The EMBO Journal vol.11 no.13 pp.4993-5003, 1992

The DNA binding specificity of the bipartite POU domain and its subdomains

C.Peter Verrijzer, Mark J.Alkema[1], Willem W.van Weperen, Hans C.Van Leeuwen, Marijke J.J.Strating and Peter C.van der Vliet

Laboratory for Physiological Chemistry, University of Utrecht, Vondellaan 24a, 3521 GG Utrecht and [1]Division of Molecular Genetics, The Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands

Communicated by H.S.Jansz

The POU domain is a conserved DNA binding region of ~160 amino acids present in a family of eukaryotic transcription factors that play regulatory roles in development. The POU domain consists of two subdomains, the POU-specific (POU_S) domain and a POU-type homeodomain (POU_{HD}). We show here that, like the POU_{HD}, the Oct-1 POU_S domain can bind autonomously to DNA but with low affinity. DNA binding studies and *in vitro* binding site selection revealed that the POU subdomains each have a different sequence specificity. The binding consensus of the POU_S domain [gAATAT(G/T)CA] and POU_{HD} (RTAATNA) respectively overlap the 'left half' and 'right half' of the POU domain recognition sequence [a(a/t)TATGC(A/T) AAT(t/a)t]. In addition to the core sequence, which is very similar to the octamer motif (ATGCAAAT), the flanking bases make a significant contribution to the binding affinity of the POU domain. Interestingly, at some positions the sequence preferences of the isolated POU subdomains are distinct from those of the POU domain, suggesting that the POU domain binding site is more than a simple juxtaposition of the POU_S and POU_{HD} target sequences. In addition, analysis of the binding kinetics of the POU domain and POU_{HD} indicates that the POU_S domain enhances the binding affinity by reducing the dissociation rate. Our results show that the POU domain proteins have DNA binding properties distinct from those of classic homeodomain proteins. We suggest a model for the way in which an additional conserved domain adds further specificity to DNA recognition by homeodomain proteins.

Key words: DNA binding/homeodomain/Oct-1/POU domain/transcription factors

Introduction

Specificity of sequence recognition by *trans*-acting factors is a critical prerequisite for coordinated gene expression (Johnson and McKnight, 1989; Mitchell and Tjian, 1989). Many sequence-specific DNA binding proteins can be classified according to the presence of a conserved protein domain such as the helix-turn-helix, zinc finger, leucine zipper, β-ribbon and helix-loop-helix motifs (Harrison, 1991). The POU domain is the DNA binding domain of a class of transcription factors involved in developmental regulation. It was initially discovered as a conserved region in three mammalian transcription factors (Herr *et al.*, 1988), Pit-1 (GHF-1) (Bodner *et al.*, 1988; Ingraham *et al.*, 1988; Karin *et al.*, 1990), Oct-1 (Sturm *et al.*, 1988), Oct-2 (Clerc *et al.*, 1988; Ko *et al.*, 1988; Müller-Immerglück *et al.*, 1988; Scheidereit *et al.*, 1988) and the *Caenorhabditis elegans* developmental control gene *Unc-86* (Finney *et al.*, 1988). Subsequently, several other POU domain proteins have been identified (reviewed in Rosenfeld, 1991). For several members of the POU gene family it has been demonstrated that they exert critical functions in the regulation of cell-type-specific gene expression, DNA replication, cellular proliferation, determination of cell identity and developmental control (reviewed in: Rosenfeld, 1991; Ruvkun and Finney, 1991; Schöler, 1991).

The POU domain proteins form a unique subfamily of homeodomain proteins. The POU domain is a bipartite structure of ~160 amino acids in which two conserved regions can be recognized, a 74-82 amino acid N-terminal POU_S domain and a 60 amino acid POU_{HD}, connected by a 15-27 amino acid variable spacer region. The POU_{HD} is distantly related to the classic homeodomain encoded by the homeobox. The homeobox is a 180 bp sequence, first found in several developmental regulatory genes of *Drosophila* and subsequently detected in other organisms ranging from yeast to man (reviewed in Scott *et al.*, 1989). The classic homeodomain confers sequence-specific DNA binding on the homeodomain proteins. The three-dimensional structures of the *Antennapedia, engrailed* and the Matα2 homeodomain-DNA complexes have been determined (Kissinger *et al.*, 1990; Otting *et al.*, 1990; Wolberger *et al.*, 1991). Despite a limited sequence identity between these homeodomains, the protein fold is highly conserved. The homeodomain consists of a triple α-helical structure and a flexible N-terminal arm. The second and third helices form a structure similar to the helix-turn-helix motif found in many prokaryotic repressors. Helix 3 is referred to as the recognition helix, since it makes base contacts in the major groove of the DNA. The N-terminal arm binds in the adjacent minor groove. Furthermore, there are extensive side chain contacts with the sugar-phosphate backbone.

Within the homeodomain superfamily, the POU domain proteins are defined by the divergent POU_{HD} and the presence of the POU_S domain. In contrast to the classic homeodomain proteins, the POU_{HD} is incapable of high affinity, sequence-specific DNA binding, which requires both POU subdomains (Sturm and Herr, 1988; Ingraham *et al.*, 1990; Kristie and Sharp, 1990; Verrijzer *et al.*, 1990a,b). In view of the functional differences between the classic homeodomain and the POU_{HD}, it is possible that there are also some structural differences between the two domains. However, sequence homology and prediction of α-helical structures suggest that the main features of the

© Oxford University Press

protein fold of the POU_{HD} and classic homeodomain are similar. In contrast, the structure of the POU_S domain is still unknown. Based on sequence homology, the POU_S domain can be further subdivided into the POU_S-A and POU_S-B regions. Comparison of the DNA contacts made by the entire POU domain or the POU_{HD} showed that deletion of the POU_S domain leads to a decrease in the number of DNA contacts and that much of the specificity of sequence recognition is determined by the POU_S domain (Ingraham *et al.*, 1990; Verrijzer *et al.*, 1990a,b; Aurora and Herr, 1992). The extensive DNA contacts made by the full POU domain are hard to reconcile with the binding paradigm of the homeodomain. Furthermore, binding of the intact POU domain, but not of the POU_{HD}, induces bending of the DNA (Verrijzer *et al.*, 1991). Taken together, these results suggest a direct role for the POU_S domain in DNA binding. However, the mechanism of action of the POU_S domain is still unclear. It might influence the folding and docking on the DNA of the POU_{HD}, or the POU_S domain could provide additional DNA contacts.

In this report we address how the POU_S domain contributes to sequence recognition by the POU domain. We show that the POU_S domain can bind DNA independently of the POU_{HD}. By using an *in vitro* binding site selection assay we have determined the target sequences for the intact POU domain as well as for the POU_S domain and the POU_{HD}. The recognition sequences of the POU_S domain and POU_{HD} correspond to the 5′-half and 3′-half of the POU domain binding site, respectively. Taken together, our results demonstrate that the POU_S domain and the POU_{HD} form two autonomous DNA binding structures that bind cooperatively via the covalent peptide linker.

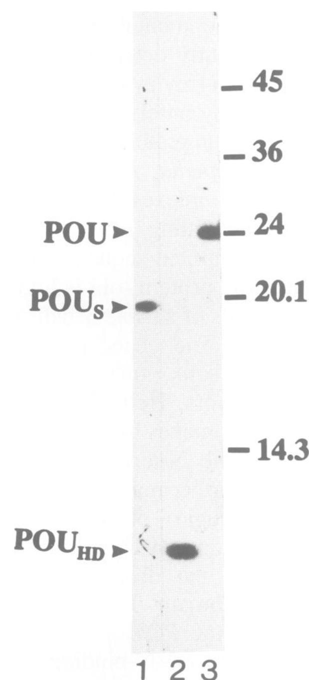

Fig. 1. Silver-stained 13% polyacrylamide gel of purified vaccinia expressed Oct-1 POU_S domain (lane 1), POU_{HD} (lane 2) and intact POU domain (lane 3). Indicated are the positions of protein size standards: ovalbumin (45 kDa), glyceraldehyde-3-phosphate dehydrogenase (36 kDa), bovine pancreas trypsinogen (24 kDa), soybean trypsin inhibitor (20.1 kDa) and bovine milk α-lactalbumin (14.3 kDa).

