligation into the *Sma*I site of pUC119 end-filled 28-bp *Xho*I-digested monomer fragments isolated from the plasmids pβ6xCoreA/CoreA, pβ6xB20 (*Sph*II–*Sph*I; Ondek et al. 1988), and pβ6xB20*dpm*8 (Tanaka et al. 1988). These fragments were resected for end-la-

beling as *Eco*RI–*Pst*I fragments. The IgH enhancer probe was a *Hin*dIII–*Pst*I fragment from pβIgH$^+$ (Sturm et al. 1987), the TAATGARAT probe was an *EcoRI–Pst*I fragment from cloned synthetic oligomers in pUC119 (Baumruker et al. 1988), and the SRE probe was an *Eco*RI–*Hin*dIII fragment from pUC119 containing a synthetic SRE site cloned into the *Hin*cII site (Gilman 1988).

Cell culture, RNA preparation, and RNase protection

The NTera-2D1, WI-38, and BJAB cells were obtained from P. Andrews (Wistar Institute), E. Moran (Cold Spring Harbor Laboratory), and P. Sharp (MIT), respectively. NTera-2D1, WI-38, HeLa, 293, NS-1, and NIH/3T3 cells were grown in 10% fetal bovine serum in Dulbecco's modified Eagle's minimal medium with 100 U/ml penicillin and 100 μg/ml streptomycin. BJAB cells were grown in 10% fetal bovine serum in RPMI medium suplemented with 50 μM mercaptoethanol. Spinner HeLa cells were grown to a density of 4×10^5 cells/ml in 5% calf serum. Cytoplasmic RNAs were prepared by NP-40 lysis, as described previously (Shepard et al. 1988), except that the DNase I digestion step was not included. The T3 RNA polymerase antisense *oct*-1 probe was prepared from a *Hae*III-digested subclone of the 5′ 1.1-kb λC5 EcoRI fragment into the *Eco*RI site of pBSM13$^+$ (Stratagene) (called pBSλC51.1$^+$), generating a probe that spanned from nucleotide 642 to 1109 in the *oct*-1 sequence shown in Figure 5A. The probe was labeled with [α-^{32}P]CTP at 400 Ci/mmole and purified by polyacrylamide gel electrophoresis. The probe was hybridized overnight to the different cytoplasmic RNAs treated with RNase A and T_1, and fractionated on 6% denaturing polyacrylamide gels, as described previously (Zinn et al. 1983).

Generation and use of rabbit anti-Oct-1 antisera

λC5 fusion protein expressed in Y1089, as described above, was purified for use as antigen by preparative SDS–polyacrylamide gel electrophoresis of the total protein lysate on 4% polyacrylamide gels. The 190-kD β-galactosidase–octamer-binding fusion protein was visualized by briefly staining the gel with 0.05% Coomassie Brilliant Blue R-250 in water and then destaining with several changes of water. For the first injection of two New Zealand white rabbits, gel slices containing ~50 μg of fusion protein were macerated by passing through a syringe repeatedly and emulsified in Freund's complete adjuvant (Harlow and Lane 1988). For subsequent injections, the fusion protein was recovered from the gel slice by electroelution into 3 ml of 0.2 M Tris-acetate (pH 7.4), 1% SDS, 100 mM DTT, and precipitation with five volumes of acetone before resuspension in water and emulsifying in Freund's incomplete adjuvant. Preimmune serum was obtained before the first injection, and the rabbits were boosted every 2 weeks with 50 μg of fusion protein after the first immunization. Five-milliliter test bleeds were obtained before each boosting to test for antibody production. One rabbit exhibited an immune response after the third injection, as assayed by Western blotting of the fusion protein. This antiserum was subsequently used for the gel retardation assays. anti-Oct-1 antisera was added to the binding reaction, together with the cell extract in the 10-min preincubation step in the absence of probe.

