Nucleic Acids Research, 1995, Vol. 23, No. 16 3189–3197

Mutation of the Oct-1 POU-specific recognition helix leads to altered DNA binding and influences enhancement of adenovirus DNA replication

Hans C. van Leeuwen, Marijke J. Strating, Michel Cox+, Robert Kaptein[1] and Peter C. van der Vliet*

Laboratory for Physiological Chemistry, Utrecht University, Stratenum, 3508 TA Utrecht, The Netherlands and [1]Bijvoet Center for Biomolecular Research, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands

Received May 22, 1995; Revised and Accepted July 10, 1995

ABSTRACT

To assess which residues of Oct-1 POU-specific (POUs) are important for DNA recognition and stimulation of adenovirus DNA replication we have mutated 10 residues of the POUs helix–turn–helix motif implicated in DNA contact. Seven of these turned out to have reduced DNA binding affinity. Of these, three alanine substituted proteins were found to have a changed specificity using a binding site selection procedure. Mutation of the first residue in the recognition helix, Gln44, to alanine led to a loss of specificity for the first two bases, TA, of the wild-type recognition site TATGC(A/T)AAT. Instead of the A, a T was selected, suggesting a new contact and a novel specificity. A change in specificity was also observed for the T45A mutant, which could bind to TATAC(A/T)AAT, a site hardly recognized by the wild-type protein. Mutation of residue Arg49 led to a relaxed specificity for three consecutive bases, TGC. This residue, which is critical for high affinity binding, is absent from the structurally homologous lambdoid helix–turn–helix motifs. Employing a reconstituted system all but two mutants could stimulate adenovirus DNA replication upon saturation. Mutation of residues Gln27 and Arg49 impairs the ability of the Oct-1 POU domain protein to enhance replication, with a concomitant loss of DNA contacts. Since the POU domain binds the precursor terminal protein–DNA polymerase complex and guides it to the origin, lack of stimulation may be caused by incorrect targetting of the DNA polymerase due to loss of specificity.

INTRODUCTION

The POU domain is a DNA binding motif shared by a family of transcription factors involved in tissue-specific gene expression and cell differentiation. Besides transcriptional activation POU domain proteins are also involved in DNA replication (1–4). The isolated Oct-1 POU domain, which lacks the transactivation domain, enhances adenovirus DNA replication through an interaction with the adenovirus pTP–DNA polymerase complex (5). The POU domain consists of two separate DNA binding subdomains, a C-terminal 60 amino acid POU homeodomain (POUhd), closely related to the classical homeodomains (6,7), and an N-terminal 75 amino acid POU-specific domain (POUs), unique to the POU domain gene family. These subdomains are connected by a 14–29 amino acid flexible linker region which results in combined specificity of both subdomains and increased affinity (8,9). The intact Oct-1 POU domain recognizes the conserved sequence, TATGC($^{A}/_{T}$)AAT (10), as determined in a binding site selection procedure. The 5′ half of this site, TATGCA, is recognized by the POUs domain and the 3′-end, TAAT, by POUhd. The A/T overlap in the recognition sequence seems to be contacted by both subdomains (11,12).

Extensive structural studies of the isolated subdomains (13–16) and the intact POU domain in complex with DNA (12) have established the folding topology of the POU domain. The POUhd structure is a tri-α-helical structure with helices 2 and 3 forming a helix–turn–helix (HTH) recognition motif. POUs is a tetra-α-helical structure which also uses helices 2 and 3 as a HTH DNA recognition motif. The other helices (helix 1 in POUhd and helices 1 and 4 in POUs) stabilize the HTH motifs through extensive hydrophobic packing of the core. Most specific DNA contacts are made by helix 3 of both subdomains, which extend into the major groove.

The POUhd DNA binding helix differs from the POUs recognition helix in several respects. Upon binding in the major groove the POUhd recognition helix becomes extended by two α-helical turns at the C-terminus, allowing it to make extra contacts (12,16). Furthermore, the POUhd recognition helix contains a basic C-terminus typical of eukaryotic homeodomain proteins (17), which fixes the helix into the major groove. The POUs HTH element on the other hand is structurally similar to the DNA binding domain of bacteriophage λ and 434 proteins (13,14), with the exception of the turn between both helices, which is four amino acids longer in POUs. Besides this structural resemblance, the POUs domain also docks onto the DNA in a similar way. Analogous amino acids make identical side chain

* To whom correspondence should be addressed

+Deceased

base contacts (11), suggesting a shared DNA recognition mechanism between bacteriophage and eukaryotic HTH motifs.

In this study we have investigated the role of a number of residue side chains of Oct-1 POUs that have been implicated in making DNA contacts based on structural data (14). These solvent-exposed amino acids were exchanged for alanines and probed for reduced affinity. Moreover, relaxed or altered specificities of the mutant proteins Q44A, T45A, R49A and L55A were assayed using a random binding site selection procedure, allowing a deduction of each specific residue base contact. Comparison of the binding characteristics of POUs with the homologous lambdoid HTH motifs showed that there are essential differences. Furthermore, all substituted proteins were assayed for their ability to stimulate *in vitro* adenovirus replication.

MATERIALS AND METHODS

DNA constructs and site-directed mutagenesis

The wild-type His_6 Oct-1 POU domain expression construct was obtained by PCR amplification using Oct-1-specific primers that contained *Nde*I and *Bam*HI sites (5′ primer, GACCCGGCCCATATGGACCTTGAGGAGCTTG: 3′ primer, GGCCGGATCCTTAGTTGATTCTTTTTTCTTTC; restriction sites underlined). The His_6 POUHd construct was made using the same 3′ primer and a 5′ linker-specific primer containing a *Nde*I site (5′ primer, GAGCCCCCGCATATGAATTCTC CAGGAATTG). PCR products were cloned in the pET15b vector (Novagen) and verified by dideoxy DNA sequencing. For site-directed mutagenesis the POUs domain, including a part of the pET15b plasmid, was excised by digestion with *Xba*I and *Eco*RI and cloned in the pALTER™-1 vector (Promega). Site-directed mutagenesis was performed using the oligonucleotide-directed *in vitro* mutagenesis system as recommended by the manufacturer (Promega) and verified by sequence analysis.