Results

Sequence requirements for optimal binding by the Oct-1 POU domain

We used a library of random sequence oligonucleotides and purified Oct-1 POU domain (Figure 1) to select for high affinity target sequences (Oliphant *et al.*, 1989; Pollock and Treisman, 1990; Thiessen *et al.*, 1990). The oligonucleotides were random at 24 positions, flanked by primer and cloning sequences. After incubation with a pool of these oligonucleotides, the POU domain was immunoprecipitated. Oligonucleotides that coprecipitated were eluted, amplified by the polymerase chain reaction (PCR) and used in a subsequent round of selection. The binding site selection was carried out four times. DNA recovered after each round of selection was used in a bandshift assay which revealed a strong enrichment in sequences recognized by the POU

A

1	CACAATATGCTAATA	29	TCAATGCAAATTAGTA
2	CCCATTATGCAAATTA	30	TGTTATGCAAATTCGGC
3	CCCATTATGCTAATGT	31	CTTATGCAAATTGGAT
4	CGATATGCTAATGCATG	32	AGTTATGCAAATCTGGG
5	CTACATATGCTAATGTATG	33	ATATGCAAATTCTGC
6	TGGATTATGCAAATATCGC	34	GTTTTATGCAAAACGGTC
7	ATGATTATGCAAATGTCGT	35	AATATGCAAATGAGCA
8	TAGATTATGCAAATAGCG	36	AATATGCAAATGTGAA
9	ATCGTTATGCAAATTTC	37	TATGCTAATCGTCC
10	GCAAGTATGCTAATTTCAA	38	CCATTATGCTAATATCGG
11	TTTCGTATGCTAATTTCAA	39	ACCATATGCAAATATCCG
12	TAAAGTATGCTAATTTTTG	40	TATGCAAATGTGCG
13	CTTCGTATGCAAATTCC	41	GGAGGTATGCAAATTTCCA
14	ATAAATATGCAAATCT	42	ATTTATGCTAATTTTGC
15	GCACATATGCAAATCTGGA	43	TGCATATGCTAATATGTC
16	ACTGATATGCAAATTGGC	44	GCTTTATGCAAATTTTGC
17	AGGAATATGCAAATATCCG	45	GCCGTATGCAAATTTTAT
18	CGGAATATGCAAATATGCG	46	CGCCGTATGCAAATATCGA
19	ATGCATATGCAAATCAGAG	47	GCCGTATGCAAATGTTCT
20	GTCAATATGCAAATCCCAA	48	TCCAATATGCAAATATACT
21	ATGAATATGCAAAT	49	TGAGATATGCAAATATCAC
22	ATACATATGCAAATACAAG	50	AATATGCAAATTTCAT
23	AAAATATGCAAATTCAAT	51	AAATATGCAAATTGCAT
24	TGGTATATGCAAATGCAA	52	ATATGCTAATTTCGT
25	TTGAATATGCAAAAACAGA	53	ACGTATATGGTAAT
26	TCTGCATGCAAATTTCTG	54	TAATTTATGCAAGGATCTT
27	TTTATGCAAATTCGTA	55	ATTATGTAAATTTACG
28	GTATGCAAATTTTTA	56	CTGATTATGTAAAATCTGG

B

	-6	-5	-4	-3	-2	-1	1	2	3	4	5	6	7	8	9	10	11	12	13
A	9	6	13	24	26	1	56	0	0	0	43	56	55	3	14	4	8	13	12
G	4	11	12	7	9	0	0	0	56	1	0	0	1	1	9	6	14	14	14
C	8	8	14	11	2	1	0	0	0	53	0	0	0	0	7	11	19	11	9
T	9	14	5	9	17	54	0	56	0	2	13	0	0	52	24	32	9	10	10
total	30	39	44	51	54	56	56	56	56	56	56	56	56	56	54	53	50	48	45

consensus a a/t T A T G C A/T A A T t/a t

C

site		K_D (M)	relative affinity
#50	5′AATATGCAAATTT3′	9 (±0.1) x10^{-11}	1
Ad4	AATATGCAAATAA	1 (±0.1) x10^{-10}	0.9
#20	AATATGCAAATCC	1.3 (±0.1) x10^{-10}	0.7
#19	CATATGCAAATCA	1.3 (±0.1) x10^{-10}	0.7
#47	CGTATGCAAATGT	1.8 (±0.1) x10^{-10}	0.5
U2	GCTATGCAAATAG	2.3 (±0.1) x10^{-10}	0.4
#26	TGCATGCAAATTT	1.8 (±0.1) x10^{-10}	0.5
#29	TCAATGCAAATTA	2.3 (±0.1) x10^{-10}	0.4
#5	CATATGCTAATGT	1.3 (±0.1) x10^{-10}	0.7
position	-3 1 5 10		

Fig. 2. Selected binding sites for the Oct-1 POU domain. (**A**) Aligned oligonucleotide sequences. Oct-1 POU domain binding sites were selected from a pool of random oligonucleotides as described in Materials and methods. The sequences of 56 cloned binding sites derived after four rounds of selection are aligned. (**B**) Determination of consensus sequence. At each position the frequency of the bases is shown. This quantification was used to obtain a consensus sequence. Upper case bases indicate the highly constrained core sequence, lower case indicates bases that are moderately constrained. (**C**) The importance of flanking bases for the binding affinity. K_D values of the Oct-1 POU domain for various binding sites were determined as described in Materials and methods. The K_D values are the average of three independent experiments and the standard error is indicated. The relative binding affinities of the various sites compared with the optimal binding sequence (#50) are shown.

domain (not shown). Shifted oligonucleotides from the fourth round of selection were eluted, subcloned and sequenced. All the clones analysed had differences in sequence, demonstrating that they arose from independently selected oligonucleotides. All but four of the 60 sequences obtained could be aligned and they contained either an octamer motif or a very similar sequence (Figure 2A).

In order to detect sequence constraints, selected sequences were tabulated for statistical analysis (Figure 2B). The consensus sequence derived is 5′-a(a/t)TATGC(A/T)AAT-(t/a)t-3′. The core sequence is very similar to the octamer element, and there is a very high constraint on the T at position −1. Furthermore, at position 5 within the octamer sequence there is a strong secondary preference for a T when no A is present. There is a moderate preference for A(A/T) in front of the highly constrained nonamer core and for (T/A)T at the right-hand side. Surprisingly, there is no preference for an A residue at position 10, which is conserved among many octamer sites, especially in the immunoglobulin promoters (Falkner and Zachau, 1984). Outside the indicated 13 bp region, no significant sequence constraints were detected.

Previously, mutagenesis has underscored the importance of the bases within the octamer motif for Oct-1 binding (e.g. Staudt *et al.*, 1986). In order to quantify the correlation between sequence constraints in flanking regions and their effect on the binding affinity, we determined the K_D values of POU domain binding to various binding sites (Figure 2C). The K_D values are compared with the consensus binding sequence (#50) and expressed as relative binding affinities. The A to T transition at position 5 (#5) had a small negative effect on the binding affinity. A change from T residues at positions 9 and 10 (#50) into two A residues (Ad4) had a slightly negative effect on the binding affinity and the same is true for a change to two C residues (#20) at these positions. At the 5′-end, a change from AAT (#50) to TGC (#26) results in a halving in affinity. Other mutations at flanking bases also cause weak but significant reductions in binding affinity. Although we did not perform an extensive analysis, those reductions are likely to have a cumulative effect. We conclude that the sequence requirements for binding by the Oct-1 POU domain correspond to sequence constraints in the binding site selection assay. Optimal DNA binding depends on a 13 bp region. Such an extended region has also been detected by contact point analysis (Baumruker *et al.*, 1988; Pruijn *et al.*, 1988; Verrijzer *et al.*, 1990b). In addition to the core octamer motif, flanking sequences are essential for high affinity binding by the Oct-1 POU domain.

The POU_S domain contributes to the flexibility and specificity of sequence recognition

Although its optimal binding sequence is clearly constrained, Oct-1 has a remarkable flexibility in recognition of suboptimal binding sites which is still with relatively high affinity and specificity (Baumruker *et al.*, 1988). We have compared the binding of the full POU domain and the POU_{HD} to diverse natural response elements, some of which, such as the heptamer element (CTCATGA) or the ICP4 TAATGARAT motif, have only very limited similarity to the consensus octamer sequence (ATGCAAAT). The POU domain and POU_{HD} were purified to near homogeneity from recombinant vaccinia virus infected HeLa cells (Figure 1). DNA binding was studied by a bandshift assay (Figure 3). The POU domain binds to the different sequences with varying affinity. Strong binding was observed for the U2, abII, Ad4, Ad2 and the H^+O^+ motif. The Oct-1 POU domain has a lower affinity for the ICP4 site while several target sequences for classic homeodomain proteins are not recognized. The heptamer sequence alone is a very weak binding site for the POU domain (lane 3) but binding to the heptamer element is facilitated when the adjacent octamer site is occupied (lane 4) (Kemler *et al.*, 1989; Poellinger *et al.*, 1989; LeBowitz *et al.*, 1989; Verrijzer *et al.*, 1992). The first complex in lane 4 is due to binding to the octamer sequence, the second results from subsequent binding to the heptamer site.

The POU_{HD} binds only a small subset of the POU

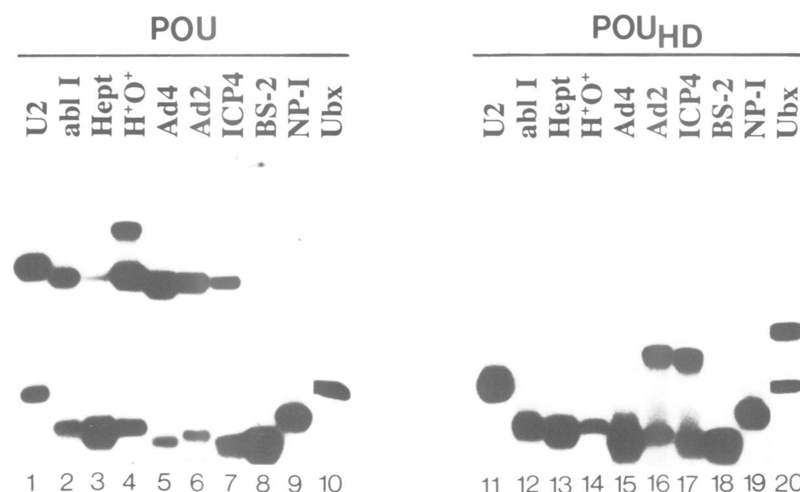

Fig. 3. The Oct-1 POU domain and POU_{HD} have different binding specificities. Binding of the purified intact POU domain (5×10^{-10} M) and POU_{HD} (3×10^{-10} M) was assayed in a bandshift experiment. Tested were labelled oligonucleotide probes containing the octamer sequence from the *Xenopus laevis* U2 enhancer (U2), the ablI promoter (ablI), the heptamer motif from the immunoglobin heavy chain (IgH) promoter (Hept), the intact octamer-heptamer site from the IgH promoter (H^+O^+), the canonical octamer present in the adenovirus 4 origin (Ad4), the degenerate adenovirus 2 octamer site (Ad2), the HSV ICP4 TAATGARAT motif (ICP4) and the recognition sites for three classic *Drosophila* homeodomain proteins, *Antennapedia* (BS-2), *engrailed* (NP-1) and *Ultrabithorax* (Ubx). References are given in Materials and methods and the sequences are shown in Figure 5C.