Sequencing of oct-*1 cDNAs*

Due to the presence of an internal *Eco*RI site, the cDNA insert of λC5 was excised as two *Eco*RI fragments and subcloned into the *Eco*RI site of pUC118. The nucleotide sequence of each fragment was determined by the shotgun method (Bankier and Barrell 1983). To prepare the random shotgun subclones, each *Eco*RI insert was isolated preparatively on a polyacrylamide gel, self-ligated with T4 ligase, sonicated, end-repaired, and fragments size-selected (~400–1000 bp) on a polyacrylamide gel. After elution, the fragments were blunt-end ligated into *Sma*I-digested and phosphatased M13mp8 phage vector. M13 single-stranded DNA templates were sequenced using the dideoxy method (Sanger et al. 1977), and the data collected and compiled in the Intelligenetics GENED program. Both fragments were sequenced in entirety on both strands. The junction between the two *Eco*RI sites was ascertained by sequencing across the site in the pBS*oct*-1$^+$ clone described below.

To isolate additional *oct*-1 cDNAs, a λgt10 library of cytoplasmic polyadenylated RNA from NTera-2D1 human teratocarcinoma cells (Skowronski et al. 1988) was screened for *oct*-1 cDNA clones by hybridization using standard methods (Maniatis et al. 1982). Both of the 1.1- and 1.3-kb *Eco*RI fragments from λC5 were nick-translated and used as probe. After three rounds of plaque purification, DNA was prepared from each clone (Manfioletti and Schneider 1988), and the size of the cDNA inserts was determined by *Eco*RI digestion. This procedure resulted in the isolation of 16 new *oct*-1 cDNA clones. The longest clone obtained, λC7, was sequenced by the dideoxy method (Sanger et al. 1977), using forward and reverse λgt10 primers flanking the *Eco*RI cloning site according to the manufacturer's specifications (New England Biolabs), and also with a 21-nucleotide-long oligonucleotide corresponding to 2389–2409 of the *oct*-1 cDNA sequence shown in Figure 5 and directed toward the poly(A) sequence.

In vitro transcription/translation of oct-*1 cDNA*

To synthesize protein for the mapping of the Oct-1 DNA-binding domain, an *Aval–Hpa*II restriction fragment that spans the entire *oct*-1-coding sequence and includes 13 bp of downstream λgt11 sequences from λC5 was cloned into the *Hin*cII site of pBSM13$^+$ (Stratagene) by blunt-end ligation after end-filling of the 5′ overhanging termini. The *Aval* site was created by the ligation of the *Eco*RI linker to the 5′ end of the *oct*-1 sequences in λC5 (CC/CGAG) and in the pBS subclone lies adjacent to the pBS *Eco*RI site. This plasmid is called pBS*oct*-1$^+$. A *Bst*XI–*Hin*cII deletion, pBS*oct*-1ΔBH, from this clone was made by digestion of plasmid DNA with these enzymes, blunt-ending of the 3′ *Bst*XI overhang with the large fragment of DNA polymerase I, and recircularization with DNA ligase. Generation of the correct junction sequence was ascertained by DNA sequencing. Plasmid DNA from these constructs was digested with restriction endonucleases and used as template for T7 RNA polymerase to transcribe uncapped RNA suitable for in vitro translation. The RNA templates (about 1 μg) were used as mRNA in a rabbit reticulocyte lysate (Promega; 50 μl total volume) to translate [^{35}S]methionine-labeled protein, according to the manufacturer's specifications, and protein production confirmed by analysis of the reaction products on SDS–polyacrylamide gels. Gel retardation analysis with ^{32}P-radiolabeled DNA fragments was performed directly using 1 μl of the translation mix as described previously (Sturm et al. 1987) with the preincubation step described above. The gels were acid-fixed, dried, and exposed to two sheets of X-ray film to shield the top sheet from ^{35}S radiation but still allow the ^{32}P-labeled DNA fragments to be detected.