Expression and purification of wild-type and mutant POU proteins

Oct-1 POU constructs in pET15b were expressed in *Escherichia coli* BL21(DE3) plysS cells using the T7 expression system (pET15b) (18). The His-tagged fusion proteins were purified on Ni–NTA columns essentially as described by the manufacturer (Qiagen) with the following modifications. Overnight cultures were 1:100 diluted in LB medium (50 μg/ml ampicillin, 30 μg/ml chloramphenicol) to a volume of 0.5 l and grown at 37°C until an OD_{600} of 0.7 was reached. Protein expression was induced by addition of isopropyl thiogalactopyranoside to 1 mM. Four hours after induction cells were harvested (3000 *g*, 10 min, 4°C) and resuspended in 10 ml cold lysis buffer [50 mM sodium phosphate buffer, pH 8.0, 5 mM β-mercaptoethanol, 5 mM sodium metabisulfite (NaMBS), 0.1 mM phenylmethylsulphonyl fluoride (PMSF)]. After a freeze/thaw step (overnight –20°C) 0.5 ml 10 mg/ml lysozyme was added and incubated for 60 min at 4°C, followed by sonification (30 s, 300 W). Dnase I and Rnase A were added to a final concentration of 5 μg/ml and 1 μg/ml, respectively, and the mixture was incubated for 30 min at 4°C on a rotating wheel. NP-40 was added to final concentration of 0.1% and NaCl to a final concentration of 300 mM. Cell fragments and insoluble material were removed by centrifuging twice (10 min at 10 000 *g*, 20 min at 40 000 *g*). Extracts were loaded on a 5 ml DEAE–Sephacel column. The flow-through was loaded on a 1.5 ml Ni–NTA column (Qiagen). The column was washed with 10 ml buffer A (50 mM sodium phosphate buffer pH 6.0, 300 mM NaCl, 5 mM β-mercaptoethanol, 5 mM NaMBS, 0.1 mM PMSF, 10% glycerol) followed by 10 ml buffer A containing 15 mM imidazole. The His-tagged proteins eluted at ~200 mM imidazole employing a 25 ml linear gradient ranging from 15 to 500 mM imidazole. Fractions containing the His-tagged proteins were pooled and diluted to 100 mM NaCl with buffer B (50 mM Tris pH 8.0, 10% glycerol, 5 mM β-mercaptoethanol, 0.5 mM DTT, 0.1 mM PMSF) and loaded onto a 1.5 ml Sepharose Fast Flow S column. The column was washed with buffer B, 50 mM NaCl and eluted with a 30 ml linear gradient of 50–1000 mM NaCl. Wild-type Oct-1 POU domain and the mutants all eluted at ~400 mM NaCl. Protein concentrations were determined using the BioRad dye agent with bovine serum albumin as standard. All proteins were completely homogeneous and did not contain any breakdown products, as shown with Coomassie staining of a SDS–polyacrylamide gel.

DNA binding studies

The probes used for band shift assays were double-stranded oligonucleotides containing the adenovirus 2 (Ad2) or adenovirus 4 (Ad4) binding site, end-labelled with T4 polynucleotide kinase and purified by preparative polyacrylamide gel electrophoresis. The sequences are: Ad4, CGAATATGCAAATAAGGC; Ad2, AGGCCAATATGATAATGAGGGGGT. Sites selected from the binding site selection procedure were digested with *Eco*RI and *Xba*I and end-labelled with [α-^{32}P]dATP using the Klenow fragment. The DNA concentrations were determined by absorption at 260 nm. The concentration of input DNA was 1 nM. Binding reactions were carried out for 60 min on ice in 20 μl binding buffer (20 mM HEPES–KOH, pH 7.5, 1 mM EDTA, 1 mM DTT, 0.025% Nonidet P-40, 4% Ficoll). Free DNA and protein–DNA complexes were separated on a 15% polyacrylamide gel (37.5:1) run in 0.5× TBE at 4°C. Dried gels were exposed and quantified by liquid scintillation counting of gel slices corresponding to free and bound DNA. The equilibrium dissociation constant (K_d) was calculated at half saturation from the equation $K_d = P_t - D_b$ (19). The total protein concentration (P_t) was calculated using a deduced molecular mass of 19.7 kDa for all His_6–POU proteins and 10.8 kDa for His_6-POUHd.

Binding site selection

The random site used in the first selection round (5′-CTCGCACCGAATTCCTCGAAATTNNNNNNNNCGACAGGATCCGCTGAACTGACCTGACC-3′) was rendered double-stranded using radiolabelled primer A (5′-GGTCAGGTCAGTTCAGCG-3′), Klenow fragment and dNTPs (20). Binding reactions were performed as described above with 1 μg poly(dI·dC), 0.5 ng probe and 5 ng purified POU domain. Protein–DNA complexes were separated on a 10% polyacrylamide gel (37.5:1) run in 0.5× TBE at 4°C. Following exposure of the wet gel, the bound DNA fragments were cut out and eluted. After ethanol precipitation ~5% of the selected DNA was amplified in 15 PCR cycles with 150 ng primer A and primer B (5′-CTCGCACCGAATTCCTC-3′). An aliquot (0.5 ng) of this amplified material was used in the next round of selection. After eight rounds of selection bound DNA fragments were eluted, digested with *Eco*RI and *Bam*HI and cloned into the vector pUC18, followed by sequencing of the insert with T7 sequenase (USB).