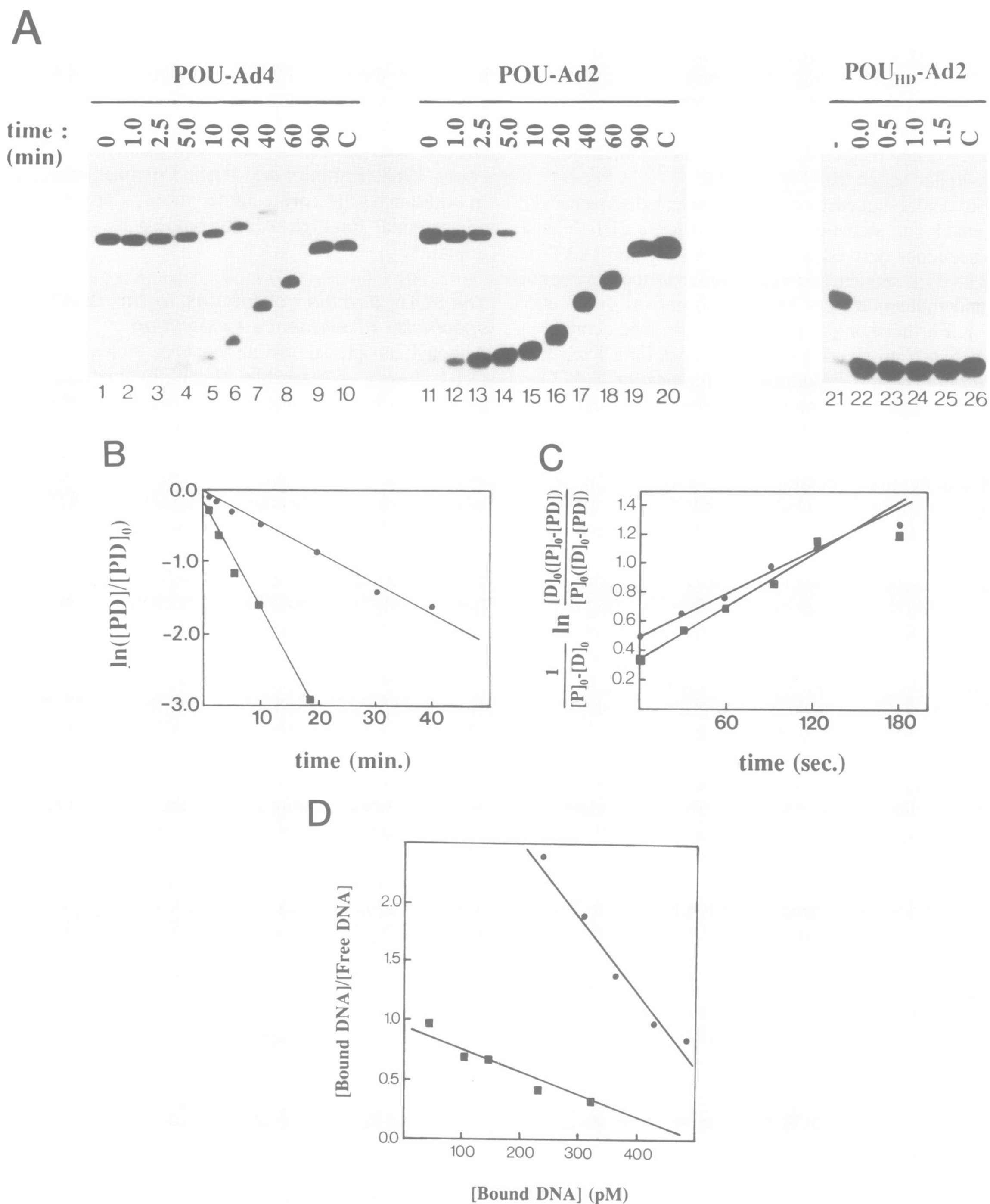

Fig. 4. Kinetic analysis of POU domain–DNA interactions. The interaction of the POU domain with two different recognition sequences was studied: the canonical octamer motif from the Ad4 origin (ATGCAAAT) and the degenerate Ad2 octamer (ATGATAAT). (**A**) Stabilities of protein–DNA complexes were assayed in bandshift experiments. The POU domain concentration was 7.5×10^{-10} M and the DNA concentration was 2×10^{-10} M. After 1.5 h at 0°C, a 200-fold excess of unlabelled competitor DNA was added and samples were loaded onto a running polyacrylamide gel at the indicated time points. As a control (C) the competitor DNA was added prior to addition of the probe (lanes 10, 20 and 26). A similar protocol was followed to measure the dissociation rate of the POU_{HD} (lanes 21–26). Lane 21 is without competitor DNA (–). (**B**) Determination of the POU domain dissociation rate constant (k_d). The results of the experiments shown in panel A were quantified as described in Materials and methods. $\ln([PD]/[PD]_0)$ was plotted as a function of time. From the slopes $k_d = 8\times10^{-4}\ s^{-1}$ and $k_d = 2.9\times10^{-3}\ s^{-1}$ were obtained for the Ad4 (●) and Ad2 (■) sites, respectively. (**C**) Determination of the POU domain association rate constant (k_a). POU domain and DNA were mixed at a final concentration of 7.5×10^{-10} M and 2×10^{-10} M, respectively. At various time points reactions were quenched by addition of an excess of unlabelled competitor and directly layered onto a running polyacrylamide gel. After quantification, a line of best fit of $1/([P]_0 - [D]_0)\ln\{[D]_0([P]_0 - [PD])/[P]_0([D]_0 - [PD])\}$ was plotted as a function of time. The slope of the lines resulted in $k_a = 5\times10^6\ M^{-1}s^{-1}$ and $k_a = 5.8\times10^6\ M^{-1}s^{-1}$ for Ad4 (●) and Ad2 (■), respectively. (**D**) Scatchard analysis of the interaction between the POU domain and DNA. In order to determine the equilibrium dissociation constant (K_D), a fixed amount of POU domain (6×10^{-10}) was incubated with increasing amounts of labelled DNA. After 1.5 h, bound and free DNA were separated on a native polyacrylamide gel. After quantification, [bound DNA]/[free DNA] was plotted against [bound DNA]. At equilibrium, $-1/K_D$ is equal to the slope. We determined $K_D = 1.5\times10^{-10}$ M for the Ad4 site (●) and $K_D = 5.5\times10^{-10}$ M for the Ad2 site (■).

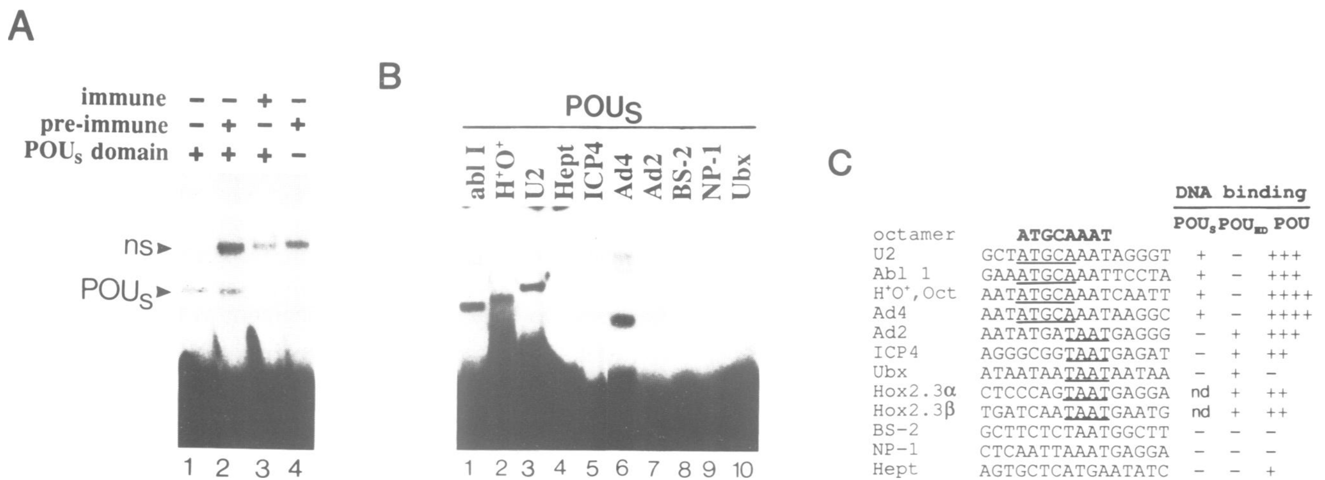

Fig. 5. The POU_S domain binds DNA with low affinity. (**A**) The POU_S domain can bind independently to the octamer sequence. Under conditions described in Materials and methods, binding to the Ad4 octamer sequence could be observed when 1 μg POU_S domain was added (lane 1). Immune serum directed against the POU domain or pre-immune serum was added as indicated (Verrijzer *et al.*, 1992). The POU_S domain–DNA complex and a non-specific band (n.s.) which results from serum addition are indicated. (**B**) Binding of the POU_S domain is specific for the canonical octamer. Several probes were tested for binding to the POU_S domain. The probes are indicated and are described in the legend to Figure 3. All binding assays were in the presence of 1 μg poly(dI-dC). (**C**) Comparison of binding sites of the POU_S domain, the POU_{HD} and the intact Oct-1 POU domain. Binding (+) or non-binding (–) of the POU_S domain and the POU_{HD} are indicated. Note that the binding affinity of the POU_{HD} is always higher than that of the POU_S domain. The data for the two Hox2.3 sites is from Verrijzer *et al.* (1990b). The POU_{HD} binds to its sites with a K_D of $\sim 3\times10^{-9}$ M. The K_D of the POU_S domain could not be determined. For the intact POU domain the relative binding affinity is indicated; the approximate binding affinities are: ++++, 1×10^{-10} M; +++, 3.5×10^{-10} M; ++, 1.5×10^{-9} M; +, 2×10^{-8} M.