Acknowledgments

We thank S. McKnight, C. Vinson, and K. La Marco for teaching the 'Rapid Cloning of Genes encoding Sequence-Specific DNA Binding Proteins' course at the Waksman Institute, April

21–22, 1988. We are indebted to J. Skowronski for the gift of the NTera-2D1 λgt11 and λgt10 cDNA libraries and for advice. We thank J. Clarke and M. Cleary for excellent technical assistance; M. Goodwin and J. Liebowitz for help in preparation of the manuscript; J. Duffy, D. Greene, and M. Ockler for artwork and photography; G. Otto and T. Burglin for computer analyses; M. Iacovacci and M. Zoller for oligonucleotide synthesis; B. Knowles, P. Andrews, B. Moran, and P. Sharp for gifts of cell lines; W. Ryan for the gift of SRE probe and SRF-containing extract; and E. Harlow for advice on antibody production and help with immunoprecipitation. We thank R. Clerc, P. Sharp, H. Ingraham, M. Rosenfeld, M. Finney, and G. Ruvkun for exchanging unpublished sequences and helpful discussions, and N. Hernandez, B. Stillman, and M. Gilman for critical readings of the manuscript. Last, we thank J.D. Watson for encouraging us to initiate this project. R.A.S. is a recipient of a Cancer Research Institute Fellowship (New York); G.D. was supported by a Long Island Biological Association (LIBA) postdoctoral fellowship; W.H. is a Rita Allen Foundation Scholar. This work was supported by U.S. Public Health Services grant CA-13106 from the National Cancer Institute.

Note

Sequence data described in this paper have been submitted to the EMBL/GenBank Data Libraries under accession number X13403.

References

Ares, M. Jr., M. Mangin, and A.M. Weiner. 1985. Orientation-dependent transcriptional activator upstream of a human U2 snRNA gene. *Mol. Cell. Biol.* **5**: 1560–1570.

Ares, M. Jr., J.-S. Chung, L. Giglio, and A.M. Weiner. 1987. Distinct factors with Sp1 and NF-A specificities bind to adjacent functional elements of the human U2 snRNA gene enhancer. *Genes Dev.* **1**: 808–817.

Bankier, A.T. and B.G. Barrell. 1983. Shotgun DNA sequencing. In *Techniques in the life sciences B5*, Elsevier, Ireland.

Baumruker, T., R. Sturm, and W. Herr. 1988. OBP100 binds remarkably degenerate octamer motifs through specific interactions with flanking sequences. *Genes Dev.* **2:**. 1400–1413.

Brent, R. and M. Ptashne. 1985. A eucaryotic transcriptional activator bearing the DNA specificity of a procaryotic repressor. *Cell* **43:** 729–736.

Carbon, P., S. Murgo, J.-P. Ebel, A. Krol, G. Tebb, and I.W. Mattaj. 1987. A common octamer motif binding protein is involved in the transcription of U6 snRNA by RNA polymerase III and U2 snRNA by RNA polymerase II. *Cell* **51:** 71–79.

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:** 1570–1581.

Davidson, I., C. Fromental, P. Augereau, A. Wildeman, M. Zenke, and P. Chambon. 1986. Cell-type specific protein binding to the enhancer of simian virus 40 in nuclear extracts. *Nature* **323:** 544–548.

Desplan, C., J. Theis, and P.H. O'Farrell. 1985. The Drosophila developmental gene , engrailed, encodes a sequence-specific DNA binding activity. *Nature* **318:**630–635.

Dignam, J.D., R.M. Lebovitz, and R.G. Roeder. 1983. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Res.* **11:** 1475–1489.

Dressler, G.R. and P. Gruss. 1988. Do multigene families regulate vertebrate development? *Trends Genet.* **4:** 214–219.

Falkner, F.G. and H.G. Zachau. 1984. Correct transcription of an immunoglobulin κ gene requires an upstream fragment containing conserved sequence elements. *Nature* **310:** 71–74.

Finney, M., G. Ruvkun, and H.R. Horvitz. 1988. The C. elegans cell lineage and differentiation gene *unc-86* encodes a protein containing a homeo domain and extended sequence similarity to mammalian transcription factors. *Cell* (in press).