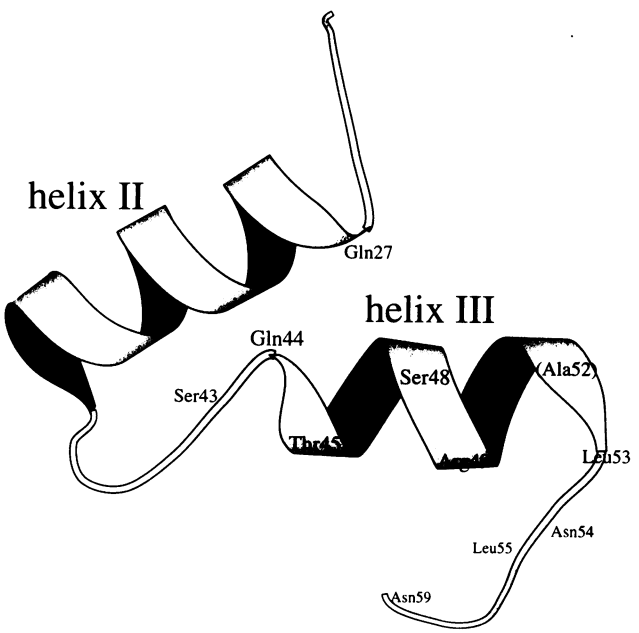

Figure 1. Ribbon presentation of the Oct-1 POUs helix–turn–helix motif using the program Molscript (38). Amino acids that were mutated are indicated in three letter code followed by numbering according to Klemm *et al.* (12). Ala52 placed between brackets has not been mutated. The helix 3 axis is placed horizontally and faces the major groove.

Ad DNA replication *in vitro*

Replication reactions (21) were performed in a total volume of 15 µl containing 30 ng *Xho*I-digested adenovirus 5 (Ad 5) TP–DNA, 25 mM HEPES, pH 7.5, 50 mM NaCl, 1.5 mM $MgCl_2$, 1 mM DTT, 40 µM dATP, dTTP, dGTP and 500 nM [α-^{32}P]dCTP (400 Ci/mmol), 15 ng precursor terminal protein–polymerase complex, 0.5 µg DNA binding protein, 2 binding units POU protein. All replication proteins were purified as described before (5). Reactions were incubated for 1 h at 37°C and stopped by addition of 1.5 µl stop mix (40% sucrose, 1% SDS, 0.1% bromophenol blue, 0.1% xylene cyanol). Replication products were analysed on a 1% agarose gel (0.5× TBE, 0.1% SDS) run at 50 V for 16 h. After electrophoresis gels were dried and exposed. One binding unit POU is defined as the amount of protein that could bind 50% of the Ad2 probe in a band shift assay (1 ng for wild-type POU domain).

RESULTS

DNA binding affinity of POUs mutants

Mutated residues present at the protein–DNA interface of the HTH motif of Oct-1 POUs are indicated in Figure 1. All residues were mutated to alanine except for residue Gln27, which was changed to Glu, because we did not want to disturb the hydrogen bond between the side chains of Gln27 and Gln44. This hydrogen bond is present both in the crystal structure of the Oct-1 POU–DNA complex (12) and in the highly related λ repressor HTH motif (22).

Wild type and mutant constructs were cloned into the pET15b expression vector (Novagen) containing at the N-terminus six consecutive histidines. Bacterially expressed POU domains were purified to homogeneity by successive application of anion exchange, nickel affinity and cation exchange chromatography,

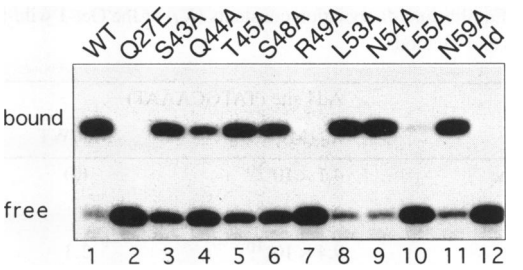

Figure 2. Effects of amino acid substitution in the Oct-1 POUs domain on binding to the Ad4 octamer sequence (TATGCAAAT). Aliquots (1 ng) of each protein were tested in the band shift assay as described in Materials and Methods. WT, wild-type; Hd, POU homeodomain.

as described in Materials and Methods. The various proteins were tested for DNA binding in a band shift assay with a probe containing the Oct-1 recognition site (TATGCAAAT), which is the optimal binding site present in the Ad4 origin. As shown in Figure 2, two of the mutant POUs domains, Q27E and R49A, did not bind detectably (lanes 2 and 7), even at longer exposures (not shown). Mutants Q44A and L55A showed strongly reduced affinity (lanes 4 and 10). Substitutions at positions 43, 45 and 48 in helix 3 resulted in intermediate affinities (lanes 3, 5 and 6). No drastic effects were seen when the side chains of Leu53, Asn54 and Asn59, located outside helix 3, were changed to a methyl group. As for the latter two, this was unexpected, since these residues have been predicted to make DNA contacts on the basis of observed line broadening and chemical shift changes in HMQC experiments (14).

In order to obtain a quantitative comparison of the mutants we determined the POU domain concentration required for half saturation of the Ad4 site ($T^{-1}A^1T^2G^3C^4A^5A^6A^7T^8$), from which the equilibrium dissociation constant (K_d) was calculated (see Materials and Methods). The results are summarized in Table 1. The mutant POU proteins Q27E and R49A, for which no binding could be observed in Figure 2, have K_d values 71- and 150-fold higher respectively than the wild-type POU domain. We also determined the equilibrium constants of the various mutant proteins for the naturally occurring Oct-1 binding site of the Ad2 origin, TATGA-TAAT, which is used for replication studies. The affinity of the intact POU domain for the Ad2 site is lower than for the Ad4 site, because the A at position 4 in the Ad2 site (underlined) is unfavourable for POUs binding (23). The Ad2 site, in contrast, is more favourable for POUhd binding than the Ad4 site, because the 3′ part of the Ad2 site, TAAT, is the preferred site for isolated POUhd. As a result, the Ad2 Oct-1 site is biased toward POUhd binding, whereas the Ad4 site is biased toward POUs binding. Thus deletion of the complete POUs domain leads to a 340-fold lower affinity on the Ad4 site, compared with only a 9.1-fold lower affinity on the Ad2 site (Table 1). In agreement with this, mutations with grossly diminished affinities for the Ad4 site (Q27E and R49A) have less reduced affinities for the Ad2 site. Interestingly, the single mutation of residue Gln27 in POUs has an even more drastic effect on binding to the Ad2 site (12-fold) than complete deletion of the POUs subdomain (9.1-fold). This is not the case on the Ad4 site, where the Q27E mutant is less affected than POUhd (71-fold compared with 340-fold). The reduced affinity due to this mutation, together with the unfavourable adenine at position 4 of the Ad2 site, has thus resulted in a total loss of POUs binding on the Ad2 site. Apparently, homeodomain binding of this mutant is hampered by the presence of an ineffective POUs domain.