domain target sequences, namely the Ad2 (ATGATAAT) and ICP4 sites. Both of these sequences contain a TAAT core which is essential in many homeodomain binding sites (Laughon, 1991). Previously, we have detected binding of the POU domain and POU_{HD} to related sequences containing a TAAT core (Verrijzer *et al.*, 1990b). However, when we tested several TAAT-containing target sequences for *Drosophila* homeobox proteins, none of these was recognized by the intact POU domain (lanes 8–10) while only the Ubx site was recognized by the POU_{HD} (lanes 18–20). This shows that additional sequences outside the TAAT core are important for binding by the POU domain or the POU_{HD}. Furthermore, the presence of the POU_S domain, which is essential for binding to the octamer, can hamper binding to other sites. The binding data are summarized in Figure 5C.

Kinetic analysis of POU domain–DNA interactions

We determined the kinetic DNA binding characteristics of the Oct-1 POU domain for two sites, the Ad4 canonical octamer (AATATGCAAATAA) and the degenerate Ad2 octamer motif (AATATGATAATGA). The Ad4 site is the strongest natural Oct-1 binding site: the Ad2 element has a 3- to 4-fold lower affinity. The contribution of the POU_S domain to the binding affinities of these sites is very different. Deletion of the POU_S region leads to an almost 1000-fold lower affinity for the Ad4 site but only a 7-fold reduction in affinity for the Ad2 site (Verrijzer *et al.*, 1990b).

The stability of the POU domain–DNA complex was measured in competition experiments. POU domain and labelled DNA were incubated on ice for 1.5 h, during which time equilibrium is reached. After addition of a 200-fold excess of cold competitor, aliquots were layered onto a running gel at various time points (Figure 4A). Analysis of the data yielded a dissociation rate constant (k_d) of $8\times10^{-4}s^{-1}$ for the Ad4 octamer and a k_d of $2.9\times10^{-3}s^{-1}$ for the Ad2 octamer (Figure 4B).

The association rate constants (k_a) were also determined. POU domain was incubated with labelled DNA and aliquots were removed over a short time course. Addition of a 200-fold excess of unlabelled competitor DNA prevented further association and samples were analysed on a polyacrylamide gel. The k_a values, determined from the plots in Figure 4C, were almost identical for the two sites, 5.8×10^6 $M^{-1}s^{-1}$ and 5.0×10^6 $M^{-1}s^{-1}$ for the Ad2 and Ad4 site, respectively.

The equilibrium dissociation constant (K_D) was determined by binding saturation experiments, in which the amount of DNA was increased in the presence of a fixed amount of protein, followed by Scatchard analysis. We obtained a K_D value of 1.5×10^{-10} M for the Ad4 site and a K_D of 5.5×10^{-10} for the Ad2 site (Figure 4D). Previously, by varying the concentration of POU domain with a fixed, limiting amount of DNA, we estimated K_D values of 9×10^{-11} M and 3.7×10^{-10} M for the Ad4 and Ad2 sites, respectively (Verrijzer *et al.*, 1990b). The K_D for a bimolecular reaction is equal to k_d/k_a. Using the data obtained from the kinetic experiments, K_D values of 1.6×10^{-10} M and 5.0×10^{-10} M were calculated for the Ad4 and Ad2 sites, respectively. These values, determined by three different methods, correspond well with each other.

The Ad2 octamer binds the POU_{HD} relatively well (K_D = 2.7×10^{-9} M) and all of the Ad2 probe could be complexed by the POU_{HD} (lane 21). When we mixed POU_{HD} bound to labelled oligonucleotides with an excess of unlabelled competitor and layered the sample directly onto a native gel, no POU_{HD}–DNA complex was detected (Figure 4A, lane 22). This indicates that the POU_{HD} dissociates very rapidly from the DNA in the slots. Due to the rapid dissociation of the POU_{HD} it was not possible to measure the association rate precisely. Nevertheless, in a binding experiment without addition of homologous competitor DNA, both the POU domain and POU_{HD} bound ~80% of the probe when they were layered directly after

mixing, but fully occupied the DNA when layered 1 min after mixing (not shown). This suggests that their association rates are approximately similar.

We conclude that the POU_{HD} has a very rapid association/dissociation rate. These experiments suggest that the POU_S domain contributes to the binding affinity by increasing the half life of the complex. The ~4-fold higher affinity of the Oct-1 POU domain for the Ad4 octamer than for the degenerate Ad2 sequence results from a difference in dissociation rate, not in association rate.

The POU_S domain specifically binds to the octamer element with low affinity

Various arguments suggest that the POU_S domain, when present in the intact POU domain, directly contacts the DNA. However, to date no independent DNA binding of the POU_S domain has been observed (Sturm and Herr, 1988; Ingraham *et al.*, 1990; Kristie and Sharp, 1990; Verrijzer *et al.*, 1990b). We produced the POU_S region in the vaccinia virus expression system and, employing conventional chromatographic procedures, obtained highly purified POU_S domain (Figure 1, lane 1). We used high concentrations of purified POU_S domain to assay its DNA binding capacity. Under conditions that reduce dissociation during electrophoresis (Materials and methods), weak binding to the octamer could be detected (Figure 5A, lane 1). The binding activity cofractionated with the POU_S domain during column chromatography (not shown). Antiserum directed against the POU domain abolished DNA binding (lane 3). In contrast, binding was not affected by addition of pre-immune serum (lane 2). Addition of serum gave rise to a non-specific band which was also present in the absence of the POU_S domain (lane 4). Several probes were tested for POU_S domain binding but only those containing a canonical octamer sequence were recognized (Figure 5B). This shows that although the binding affinity of the POU_S domain is weak, sequence recognition is clearly specific. These experiments revealed that the POU_S domain can bind autonomously to the octamer sequence.

Figure 5C shows an alignment of the binding sequences of the intact POU domain, the POU_{HD} and the POU_S domain. The relative binding affinity of the POU domain for the various sites is indicated. For the POU_S domain and the POU_{HD} only binding (+) or non-binding (−) are indicated. It should be noted that the binding affinity of the POU_{HD} is higher than that of the POU_S domain and that the POU_{HD} has a relatively high affinity for non-specific DNA (K_D = 10^{-7}-10^{-6} M; Verrijzer *et al.*, 1990b; data not shown). High affinity, sequence-specific DNA binding requires both subdomains. This is illustrated by a comparison of binding to the U2 and Ad2 sites. The U2 site is recognized by the POUS domain but not by the POU_{HD} whereas the