Fletcher, C., N. Heintz, and R.G. Roeder. 1987. Purification and characterization of OTF-1, a transcription factor regulating cell cycle expression of a human histone H2b gene. *Cell* **51:** 773–781.

Frederiksen, S., P. Hellung-Larsen, and E. Gram Jensen. 1978. The differential inhibitory effect of α-amanitin on the synthesis of low molecular weight RNA components in BHK cells. *FEBS Lett.* **87:** 227–231.

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:** 1323–1330.

Gilman, M.Z. 1988. The c-*fos* serum response element responds to protein kinase C-dependent and -independent signals but not to cyclic AMP. *Genes Dev.* **2:** 394–402.

Gram Jensen, E., P. Hellung-Larsen, and S. Frederiksen. 1979. Synthesis of low molecular weight RNA components A, C, and D by polymerase II in α-amanitin resistant hamster cells. *Nucleic Acids Res.* **6:** 321–330.

Hanke, J.H., N.F. Landolfi, P.W. Tucker, and J.D. Capra. 1988. Identification of murine nuclear proteins that bind to the conserved octamer sequence of the immunoglobulin promoter region. *Proc. Natl. Acad. Sci.* **85:** 3560–3564.

Harlow, E. and D. Lane. 1988. *Antibodies: A laboratory manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

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:** 0000–0000.

Hoey, T. and M. Levine. 1988. Divergent homeo box proteins recognize similar DNA sequences in Drosophila. *Nature* **332:** 858–861.

Hoey, T., R. Warrior, J. Manak, and M. Levine. 1988. The DNA binding activities of the *Drosophila melanogaster* even-skipped protein are mediated by its homeo domain and influenced by protein context. *Mol. Cell. Biol.* **8:** 4598–4607.

Hope, I.A. and K. Struhl. 1985. GCN4 protein, synthesized in vitro, binds *his*3 regulatory sequences: Implications for general control of amino acid biosynthetic genes in yeast. *Cell* **43:** 177–188.

Ingraham, H.A., R. Chen, H.J. Mangalam, H.P. Elsholtz, S.E. Flynn, C.R. Lin, D.M. Simmons, L.W. Swanson, and M.G. Rosenfeld. 1988. A tissue specific transcription factor containing a homeo domain specifies a pituitary phenotype. *Cell* **55:** 519–529.

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:** 1079–1090.

Kakidani, H. and M. Ptashne. 1988. GAL4 activates gene expression in mammalian cells. *Cell* **52:** 161–167.

Kozak, M. 1984. Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. *Nucleic Acids Res.* **12:** 857–872.

Kuenzel, E.A., J.A. Mulligan, J. Sommercorn, and E.G. Krebs.

1987. Substrate specificity determinants for casein kinase II as deduced from studies with synthetic peptides. *J. Biol. Chem.* **262:** 9136–9140.

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:** 32–39.

Landolfi, N.F., J.D. Capra, and P.W. Tucker. 1986. Interaction of cell-type-specific nuclear proteins with immunoglobulin V_H promoter region sequences. *Nature* **323:** 548–551.

Landschulz, W.H., P.F. Johnson, and S.L. McKnight. 1988a. The leucine zipper: A hypothetical structure common to a new class of DNA binding proteins. *Science* **240:** 1759–1764.

Landschulz, W.H., P.F. Johnson, E.Y. Adashi, B.J. Graves, and S.L. McKnight. 1988b. Isolation of a recombinant copy of the gene encoding C/EBP. *Genes Dev.* **2:** 786–800.

Lenardo, M., J.W. Pierce, and D. Baltimore. 1987. Protein-binding sites in Ig gene enhancers determine transcriptional activity and inducibility. *Science* **236:** 1573–1577.

Macdonald, P.M., P. Ingham, and G. Struhl. 1986. Isolation, structure, and expression of even-skipped: A second pair-rule gene of Drosophila containing a homeo box. *Cell* **47:** 721–734.

Manfioletti, G. and C. Schneider. 1988. A new and fast method for preparing high quality lambda DNA suitable for sequencing. *Nucleic Acids Res.* **16:** 2873–2884.