Table 1. Equilibrium dissociation constants (K_d) of the Oct-1 wild-type and mutated POU domains for the Ad4 and Ad2 sites, calculated as described in Materials and Methods

	Ad4 site (TATGCAAAT)		Ad2 site (TATGATAAT)		
	K_d (M)	Mut/WT	K_d (M)	Mut/WT	Ratio Ad2:Ad4
Wild type	4.1×10^{-10}	1.0	1.4×10^{-9}	1.0	3.4
Q27E	2.9×10^{-8}	71	1.7×10^{-8}	12	0.6
S43A	9.4×10^{-10}	2.3	4.3×10^{-9}	3.1	4.6
Q44A	1.4×10^{-9}	3.3	6.6×10^{-9}	4.7	4.7
T45A	9.8×10^{-10}	2.4	4.6×10^{-9}	3.3	4.7
S48A	1.2×10^{-9}	2.9	4.3×10^{-9}	3.0	3.6
R49A	6.1×10^{-8}	150	9.4×10^{-9}	6.7	0.2
L53A	4.9×10^{-10}	1.2	1.9×10^{-9}	1.4	3.9
N54A	4.7×10^{-10}	1.1	1.8×10^{-9}	1.3	3.8
L55A	1.9×10^{-9}	4.6	3.5×10^{-9}	3.2	1.8
N59A	5.9×10^{-10}	1.5	2.1×10^{-9}	1.6	3.6
Homeodomain	1.4×10^{-7}	340	1.2×10^{-8}	9.1	0.1

Experiments were performed in triplicate and deviation was never >15%. Mutants are indicated in the single letter code followed by its position and the identity of the mutation. Mut/WT indicates binding affinity relative to the wild-type POU protein on the indicated site. In the last column the ratio of the binding affinities on the two binding sites is indicated.

Alanine substituted proteins have a changed specificity

Each of the alanine substitutions might lead to a change in specificity of DNA recognition due to loss of a contact or gain of a new contact. To determine this we employed a binding site selection procedure for several of the mutants. We used a pool of oligonucleotides in which part of the POUs recognition site was randomized and the homeodomain recognition sequence was kept intact, i.e. NNNNNNNNAAT. Selection was performed for the wild-type and mutant proteins which showed strongly reduced DNA binding: Q44A, T45A, R49A and L55A. Mutant Q27E was not chosen, since in both the λ repressor–DNA complex (24) and Oct-1 POU–DNA complex (12) this residue makes phosphate backbone contacts. Proteins were incubated with the labelled pool of oligonucleotides. After separation on a polyacrylamide gel bound DNA fragments were eluted, PCR amplified and used in a subsequent selection round (see Materials and Methods). After eight rounds of enrichment selected sites were cloned and sequenced. Figure 3A shows the individual sites selected by the wild-type and the four mutants and the frequency of bases at each position is shown in Figure 3B.

Most of the sequences selected by the wild-type POU domain correspond to the optimal octamer binding site [TATGC (A/T)AAT], except for position 4, where G, T or C were selected. Binding studies with the individual sites revealed that the variation in affinities of the POU domain for the different sites are <2-fold. The highest affinity was obtained with a C at position 4. Only site wt.43, TATTCTAAT, with a T selected at position 3, is bound 2-fold more weakly.

For the Q44A mutant another pattern was observed. No T was selected at position –1, but a random distribution was found. At position 1 a T is selected instead of an A in 18 out of 19 cases [consensus NTTGC(A/T)AAT]. This indicates that a contact between Q44 and the A1 residue exists which is lost upon mutation. Moreover, a new contact between the alanine and the predominantly selected T residue seems to have been gained. This is confirmed by a quantitative determination of the K_d employing one of the Q44A selected sites, GTTGCTAAT (44.03), with a T at position 1. The wild-type POU domain binds with a 3-fold lower affinity to this site relative to the canonical octamer site (Fig. 4A, lanes 1 and 3), whereas Q44A has a 3-fold higher affinity for this site, thus restoring binding almost to wild-type levels (lane 4).

The consensus site of the T45A mutant differs from the wild-type consensus site mainly at position 3 (Fig. 3B), where a purine was preferentially selected. Comparison of site TATACAAAT (45.42) with TATGCAAAT (wt.38) shows reduced binding for the wild-type POU domain when an adenine is present, whereas T45A is indifferent to an A or G at this position (Fig. 4B). A cytosine or a thymine at position 3 reduces binding for both the wild-type and mutant T45A (data not shown). The preference for an A by mutant T45A suggests an interaction between alanine and the T methyl on the opposite strand (see Discussion).

Specificity is lost or changed at positions 2–4 in the sites selected by mutant R49A, TAANG(A/T)AAT (Fig. 3B). This may explain the 150-fold lower affinity of this mutant for the Ad4 site. The preference for a G residue at position 4 presumably originates from the homeodomain, since the isolated homeodomain prefers GTAAT (23). Variations at positions 2 and 3 in sites TAACGTAAT (49.45), TATCGTAAT (49.09) and TATGGTAAT (49.22) were tested for R49A binding and compared with the isolated homeodomain (Fig. 5). Binding to site CTTGGTAAT (44.13) with different bases at positions –1 and 1 was also tested. The highest affinity site for the R49A mutant is TATGGTAAT with G at position 3, which is also selected by the wild-type protein (Fig. 5, lane 2). The affinity for this site, however, is only slightly higher than for sites containing a non-consensus base pair at position 3 (Fig. 5, lane 3), explaining why it is not preferentially selected (Fig. 3B). Variation of position 2 in the sites TAACGTAAT and TATCGTAAT (Fig. 5, lanes 3 and 4) does not result in different affinities of mutant R49A.

wt (n=18):