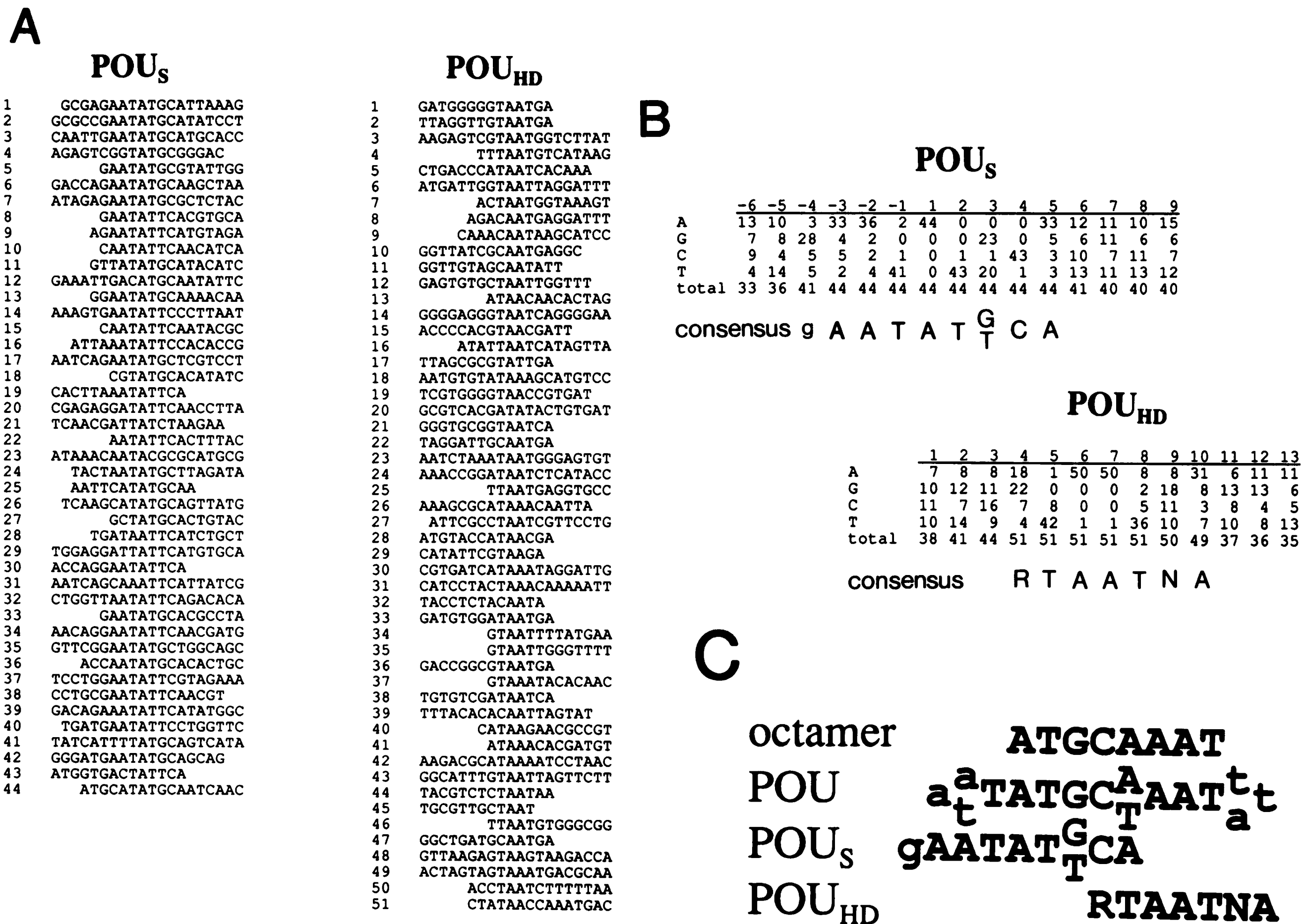

Fig. 6. Sites selected by the POU_S domain and the POU_{HD}. **(A)** DNA recovered after five rounds of selection by either GST-POU_S domain or GST-POU_{HD} fusion proteins were subcloned and sequenced. The sequences obtained were aligned. **(B)** Consensus sequences were determined by quantification of the bases present at each position. The derived consensus sequence for the POU_S domain and POU_{HD} are shown. **(C)** Alignment of the POU domain, POU_S domain and POU_{HD} target sequences.

POU_{HD}, but not the POU_S domain, binds to the Ad2 site. Nevertheless, both sites are recognized with a much higher affinity by the full POU domain, showing that the POU_S domain and POU_{HD} cooperate in DNA binding. Comparison of the various binding sites suggests that the POU_S domain recognizes the 'left part' of the canonical octamer sequence (5'-ATGCA-3') whereas the POU_{HD} contacts the AT-rich 'right part' of the octamer.

Target sequences of the POU_{HD} and POU_S domain

In order to determine directly the target sequences of the POU_S domain and the POU_{HD}, we have used the *in vitro* binding site selection similarly as for the intact POU domain. However, the coprecipitation of DNA bound by either the POU_S domain or the POU_{HD} with antiserum directed against the POU domain turned out to be very inefficient. Most epitopes are outside the POU_{HD} and therefore a high serum concentration was needed to immunoprecipitate the POU_{HD} which interfered with DNA binding. Moreover, the antiserum also affected the POU_S domain–DNA interactions. To circumvent these problems we expressed the POU domain, POU_S domain and POU_{HD} as fusion proteins with glutathione-S-transferase (GST) in bacteria. These polypeptides were purified to near homogeneity and their DNA binding specificities were indistinguishable from those of the isolated POU domain and POU subdomains expressed in the vaccinia system (data not shown). The fusion proteins were incubated with a pool of random oligonucleotides and, after DNA binding, were precipitated by binding of the GST moiety to glutathione–agarose beads. Complexed oligonucleotides were recovered, amplified by PCR and used in the next round of selection. Binding studies revealed a strong enrichment in POU_S domain or POU_{HD} binding sequences. After five rounds of selection, the oligonucleotides were subcloned and sequenced. All clones contained unique sequences, demonstrating that all binding sites were selected independently. Figure 6A shows the alignments of all POU_S domain or POU_{HD} recognition sequences obtained. Two distinct consensus sequences were derived from statistical analysis, 5'-gAATAT(G/T)CA-3' for the POU_S domain and 5'-RTAATNA-3' (R = purine) for the POU_{HD} (Figure 6B).

As illustrated in Figure 6C, the POU_S domain target site corresponds to the 5'-half of the POU domain binding sequence whereas the POU_{HD} target site corresponds to the 3'-half. In the centre of the POU binding site the POU_S domain and POU_{HD} sequences overlap at two positions. Interestingly, at some positions the sequence preferences of the isolated POU subdomains differ from those of the intact POU domain. For example, the G residue at position 3 is absolutely invariant in all isolated POU domain binding sites. The POU_S domain, however, frequently selects a T residue instead of the G. At position 4 both in the POU_S domain and in the POU domain binding sites, the C residue is highly constrained. In contrast, in the POU_{HD} binding sites there is a strong preference for a purine residue. Furthermore, at the flanking positions outside the octamer homology, there are also some differences in base preferences of the full POU domain and the isolated subdomains. In particular, the POU_S domain prefers a G residue at position -4 whereas no constraint at this position was detected for the intact POU domain. At position 9 the POU domain has a moderate preference for a T or A residue. In contrast, no such constraint was observed for the POU_{HD}. Finally, at position 10, the POU_{HD} prefers an A but the intact POU domain slightly prefers a T residue.

Discussion

Specific, high affinity binding of POU proteins requires, in addition to the POU_{HD}, an intact POU_S domain (Sturm and Herr, 1988; Ingraham *et al.*, 1990; Kristie and Sharp, 1990; Verrijzer *et al.*, 1990a,b). Both bandshift assays and binding site selection presented here demonstrate that the POU_S domain can bind DNA autonomously. Concordantly, the POU_S domain forms a stable fold in solution (Botfield *et al.*, 1992; N.Dekker, personal communication). The DNA binding affinity of the POU_S domain is low but highly sequence-specific.

Sequence specificity of the POU domain and its subdomains

In order to determine the optimal binding sequence for the Oct-1 POU domain, we selected binding sites from a library of random oligonucleotides. The derived consensus sequence is 5'-a(a/t)TATGC(A/T)AAT(t/a)t-3'. The selected optimal binding site is extremely A/T-rich and is tightly bound by the Oct-1 POU domain ($K_D = 9\times10^{-11}$ M). Positions outside the consensus core (TATGC(A/T)AAT) have a moderate but significant effect on the binding affinity and are moderately constrained in the binding site selection. It should be noted that degenerated elements recognized by Oct-1, like the ICP4 TAATGARAT motif, all have a lower affinity for Oct-1. However, in agreement with their function as natural response elements, binding to these sites is specific and with a relatively high affinity. The flexibility of sequence recognition by Oct-1 critically depends on the integrity of the POU domain. When isolated, the POU_S domain as well as the POU_{HD} are restricted in their sequence recognition.

We have used the site selection procedure to determine the Oct-1 POU_S domain and POU_{HD} recognition sequences. The sequences derived are 5'-gAATAT(G/T)CA-3' for the POU_S domain and 5'-RTAATNA-3' for the POU_{HD} which correspond to the 'left half' and 'right half' of the POU domain binding site respectively. This conformation of the POU domain on the DNA agrees well with a previous comparison of DNA contacts between the Oct-1 POU domain and the POU_{HD} (Verrijzer *et al.*, 1990b). This study showed that deletion of the POU_S domain leads to a loss of contacts in the left half of the POU domain binding site. Moreover, the comparison of POU domain- and POU subdomain-binding sites in bandshift assays suggested a similar positioning on the DNA. In the centre of the POU domain binding site, the POU_S domain and POU_{HD} contacts partially overlap. The isolated Oct-1 POU_{HD} prefers binding sites that have a TAAT sequence core. In this respect, it resembles the classic homeodomain proteins (Laughon, 1991; Scott *et al.*, 1989). However, only one of the three classic homeodomain target sequences tested here is recognized by the POU_{HD}, indicating that flanking sequences are also critical for binding; this is confirmed by the binding site selection assay.