Mangin, M., M. Ares, Jr., and A.M. Weiner. 1986. Human U2 small nuclear RNA genes contain an upstream enhancer. *EMBO J.* **5:** 987–995.

Maniatis, T., E.F. Fritsch, and J. Sambrook. 1982. *Molecular cloning: A laboratory manual.* Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Mason, J.O., G.T. Williams, and M.S. Neuberger. 1985. Transcription cell type specificity is conferred by an immunoglobulin V_H gene promoter that includes a functional consensus sequence. *Cell* **41:** 479–487.

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:** 163–167.

McGinnis, W., M.S. Levine, E. Hafen, A. Kuroiwa, and W.J. Gehring. 1984. A conserved DNA sequence in homoeotic genes of the Drosophila Antennapedia and bithorax complexes. *Nature* **308:** 428–433.

Miskimins, W.K., M.P. Roberts, A. McClelland, and F.H. Ruddle. 1985. Use of a protein-blotting procedure and a specific DNA probe to identify nuclear proteins that recognize the promoter region of the transferrin receptor gene. *Proc. Natl. Acad. Sci.* **82:** 6741–6744.

Nelson, C., V.R. Albert, H.P. Elsholtz, L.I.-W. Lu, and M.G. Rosenfeld. 1988. Activation of cell-specific expression of rat growth hormone and prolactin genes by a common transcription factor. *Science* **239:** 1400–1405.

Ondek, B., L. Gloss, and W. Herr. 1988. The SV40 enhancer contains two distinct levels of organization. *Nature* **333:** 40–45.

Ondek, B., A. Shepard, and W. Herr. 1987. Discrete elements within the SV40 enhancer region display different cell specific enhancer activities. *EMBO J.* **6:** 1017–1025.

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

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 NFIII. *Science* **241:** 1210–1213.

Pabo, C.O. and R.T. Sauer. 1984. Protein–DNA recognition. *Annu. Rev. Biochem.* **53:** 293–321.

Parslow, T.G., D.L. 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.* **81:** 2650–2654.

Poole, S.J., L.M. Kauvar, B. Drees, and T. Kornberg. 1985. The engrailed locus of Drosophila: Structural analysis of an embryonic transcript. *Cell* **40:** 37–43.

Pruijn, G.J.M., W. van Driel, and P.C. van der Vliet. 1986. Nuclear factor III, a novel sequence-specific DNA-binding protein from HeLa cells stimulating adenovirus DNA replication. *Nature* **322:** 656–659.

Ptashne, M. 1988. How eucaryotic transcriptional activators work. *Nature* **363:** 683–689.

Rosales, R., M. Vigneron, M. Macchi, I. Davidson, J.H. Xiao, and P. Chambon. 1987. In vitro binding of cell-specific and ubiquitous nuclear proteins to the octamer motif of the SV40 enhancer and related motifs present in other promoters and enhancers. *EMBO J.* **6:** 3015–3025.

Rosenfeld, P.J., E.A. O'Neill, R.J. Wides, and T.J. Kelly. 1987. Sequence-specific interactions between cellular DNA-binding proteins and the adenovirus origin of DNA replication. *Mol. Cell. Biol.* **7:** 875–886.

Sanger, F., S. Nicklen, and A.R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci.* **74:** 5463–5467.

Santoro, C., N. Mermod, P.C. Andrews, and R. Tjian. 1988. A family of human CCAAT-box-binding proteins active in transcription and DNA replication: Cloning and expression of multiple cDNAs. *Nature* **334:** 218–224.

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

Schneuwly, S., A. Kuroiwa, P. Baumgartner, and W.J. Gehring. 1986. Structural organization and sequence of the homeotic gene Antennapedia of Drosophila melanogaster. *EMBO J.* **5:** 733–739.

Scott, M.P. and A.J. Weiner. 1984. Structural relationships among genes that control development: sequence homology between the Antennapedia, Ultrabithorax, and fushi tarazu loci of Drosophila. *Proc. Natl. Acad. Sci.* **81:** 4115–4119.