	-3	-2	-1	1	2	3	4	5
G	11	28	0	0	0	95	32	0
A	22	22	6	100	0	0	0	24
T	56	22	89	0	100	5	28	76
C	6	28	6	0	0	0	39	0
	t	-	**T**	**A**	**T**	**G**	-	**a** **T**

R49A (n=19):

	-3	-2	-1	1	2	3	4	5
G	32	16	5	11	5	21	68	0
A	11	11	16	74	58	5	0	42
T	16	26	74	0	32	26	16	58
C	42	47	5	16	5	47	16	0
	-	-	**T**	**A**	a	-	G	**a** **t**

Q44A (n=19):

	-3	-2	-1	1	2	3	4	5
G	31	11	37	0	0	84	11	0
A	37	47	11	0	0	0	0	26
T	11	32	26	95	100	16	16	74
C	21	11	26	5	0	0	74	0
	-	-	-	T	**T**	**G**	**C**	**a** **T**

L55A (n=13):

	-3	-2	-1	1	2	3	4	5
G	46	31	8	8	15	54	8	8
A	15	23	8	69	0	8	8	31
T	31	31	69	8	69	15	15	54
C	8	15	15	15	15	23	69	8
	-	-	**T**	**A**	**T**	**g**	**C**	**a** **t**

T45A (n=15):

	-3	-2	-1	1	2	3	4	5
G	27	20	20	0	0	47	7	0
A	20	27	0	100	0	33	7	47
T	33	27	53	0	93	0	13	53
C	20	27	27	0	7	20	73	0
	-	-	**t**	**A**	**T**	Pu	**C**	**a** **t**

A

Wild type selected:

```
      -3-2-1 1 2 3 4 5
01:   TTTATGGT-AATT...
04:   TCTATGTT-AATT...
07:   CATATGGT-AATT
09:   AACATGCA-AATT.
21:   AGTATGGT-AATT
26:   AGAATGCT-AATT
38:   TGTATGCA-AATT.
42:   TGTATGTT-AATT
43:   GATATTCT-AATT
45:   GCTATGCA-AATT
```

Arg49 ⇒ Ala 49 selected:

```
      -3-2-1 1 2 3 4 5
02:   GTTCATTT-AATT
03:   CCTAACGA-AATT...
09:   CATATCGT-AATT
11:   TGCGAGGT-AATT
12:   ACTAACTT-AATT
22:   GTTATGGT-AATT..
24:   TCTAACGA-AATT
27:   TGACATCA-AATT
31:   ACTATTGA-AATT
32:   GTGGGTCT-AATT
36:   CCACCCCT-AATT
37:   CAAATTGA-AATT
45:   GCTAACGT-AATT
46:   CGTAAATT-AATT
```

Qln44 ⇒ Ala 44 selected:

```
      -3-2-1 1 2 3 4 5
01:   AAATTGCT-AATT
02:   GACTTTCT-AATT.
03:   GTGTTGCT-AATT...
04: TAATTTGCA-ATT  .
13:   CGCTTGGT-AATT
14:   CCTCTGTT-AATT
21: GACGTTGGA-ATT
32:   AAGTTGCA-AATT
35:   AAGTTGCT-AATT
36:   TTTTTGCT-AATT
37:   AAATTGTA-AATT
41:   CTCTTTCT-AATT
42:   CACTTGTT-AATT
44:   TGTTTGCT-AATT
```

Leu55 ⇒ Ala 55 selected:

```
      -3-2-1 1 2 3 4 5
01:   TACATGCT-AATT
03:   ATTATTCT-AATT
06:   GTTATGCA-AATT.
07:   GGTATGCT-AATT
09:   TAACTGAT-AATT
11:   GGTTGCGC-AATT
13:   TATATGCT-AATT
25:   ACCCCCCA-AATT
33:   CGTAGGTT-AATT
35:   GTTATCCG-AATT
41:   TCGGCATA-AATT
42:   GGTATTCT-AATT
```

Thr45 ⇒ Ala 45 selected:

```
      -3-2-1 1 2 3 4 5
01: TCCTATGTA-ATT
02:   TGCATACT-AATT
04:   GTGATGCT-AATT.
05:   GTGATCAA-TAATT
07:   AACATGCA-AATT.
23:   TCTACCGT-AATT
24:   TCTATGTA-AATT
42:   CATATACA-AATT.
43:   TCCATCCT-AATT
44:   AGTATACT-AATT
51:   GTTATACT-AATT
52:   TGTATGCT-AATT
```

B

Figure 3. Results of binding site selection for the Oct-1 POU domain and POU mutants. Bold letters indicate selected bases at corresponding positions of the octamer site [TATG(A/T)AAT]. (**A**) Aligned oligonucleotides. Sites were selected as described in Materials and Methods. •, identical sequence independently selected. (**B**) Relative frequency (%) of the bases at each position. Upper case indicates high constraint (≥65%). Lower case bases indicates moderate constraint (50–65%). – no strong selection. Pu, purine.

In most sites selected by R49A the consensus bases are selected at positions –1 (T) and 1 (A). When C and T bases are present at these positions, CTTGGTAAT (44.13), instead of the consensus

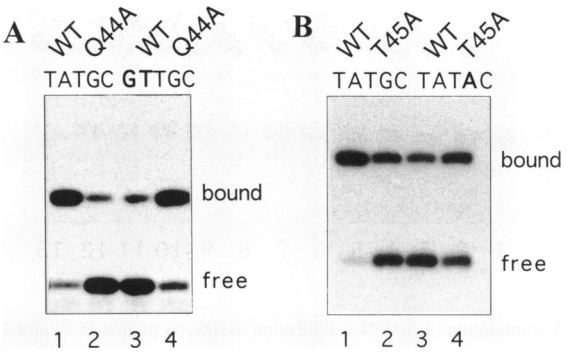

Figure 4. Mutants Q44A and T45A binding specificities different from the wild-type. (**A**) Wild-type and mutant Q44A binding was assayed in a band shift experiment using two probes, a wild-type selected site, TATGCTAAT (lanes 1 and 2), and a mutant selected site, gtTGCTAAT (lanes 3 and 4). (**B**) Wild-type and mutant T45A binding was assayed in a band shift experiment using two probes, a wild-type selected site, TATGCAAAT (lanes 1 and 2), and a mutant selected site, TATACAAAT (lanes 3 and 4).