The POU domain is a bipartite DNA binding structure

Both the protein fold and the docking on the DNA of divergent homeodomains, *engrailed*, MATα2 and

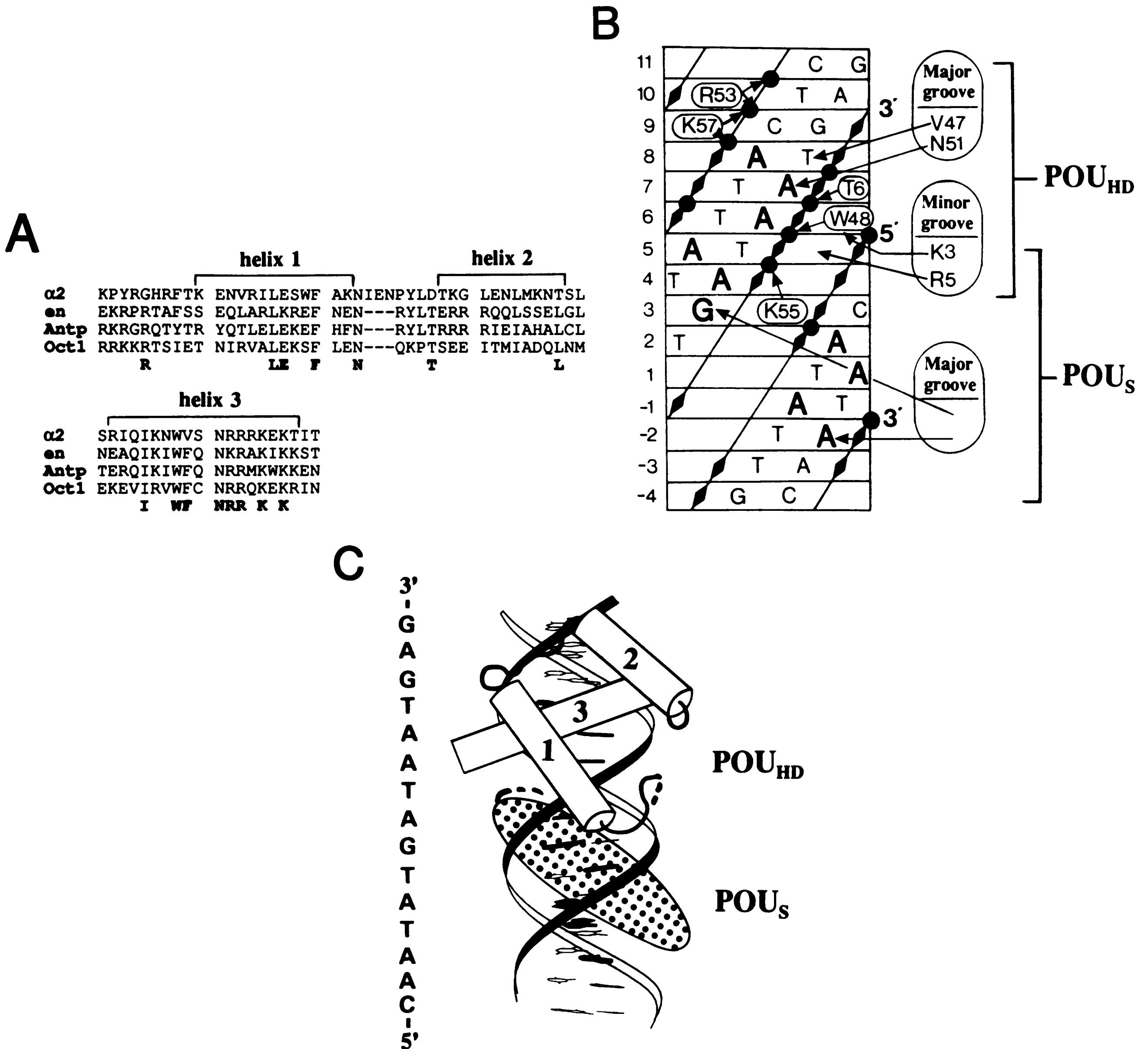

Fig. 7. Model for the interaction of the POU domain with DNA. (**A**) Sequence alignment of the MATα2 (α2), *engrailed* (*en*), *Antennapedia* (*Antp*) and Oct-1 homeodomains. Sequences were taken from Scott *et al.* (1989). Residues identical in Oct-1 and at least two other homeodomain proteins are shown below the sequences in bold. The three α-helices found in the classic homeodomain proteins are indicated. (**B**) Diagram of contacts made by the Oct-1 POU domain on the Ad2 octamer element. The DNA is represented as a cylindrical projection. Target sequences for the POU_{HD} and POU_S domain were determined from the experiments in Figures 3, 5 and 6 as well as by contact point analysis (Pruijn *et al.*, 1988; Verrijzer *et al.*, 1990b). Bases contacted are shown in bold, sugar contacts as filled diamonds and phosphate contacts as filled circles. The DNA contacts were used to model the POU_{HD} on the DNA assuming a protein fold similar to that of *engrailed* and MATα2 homeodomain DNA complexes (Wolberger *et al.*, 1991). Putative contacts by conserved residues are indicated. In this way the TAAT sequence in the Ad2 octamer and the *engrailed* binding site align perfectly. Note that a precise prediction of contacts, especially those made by residues with long side chains like arginine, is difficult to make. Furthermore, any of the residues in the basic N-terminal region of the POU_{HD} (positions 1–5) could probably make the indicated contacts with the minor groove of the DNA. Moreover, the pattern of backbone contacts made by the POU_{HD} is slightly different from those made by the *engrailed* or MATα2 homeodomains (Wolberger *et al.*, 1991). (**C**) Model for the POU domain–DNA complex. The contacts indicated in (B) are used to dock the POU_{HD} onto the DNA. Contacted bases that are identified are shown in bold. The detection of major groove contacts suggests that the POU_S domain contacts the DNA in the adjacent major groove, opposite to the POU_{HD}. The POU_S domain is schematically depicted as an ellipsoid.

Antennapedia are very similar (Kissinger *et al.*, 1991; Otting *et al.*, 1991; Wolberger *et al.*, 1991). These involve a set of residues that is highly conserved among all homeodomains, suggesting a general model for homeodomain–DNA interactions (Wolberger *et al.*, 1991). Figure 7A shows a comparison of the sequences of the Oct-1 POU_{HD} with MATα2, *engrailed* and *Antennapedia*. Conserved residues either pack in the hydrophobic core of the polypeptide and stabilize the overall fold or make critical contacts with the DNA. Interactions with the DNA made by the POU_{HD} are in good agreement with the structural data of homeodomain–DNA complexes. These protein–DNA contacts were used to position the POU_{HD} on its target sequence (Figure 7). This model suggests that DNA binding by the POU_{HD} is similar to that of the classic homeodomains. The DNA sequence contacted by the POU_S

domain is also indicated. Contact point analysis (Verrijzer *et al.*, 1990b) as well as UV cross-linking of the POU_S domain to BrdU-substituted DNA (Aurora and Herr, 1992) indicate that contacts are made within the major groove of the DNA. Taken together, all of these data suggest that the POU_S domain contacts the DNA opposite to the POU_{HD} in the adjacent major groove. Not depicted in the cartoon is that binding of the POU domain induces bending of the DNA (Verrijzer *et al.*, 1991).

The bipartite nature of POU domain–DNA interactions, first proposed by Sturm and Herr (1988), is now firmly established by several lines of evidence. (i) Each POU subdomain can bind independently to DNA and has its own recognition sequence (this paper). (ii) The POU_S domain and POU_{HD} form two autonomously folded subdomains which do not appear significantly to influence each other's folding within the POU domain (Botfield *et al.*, 1992; N.Dekker, personal communication). (iii) Cooperative DNA binding by the POU_S domain and the POU_{HD} requires a covalent linkage via the flexible peptide linker (Sturm *et al.*, 1988). Moreover, there are no interactions in solution between the POU_S domain and the POU_{HD}.

It is important to note that the POU subdomains do not interact with the DNA completely independently of each other. The POU_S domain can interfere with, or reprogram, binding of the POU_{HD}. For example, the Ubx site is bound better by the isolated POU_{HD} than by the intact POU domain. Similar observations have been made for Pit-1 (Ingraham *et al.*, 1990) and a detailed comparison of the DNA binding properties of chimeric POU domains, in which different protein segments are derived from either Pit-1 or Oct-1, indicated that segments of the POU domain can influence one another's sequence recognition specificity (Aurora and Herr, 1992). Furthermore, the comparison of the Oct-1 POU domain and POU subdomain recognition sites also demonstrates that the POU_S domain and POU_{HD} can modulate each other's sequence specificity. These effects might be caused by direct interactions that induce small but functional structural alterations in the POU subdomains. However, we consider it more likely that there is an indirect mechanism. For example, efficient binding of one POU subdomain might require docking on the DNA of the other subdomain. In addition, changes in the DNA structure, induced by binding of the POU domain but not by the subdomains, could alter their sequence requirements. In this respect it is noteworthy that binding of the POU domain induces bending of the DNA. In contrast, the POU_{HD} does not seem to affect the DNA structure (Verrijzer *et al.*, 1991).

Conserved DNA binding domains associated with the homeodomain

Despite the overall resemblance in DNA binding between the POU_{HD} and classic homeodomains, there are some essential functional differences. Most importantly, the binding affinity of the POU_{HD} for its target sequence is at least two orders of magnitude lower than that of classic homeodomain proteins for their target sequences (Affolter *et al.*, 1990; Percival-Smith *et al.*, 1990; Ekker *et al.*, 1991; Florence *et al.*, 1991). Also, the binding kinetics of the POU_{HD} are very different from those of the homeodomain proteins. The intact POU domain, however, binds its target sequences with an affinity comparable to that of the classic homeodomains ($K_D = 9\times10^{-11}-4\times10^{-10}$ M). Thus, the combination of a POU-type homeodomain with an additional DNA binding domain (the POU_S domain) does not lead to a polypeptide with a much higher DNA binding affinity than a single classic homeodomain. Instead, in the POU domain, the POU_{HD} is part of an integrated DNA binding structure that binds to a different class of recognition sequences than the classic homeodomain proteins.