Scott, M.P., J.W. Tamkun, and G.W. Hartzell. 1989. The structure and function of the homeo domain. *BBA Rev. Cancer* (in press).

SenGupta, D.N., B.Z. Zmudzka, P. Kumar, F. Cobianchi, J. Skowronski, and S.H. Wilson. 1986. Sequence of human DNA polymerase βmRNA obtained through cDNA cloning. *Biochem. Biophys. Res. Commun.* **136:** 341–347.

Shepard, A., J. Clarke, and W. Herr. 1988. Simian virus 40 revertant enhancers exhibit restricted host ranges for enhancer function. *J. Virol.* **62:** 3364–3370.

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:** 154–158.

Singh, H., J.H. LeBowitz, A.S. Baldwin, Jr., and P.A. Sharp. 1988. Molecular cloning of an enhancer binding protein: Isolation by screening of an expression library with a recognition site DNA. *Cell* **52:** 415–423.

Skowronski, J., T.G. Fanning, and M.F. Singer. 1988. Unit-length line-1 transcripts in human teratocarcinoma cells. *Mol. Cell. Biol.* **8:** 1385–1397.

Staudt, L.M., R.G. Clerc, H. Singh, J.H. Lebowitz, P.A. Sharp,

and D. Baltimore. 1988. Cloning of a lymphoid-specific cDNA encoding a protein binding the regulatory octamer DNA motif. *Science* **241:** 577–580.

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:** 640–643.

Sturm, R.A., S. Dalton, and J.R.E. Wells. 1988. Conservation of histone H2A/H2B intergene regions: a role for the H2B specific element in divergent transcription. *Nucleic Acids Res.* **16:** 8571–8586.

Sturm, R., T. Baumruker, B.R. Franza, Jr., and W. Herr. 1987. A 100-kD HeLa cell octamer binding protein (OBP100) interacts differently with two separate octamer-related sequences within the SV40 enhancer. *Genes Dev.* **1:** 1147–1160.

Tanaka, M., U. Grossniklaus, W. Herr, and N. Hernandez. 1988. Activation of the U2 snRNA promoter by the octamer motif defines a new class of RNA polymerase II enhancer elements. *Genes Dev.* **2:** (in press).

Vinson, C.R., K.L. LaMarco, P.F. Johnson, W.H. Landschulz, and S.L. McKnight. 1988. In situ detection of sequence-specific DNA binding activity specified by a recombinant bacteriophage. *Genes Dev.* **2:** 801–806.

Webster, N., J.R. Jin, S. Green, M. Hollis, and P. Chambon. 1988. The yeast UAS_G is a transcriptional enhancer in human HeLa cells in the presence of the GAL4 trans-activator. *Cell* **52:** 169–178.

Westin, G. and W. Schaffner. 1988. Heavy metal ions in transcription factors from HeLa cells: Sp1, but not octamer transcription factor requires zinc for DNA binding and for activator function. *Nucleic Acids Res.* **16:** 5771–5781.

Winter, E., F. Yamamoto, C. Almoguera, and M. Perucho. 1985. A method to detect and characterize point mutations in transcribed genes: Amplification and overexpression of the mutant c-Ki-*ras* allele in human tumor cells. *Proc. Natl. Acad. Sci.* **82:** 7575–7579.

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:** 174–178.

Young, R.A. and R.W. Davis. 1983. Yeast RNA polymerase II genes: Isolation with antibody probes. *Science* **222:** 778–782.

Zinn, K., D. DiMaio, and T. Maniatis. 1983. Identification of two distinct regulatory regions adjacent to the human β-interferon gene. *Cell* **34:** 865–879.

The ubiquitous octamer-binding protein Oct-1 contains a POU domain with a homeo box subdomain.

R A Sturm, G Das and W Herr

Genes Dev. 1988, **2:**
Access the most recent version at doi:10.1101/gad.2.12a.1582