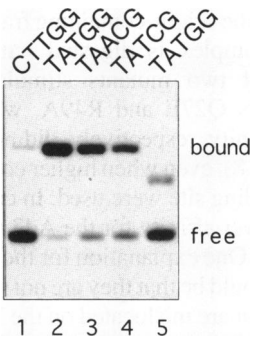

Figure 5. Mutant R49A still contacts the first two bases in the recognition site. Binding of mutant R49A (10 ng) to four different sites (lanes 1–4) compared with homeodomain binding (lane 5). Sites tested were 44.13 (lane 1), 49.22 (lanes 2 and 5), 49.45 (lane 3) and 49.09 (lane 4).

TA, TATGGTAAT (49.22), POUs binding in mutant R49A is completely lost and its affinity is even lower than homeodomain binding (Fig. 5, lanes 1, 2 and 5). This shows that these first two bases are essential to anchor the POUs domain of the R49A mutant protein.

Finally, substitution of the invariant Leu55 to alanine, which results in a 4.6-fold lower affinity for the Ad4 site (Table 1), does not result in a preference different from the canonical octamer site. Indeed, of all sites tested mutant L55A had the highest affinity for the site TATGCAAAT. The lower affinity therefore might be caused by a lost non-specific DNA contact or by an overall destabilizing effect on HTH conformation as a result of this substitution (see Discussion).

Mutants Q27E and R49A do not enhance *in vitro* adenovirus replication

Earlier experiments showed that the presence of the POUs domain is required for the POU domain to stimulate *in vitro* adenovirus DNA replication (19,25). We therefore assayed the various mutants for their ability to stimulate adenovirus type 5 DNA replication in a reconstituted system (Fig. 6). The Ad5

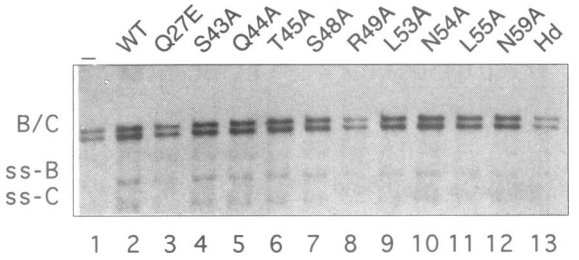

Figure 6. Stimulation of Ad DNA replication *in vitro* by mutant POU proteins. Replication was performed employing a reconstituted system (see Materials and Methods) in the absence (lane 1) or presence of two binding units of each purified protein. The origin-containing fragments B and C, as well as the single-stranded B and C fragments (ss-B and ss-C) originating from a second round of replication, are indicated.

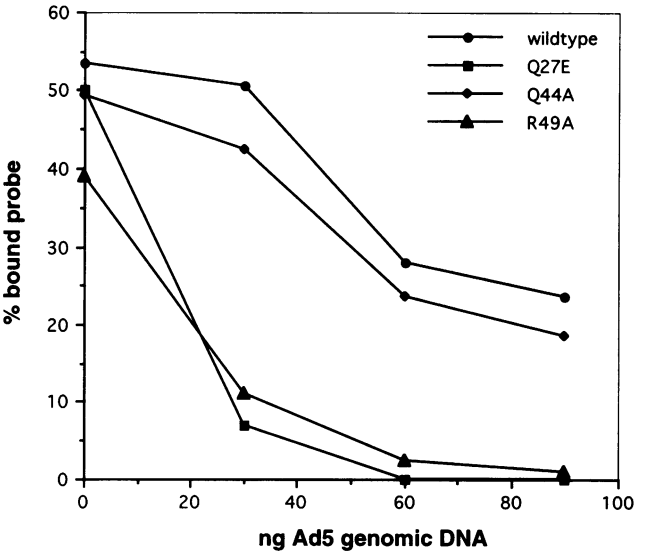

Figure 7. Wild-type Oct-1 POU protein and mutant protein Q44A bind with high specificity to the Ad2 origin recognition site, whereas mutant proteins Q27E and R49A bind non-specifically to the complete adenovirus genome. One binding unit of wild-type and mutant protein (see Materials and Methods) was incubated with 0.25 ng Ad2 labeled probe and increasing amounts of Ad5 genomic DNA. After 1 h bound and free DNA were separated on a native polyacrylamide gel. After quantification the percentage of bound Ad2 probe was plotted against the total amount of Ad5 DNA in the reaction. ●, wild-type protein; ■, mutant protein Q27E; ◆, mutant protein Q44A; ▲, mutant protein R49A.

origin contains an Oct-1 binding site identical to the Ad2 site described above. For comparison of the stimulation by each mutant we corrected for differences in DNA binding affinity and added equal binding units (see Materials and Methods). Fragments B and C are the origin-containing fragments obtained when Ad5 DNA–TP complex is digested with *Xho*I. Under these conditions all but two mutants stimulated similarly to the wild-type. Mutants Q27E and R49A, which showed 12- and 6.7-fold lower affinity, respectively, did not enhance replication (Fig. 6, lanes 3 and 8), even when higher concentrations sufficient to saturate the binding site were used. In contrast, mutant Q44A, with a 4.7-fold lower affinity for the Ad2 site, reaches wild-type stimulation levels. One explanation for the lack of stimulation by Q27E and R49A could be that they are not properly targeted to the origin sequence, but are mislocated on the 36 000 bp genome due to their lower specificity, whereas mutant Q44A retains sufficient specificity. To test this a labelled oligonucleotide (20mer) containing the Ad2 site was challenged with the intact 36 kbp Ad5 genome (Fig. 7). Whereas the wild-type and Q44A showed the profile expected for two POU binding sites per genome (one at each molecular end), Q27E and R49A were competed out very rapidly, indicating additional binding sites recognized by the homeodomain.