It is interesting to compare POU domain transcription factors with Paired homeodomain proteins. In members of the latter family of developmental regulators, a subtype of the homeodomain, the Paired homeodomain, is associated with an additional DNA binding domain: the Paired domain. The Paired domain is a region of ~128 amino acids that, like the POU_S domain, can bind DNA in conjunction with the Paired homeodomain but also on itself (Chalepakis *et al.*, 1991; Treisman *et al.*, 1991). In contrast to the POU_S domain, the Paired domain binds to DNA with high affinity and exists in some proteins in the absence of a homeodomain, suggesting an independent function. It is highly unlikely that the POU_S domain can function as an autonomous DNA binding domain *in vivo*, since its binding affinity is very low, probably because it hardly makes contacts with the sugar–phosphate backbone (Verrijzer *et al.*, 1990b). In agreement with this notion, a POU_S domain and a POU-type homeodomain are always found together in one polypeptide.

Other examples of homeodomain proteins containing an additional DNA binding domain are the liver-specific transcription factor HNF-1/LBF-1, which has a region with homology to the POU_S-A domain (Baumhueter *et al.*, 1990; Frain *et al.*, 1989; Nicosia *et al.*, 1990; Mendel *et al.*, 1991), the fungal mating type protein β1-1, which shows slight similarity to the POU domain (Tymon *et al.*, 1991) and finally, a class of homeodomain proteins that contain in addition a Cys-His domain (LIM domain) which is reminiscent of metal binding regions (Freyd *et al.*, 1990; Karlsson *et al.*, 1990).

Materials and methods

Expression and purification of proteins

The Oct-1 POU domain, POU_{HD} and POU_S domain were overproduced in a vaccinia virus expression system as described by Verrijzer *et al.* (1990a,b). The Oct-1 POU domain and POU_{HD} were purified to near homogeneity essentially as described by Verrijzer *et al.* (1990b) and Mul *et al.* (1990). The Oct-1 POU_S domain was purified from 12 l of HeLa cells infected with POU_S-expressing recombinant vaccinia viruses (Verrijzer *et al.*, 1990b). The majority of the POU_S domain accumulated in the cytoplasm (not shown). The cytoplasmic extract was prepared as described previously (Verrijzer *et al.*, 1990a), adjusted to 150 mM NaCl and applied to a 40 ml DEAE column. The flow-through, which contained the POU_S domain, was diluted to 50 mM NaCl with buffer A [25 mM MES (pH 6.8), 1 mM DTT, 0.5 mM PMSF, 0.5 mM TPCK, 0.02% NP-40 and 10% glycerol] and loaded onto a 20 ml fast-flow S column. The POU_S domain in the flow-through was further fractionated on a 20 ml heparin–Sepharose column eluted with a 240 ml linear gradient from 50 to 350 mM NaCl. The POU_S domain eluted at 175 mM NaCl. Peak fractions were diluted to 35 mM NaCl in buffer B [buffer A with 25 mM Tris–HCl (pH 8.0) instead of MES] and loaded onto a 20 ml fast-flow Q column. The column was developed with a 240 ml linear gradient from 35 to 250 mM NaCl. The POU_S domain eluted at 125 mM NaCl. Peak fractions were pooled and run on a Superdex 75 HiLoad (16/60, Pharmacia) fast-flow gel filtration column run at 300 mM NaCl. These fractions were used in the binding assays. The purification was monitored by silver-stained SDS–polyacrylamide gels and immunoblotting using antibodies directed against the POU domain (Verrijzer *et al.*, 1992). Silver-staining indicated that the final POU_S fraction was purified to near homogeneity (Figure 1). The constructs expressing POU-GST, POU_S-GST and POU_{HD}-GST have been described

by Verrijzer et al. (1992). The GST fusion proteins were expressed and purified by affinity chromatography on a glutathione–agarose column essentially as described (Smith and Johnson, 1988). In addition, POU-GST and POU_{HD}-GST were further purified on a fast-flow S column run at pH 6.5 and eluted with a linear gradient from 50 to 600 mM NaCl in buffer A. Both fusion proteins eluted at ~175 mM NaCl. POU_S-GST was further purified on a fast-flow Q column developed with a linear gradient from 50 to 500 mM NaCl in buffer B (pH 8.0). POU_S eluted at ~80 mM NaCl.

DNA binding studies

The probes used for bandshift assays were oligonucleotides, end-labelled with T4 polynucleotide kinase and purified by preparative PAGE (Sambrook et al., 1990). The Oct-1 binding sites are shown in Figure 5C. The oligonucleotides have been described before: U2 snRNA enhancer (Mattaj et al., 1985), Ad4 and Ad2 (Verrijzer et al., 1990b), ICP4 (O'Hare and Goding, 1988), Hept (= H^+O^-) and H^+O^+ (Kemler et al., 1989; Verrijzer et al., 1992), abII (Meijer et al., 1990), NP-1 and Ubx (Desplan et al., 1988) and BS2–26 (Müller et al., 1988). Binding reactions were carried out on ice in 20 μl binding buffer [20 mM HEPES–KOH (pH 7.5), 100 mM NaCl, 1 mM EDTA, 1 mM DTT, 0.025% Nonidet P-40, 4% Ficoll]. For quantitative analysis no poly(dI–dC) was added, otherwise 1 μg was added to the binding reaction. Free DNA and protein–DNA complexes were resolved on a native polyacrylamide gel (39:1) run in 0.5×TBE. After electrophoresis, the gel was dried and exposed. To detect POU_S domain–DNA interactions, 15% polyacrylamide gels run at 4°C were used. This prevented dissociation during electrophoresis. Also, more POU domain and especially POU_{HD} remained complexed to the DNA under these conditions than in standard bandshift electrophoresis on a 6% polyacrylamide gel run at room temperature (data not shown).

Quantitative analysis

The DNA concentration of the oligonucleotides used for quantitative analysis was determined fluorimetrically. The total POU domain concentration was determined by using the Bio-Rad dye reagent with BSA as a standard. Binding reactions and bandshift analysis on 15% polyacrylamide gels run at 4°C were performed as described above. Dried gels were quantified by liquid scintillation counting of gel slices corresponding to free and bound DNA. Since the POU domain binds to the DNA as a monomer (Verrijzer et al., 1992), the binding reaction can be depicted as: $P + D \rightleftharpoons PD$ in which P, D and PD represent free protein, free DNA and protein–DNA complex, respectively. For determination of the association rate constant (k_a), labelled DNA (2×10^{-10} M) was added to a saturating amount of POU domain (7.5×10^{-10} M) in a volume of 200 μl. At the indicated time points, 20 μl aliquots were taken and reactions were quenched by addition of a 200-fold excess of unlabelled DNA and directly loaded onto a running polyacrylamide gel. After quantification, a best fit of $1/([P]_0 - [D]_0)\ln\{[D]_0([P]_0 - [PD])/[P]_0([D]_0 - [PD])\}$ was plotted as a function of time, where $[P]_0$ is the concentration of free protein at time 0, $[D]_0$ is the concentration of free DNA at time 0 and [PD] is the concentration of protein–DNA complex at time t. The k_a corresponds to the slope of the plot (Fried and Crothers, 1984). To determine the dissociation rate constants (k_d), labelled DNA (2×10^{-10} M) was incubated with a saturating amount of POU domain (7.5×10^{-10} M) in a volume of 200 μl. After 1.5 h at 0°C, a 200-fold excess of cold DNA was added. At various time points 20 μl aliquots were loaded onto a running polyacrylamide gel. After quantification, the k_d was calculated from the slope of a best fit of $\ln([PD]/[PD]_0)$ plotted against time since $-k_dt = \ln([PD]/[PD]_0)$ where $[PD]_0$ is the concentration of the protein–DNA complex at time 0 (Fried and Crothers, 1984). For Scatchard analysis, a range of labelled DNA concentrations was incubated with a fixed amount of POU domain (6×10^{-10} M). After 1.5 h, bound and free DNA were separated on a polyacrylamide gel. After quantification, [bound DNA]/[free DNA] was plotted against [bound DNA]. The equilibrium dissociation constant, K_D, in a bimolecular reaction can be obtained from the slope of this plot, which is equal to $-1/K_D$. All quantitative assays were performed in triplicate and variation was within 10% of the values obtained.

Binding site selection

All steps of the *in vitro* binding site selection were performed essentially as described by Pollock and Treisman (1990). Binding reactions were performed for 30 min at 0°C with ~0.03 ng POU domain, 0.4 ng (first selection round) or 0.2 ng (subsequent selection rounds) double stranded random oligonucleotides and 200 ng poly(dI-dC). 1 μl of polyclonal antibody directed against the Oct-1 POU domain (Verrijzer et al., 1992) was added and subsequently complexes were precipitated with protein A–Sepharose (Pharmacia). Oligonucleotides that co-immunoprecipitated were eluted and amplified in 15 PCR cycles. Part of this material was used in the next round of selection. After four successive rounds of selection, the oligonucleotides were labelled during the PCR reaction and used in a bandshift assay. Material bound to the POU domain was eluted and cloned in the vector pSP72 (Promega). Sequencing reactions were performed with a T7 polymerase sequencing kit (Promega) on double stranded DNA templates with standard Sp6 and T7 primers. Site selection with the POU (sub)domain–GST fusion proteins was performed similarly. However, instead of an immunoprecipitation, fusion proteins were immobilized by binding to glutathione–agarose beads (Verrijzer et al., 1992). In agreement with its relaxed sequence-specificity, the site selection by the POU_{HD} was influenced by the amount of non-specific competitor DNA present in the binding reaction. The POU_{HD} hardly discriminated between distinct sites at low (<100 ng) amounts of poly(dI-dC). Consistently, high amounts (>1 μg) of poly(dI-dC) efficiently competed for binding to the oligonucleotides. In contrast, the concentration of poly(dI-dC) did not affect sequence recognition by the POU_S domain or POU domain (data not shown). Therefore, the amount of poly(dI-dC) present in the GST-POU_{HD} binding reactions was increased to 500 ng in order to optimize the selection of recognition sequences. To determine the K_D values of the sites derived from the *in vitro* binding site selection, fragments containing binding sites were generated, labelled and isolated by standard methods (Sambrook et al., 1989). A fixed amount of labelled DNA was incubated with increasing amounts of pure Oct-1 POU domain and complex formation was studied by bandshift analysis. After quantification, the K_D values were determined as described by Verrijzer et al. (1990b)

Acknowledgements

We are indebted to Drs Winship Herr, Dies Meijer, Peter O'Hare, Cees Vink, Ronald Plasterk, Ron de Jong and Markus Affolter for making the various plasmids available to us. We thank Rajeev Aurora and Winship Herr for communicating data prior to publication. We are grateful to Joost van Oosterhout for his contribution to the sequence analysis. Thanks are also due to Anton Berns, Maarten van Lohuizen, Niek Dekker and Yvonne Mul for discussions and helpful suggestions and to Jos Domen for computer assistance. We also thank Niek Dekker and Yvonne Mul for critical reading of the manuscript. This work was supported in part by the Netherlands Foundation for Chemical Research (SON), with financial support from the Netherlands Organization for Scientific Research (NWO) and by the Netherlands Cancer Foundation (KWF) (M.A.).