DISCUSSION

Positioning of a HTH recognition motif onto the DNA depends on interactions of amino acid side chain groups with the sugar–phosphate backbone. The primary determinants for specificity are interactions between side chains and bases. To assess which residues of the Oct-1 POUs domain are involved in these specific DNA contacts we exchanged several amino acids in the recognition motif for alanine, thereby possibly eliminating the ability to make base contacts. Selection of the optimal binding sites for these mutants out of a pool of random sites enables determination of lost or changed specificity due to the shortened side chain. Such a search also allows detection of multiple sequence changes, in case more than one base is contacted.

Contacts deduced from selection data compared with contacts in the co-crystal of POU and related DNA binding domains

Selection data for Q44A are consistent with the co-crystal structures of λ repressor, 434 repressor and Oct-1 POU domain, in which bidentate hydrogen bonding links Gln44 to an adenine at position 1 (24,26). The new specificity shown by Q44A is analogous to the Q28A mutation in the 434 repressor, which favours thymine at position 1 due to hydrophobic interactions between alanine and the T1 methyl (27). Such a contact might also exist in the POUs Q44A mutant. Employing the coordinates of the crystal structure combined with an exchange of Gln for Ala and the A:T base pair for T:A results in a distance of 3.56 between the alanine and the T1 methyl (see Fig. 8). Also likely to be involved in these hydrophobic interactions is the side chain methylene group of Ser48, which is even closer to the T1 methyl group (2.14 Å; see Fig. 8). This Ser48 residue is absolutely conserved in both the POUs class HTH motif and in the GalR/LacI HTH family (28), but not in the lambdoid family of HTH motifs. In the LacI protein–DNA complex the corresponding serine methylene group indeed contacts a T (Fig. 9). This suggests that the POUs domain has conflicting determinants for sequence-specific recognition, with the Gln44 having a stronger effect on sequence preference than Ser48. Identical alanine substitutions at position 44 were made in the Oct-2 (11) and Brn-2 (29) POUs domains. Binding studies with these proteins, however, show equal affinities for wild-type and mutant proteins to the site TTTGCAAAT, which seems to contradict our data. This paradox can be explained by the presence of a thymine methyl at position –1 (underlined) which is in van der Waals contact with the Cγ and Cδ of Gln44 (12,24). These contacts are lost in the Ala44 mutant, as can be seen from binding site selection data, where no T at position –1 is selected (see Fig. 3B). The gain of an interaction between the newly selected T1 methyl and mutant Ala44 is accompanied by the loss of contact with T at –1.

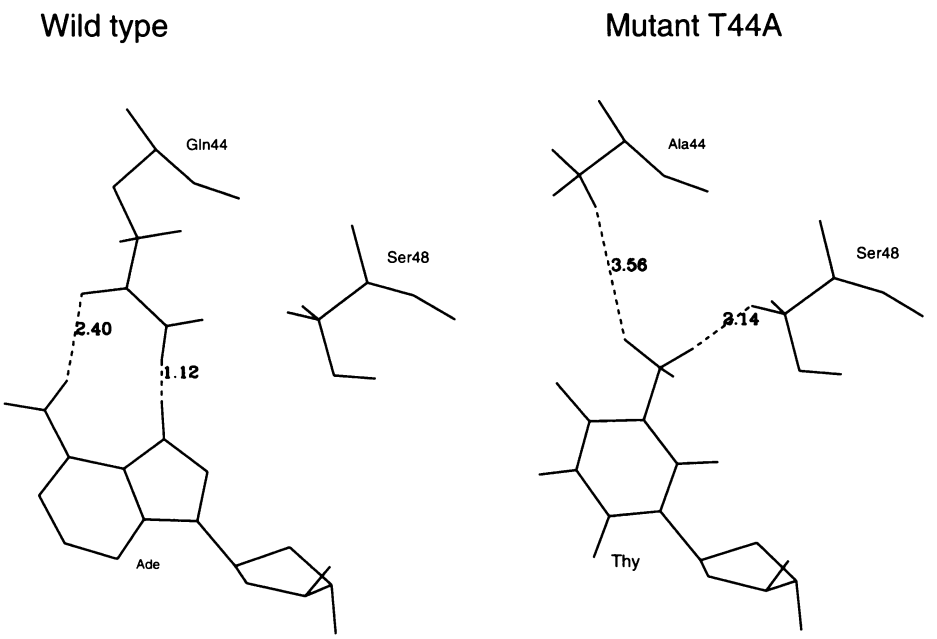

Figure 8. Base contact made in the wild-type situation and in the Q44A mutant at position 1 of the recognition sequence. Figures were generated and distances measured with the program InsightII (Biosym Technology Inc). Coordinates are from the Oct-1 POU–DNA crystal stucture (17). For the mutation Gln44 was exhanged for alanine and the A:T base pair at postion 1 for T:A in the wild-type Oct-1 POU–DNA complex. See text for further details.

The carboxyamides of Gln44 and Gln27 are linked by a hydrogen bond, but this interaction is probably not essential for Gln27–phosphate contact, since the Q44A mutant, which cannot hydrogen bond to Gln27, has only a 3-fold reduction in DNA binding, whereas the Q27E mutant has a 70-fold reduced affinity (Table 1). This severe reduction is probably caused by the negative charge effect of the Q27E mutant disrupting both side chain–side chain and side chain–backbone contacts (12).

The consensus binding site of the T45A mutant is $T^{-1}A^{1}T^{2}(G/A)^{3}C^{4}(A/T)^{5}AAT$ (Fig. 3B). The selection of an A:T base pair at position 3 suggests that the T′3 methyl group is in close contact with Ala45. Model building indeed results in a distance of 2 Å (data not shown). A mutated λ repressor with a corresponding alanine substitution binds more tightly to a site with an A:T base pair at position 3 (30). Affinity for the G:C at position 3 cannot be explained by a contact with Ala45 (31). Instead, this base pair is probably selected by Arg49 contacting G3, as has been shown in the crystal structure (see Fig. 9). Another scenario explaining the selection of either a G or A at position 3 by T45A is a complete loss of contact by this residue combined with a remaining preference for a purine by Arg49, which hydrogen bonds A3 or G3 bases via N7. In the crystal structure Thr45 also contacts T2 via O4. Thus one would expect that in the T45A mutant such a contact is lost and a T would no longer be selected. Base T2, however, is still selected in the T45A mutant and this must be due by contacts from other residues within the mutant protein or by the alanine itself. In the model where T45 is exchanged for alanine the minimal distance between the alanine and T2 methyl is 3.1 . Also within the range of van der Waals contact is the Cγ of Arg49 (2.55 Å).