References

Affolter,M., Percival-Smith,A., Muller,M., Leupin,W. and Gehring,W. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 4093–4097.

Aurora,R. and Herr,W. (1992) *Mol. Cell. Biol.*, **12**, 455–467.

Baumhueter,S., Mendel,D., Conley,P., Kuo,C., Turk,C., Graves,M.K., Edwards,C.A., Courtois,G. and Crabtree,G.R. (1990) *Genes Dev.*, **4**, 372–379.

Baumruker,T., Sturm,R. and Herr,W. (1988) *Genes Dev.*, **2**, 1400–1413.

Bodner,M., Castrillo,J.L., Theill,L.E., Deerinck,T., Ellisman,M. and Karin,M. (1988) *Cell*, **55**, 505–518.

Botfield,M.C., Jansco,A. and Weiss,M.A. (1992) *Biochemistry*, **31**, 5841–5848.

Chalepakis,G., Fritsch,R., Fickenscher,H., Deutch,U., Goulding,M. and Gruss,P. (1991) *Cell*, **66**, 873–884.

Clerc,R.G., Corcoran,L.M., LeBowitz,J.H., Baltimore,D. and Sharp,P.A. (1988) *Genes Dev.*, **2**, 1570–1581.

Desplan,C., Theis,J. and O'Farrell,P. (1988) *Cell*, **54**, 1081–1090.

Ekker,S.C., Young,R.E., Kessler,D.P. and Beachy,P.A. (1991) *EMBO J.*, **10**, 1179–1186.

Falkner,F.G. and Zachau,H.Z. (1984) *Nature*, **310**, 71–74.

Finney,M., Ruvkun,G. and Horvitz,H.R. (1988) *Cell*, **55**, 757–769.

Florence,B., Hardrow,R. and Laughon,A. (1991) *Mol. Cell. Biol.*, **11**, 3613–3623.

Frain,M., Swart,G., Monaci,P., Stampfli,S., Frank,R. and Cortese,R. (1989) *Cell*, **59**, 145–157.

Fried,M.G. and Crothers,D.M. (1984) *J. Mol. Biol.*, **172**, 263–282.

Freyd,G., Kim,S. and Horvitz,R. (1990) *Nature*, **344**, 876–879.

Harrison,S.C. (1991) *Nature*, **353**, 715–719.

Herr,W. et al. (1988) *Genes Dev.*, **2**, 1513–1516.

Ingraham,H.A. et al. (1988) *Cell*, **55**, 519–529.

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,P.F. and McKnight,S.L. (1989) *Annu. Rev. Biochem.*, **58**, 799–839.

Karin,M., Castrillo,J. and Theill,L. (1990) *Trends Genet.*, **6**, 92–96.

Karlsson,O., Thor,S., Norberg,T., Ohlsson,H. and Edlund,T. (1990) *Nature*, **344**, 879–882.
Kemler,I., Schreiber,E., Müller,M.M., Matthias,P. and Schaffner,W. (1989) *EMBO J.*, **8**, 2001–2008.
Kissinger,C.R., Liu,B., Martin-Blanco,E., Kornberg,T.B. and Pabo,C.O. (1990) *Cell*, **63**, 579–590.
Ko,H.-S., Fast,P., McBride,W. and Staudt,L.M. (1988) *Cell*, **55**, 135–144.
Kristie,T.M. and Sharp,P.A. (1990) *Genes Dev.*, **4**, 2383–2396.
Laughon,A. (1991) *Biochemistry*, **30**, 11357–11372.
LeBowitz,J.H., Clerc,R.G., Brenowitz,M. and Sharp,P.A. (1989) *Genes Dev.*, **3**, 1625–1638.
Mattaj,I.W., Lienhardt,S., Jiricny,J. and DeRobertis,E.M. (1985) *Nature*, **316**, 163–167.
Meijer,D., Graus,A., Kraay,R., Langeveld,A., Mulder,M.P. and Grosveld,G. (1990) *Nucleic Acids Res.*, **18**, 7357–7365.
Mendel,D.B., Khavari,P.A., Conley,P.B., Graves,P.B., Hansen,L.P., Admon,A. and Crabtree,G.R. (1991) *Science*, **254**, 1762–1763.
Mitchell,P.J. and Tjian,R. (1989) *Science*, **245**, 371–378.
Mul,Y.M., Verrijzer,C.P. and van der Vliet,P.C. (1990) *J. Virol.*, **64**, 5510–5518.
Müller,M., Affolter,M., Leupen,W., Otting,G., Wüthrich,K. and Gehring,W.J. (1988) *EMBO J.*, **7**, 4299–4304.
Müller-Immerglück,M.M., Ruppert,S., Schaffner,W. and Matthias,P. (1988) *Nature*, **336**, 544–551.
Nicosia,A., Monaci,P., Tomei,L., DeFransesco,R., Nuzzo,M., Stunnenberg,H. and Cortese,R. (1990) *Cell*, **61**, 1225–1236.
O'Hare,P. and Goding,C.R. (1988) *Cell*, **52**, 435–445.
Oliphant,A.R., Brandl,C.J. and Struhl,K. (1989) *Mol. Cell. Biol.*, **10**, 2944–2949.
Otting,G., Qian,Y.Q., Billeter,M., Müller,M., Affolter,M., Gehring,W.J. and Wüthrich,K. (1990) *EMBO J.*, **9**, 3085–3092.
Percival-Smith,A., Muller,M., Affolter,M. and Gehring,W.J. (1990) *EMBO J.*, **9**, 3967–3974.
Poellinger,L., Yoza,B.K. and Roeder,R.G. (1989) *Nature*, **337**, 573–576.
Pollock,R. and Treisman,R. (1990) *Nucleic Acids Res.*, **18**, 6197–6204.
Pruijn,G.J.M., van Miltenburg,R.T., Claessens,A.J. and van der Vliet,P.C. (1988) *J. Virol.*, **62**, 3092–3102.
Rosenfeld,M.G. (1991) *Genes Dev.*, **5**, 897–907.
Ruvkun,G. and Finney,M. (1991) *Cell*, **64**, 475–478.
Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual*, second edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Scheidereit,C., Cromlich,J.A., Gerster,T., Kawakami,K., Balmaceda,R.A., Currie,R.A. and Roeder,R.G. (1988) *Nature*, **336**, 551–557.
Schöler,H.R. (1991) *Trends Genet.*, **7**, 323–329.
Scott,M.P., Tamkun,J.W. and Hartzell,G.W. (1989) *Biochim. Biophys. Acta*, **989**, 25–48.
Smith,D.B. and Johnson,K.S., (1988) *Gene*, **67**, 31–40.
Staudt,L.M., Singh,H., Sen,R., Wirth,T., Sharp,P.A. and Baltimore,D. (1986) *Nature*, **323**, 640–643.
Sturm,R.A. and Herr,W. (1988) *Nature*, **336**, 601–604.
Sturm,R.A., Das,G. and Herr,W. (1988) *Genes Dev.*, **2**, 1582–1599.
Thiessen,H.-J. and Bach,C. (1990) *Nucleic Acids Res.*, **18**, 3203–3209.
Treisman,J., Harris,E. and Desplan,C. (1991) *Genes Dev.*, **5**, 594–604.
Tymon,A.M., Kües,U., Richardson,W.V.J. and Casselton,L.A. (1992) *EMBO J.*, **11**, 1805–1813.
Verrijzer,C.P., Kal,A.J. and van der Vliet,P.C. (1990a) *EMBO J.*, **9**, 1883–1888.
Verrijzer,C.P., Kal,A.J. and van der Vliet,P.C. (1990b) *Genes Dev.*, **4**, 1964–1974.
Verrijzer,C.P., van Oosterhout,J.A.W.M., van Weperen,W.W. and van der Vliet,P.C. (1991) *EMBO J.*, **10**, 3007–3014.
Verrijzer,C.P., van Oosterhout,J.A.W.M. and van der Vliet,P.C. (1992) *Mol. Cell. Biol.*, **12**, 542–551.
Wolberger,C., Verstow,A.K., Liu,B., Johnson,A.D. and Pabo,C.O. (1991) *Cell*, **67**, 517–528.

Received on May 4, 1992; revised on September 10, 1992