The R49A mutant lacks the ability to discriminate base pairs at positions 2–4 (Fig. 5). In the co-crystal Arg49 makes only two contacts ($G^{3}C^{4}$; see Fig. 9). The POU protein–DNA complex (12) shows that the long side chain of Arg49, which protrudes into the major groove to make contact, is stabilized via a complex of van der Waals interactions with the side chains of residues Thr45, Asn54 and Leu55. It could be possible that Thr45, which is involved in selecting T2 (see above), is unable to select this thymine in the absence of the stabilizing arginine side chain. Also possible is that Arg49 is indeed involved in determining the T2 contact (see above).

L55A showed a reduction in DNA binding, but no changed sequence preference. This lower affinity is apparently due to a decrease in stability of the protein in the complex. The effect of mutating Leu55, therefore, may be an indirect effect on the Arg49 side chain. A similar guiding of the arginine side chain occurs in the Lac repressor (Arg22; see Fig. 9), which is stabilized via a hydrogen bond with Gln26 (6). The involvement of residue Asn54 in guiding the essential arginine might explain the chemical shift change of the NH_2 group of Asn54 upon complex formation, even though the residue itself is not in close proximity to the DNA (14).

In conclusion, the site selection procedure employed here, coupled to quantitation of binding affinities, reveals base contacts that agree to a large extent with contacts seen in the co-crystal structure and suggests additional contacts. As shown here, a loss of specificity does not in all cases mean a loss of a direct base contact, but can also be caused by an indirect effect via a loss of stabilizing amino acid side chain interactions (e.g. the T2 contact in the R49A mutant; see above). Moreover, loss of a direct base contact does not always influence specificity when other amino acid side chains contact the same base pair.

POUs: a combination of prokaryotic and lambdoid HTH motifs?

Although the Oct-1 POUs tetra-α-helical structure is very similar to the DNA binding domain of the λ and 434 repressors and cro, and an analogous contact is made by the two conserved glutamine residues at the beginning of helices 2 and 3 with A1 in the recognition sequence (11,32), other contacts made by the POUs

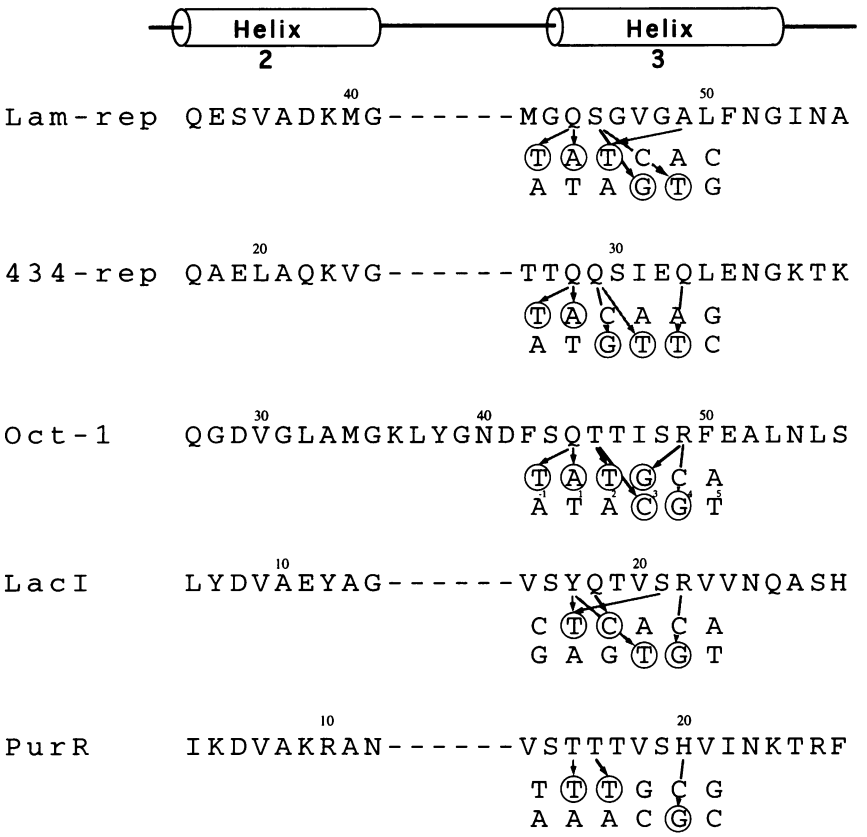

Figure 9. Comparison of the HTH motif and base contacts in the recognition helix of POUs, λ repressor (24), 434 repressor (26), LacI (6) and PurR (39). Specific contacts are indicated by black arrows. Apolar T methyl contacts are indicated with lighter shaded arrows. Numbering of the recognition helix starts from the first residue. Original numbering indicated above the sequence.

recognition helix map better to the prokaryotic recognition helix of the Lac repressor, belonging to the Lac/GalR family. The contact at position 4 in POUs is mediated via a bidentate contact to G'4. An identical arginine–guanine contact essential for DNA binding is made by the recognition helix of the prokaryotic Lac repressor (Fig. 9). Furthermore, comparison of the POUs and LacI complexes reveals that identical phosphate backbone contacts are made by corresponding serine and threonine residues in these divergent species (6,12). This suggests that the eukaryotic POUs HTH motif is a hybrid between the bacteriophage and prokaryotic HTH motifs, possibly arising from horizontal transfer (33). Alternatively, the DNA recognition motifs, as presented here, have developed independently with identical favourable side chain–base/backbone contacts.

An 'accidental' recombination/translocation event might have taken place in linking an ancestral independent POUs-like domain to a homeodomain, resulting in a protein with combined affinity and specificity.

Loss of specificity leads to a concomitant loss of replication enhancement

The function of the POUs domain seems to be to increase the specificity of the intact POU domain. A direct interaction between the POUhd and adenovirus pTP–DNA polymerase complex was shown to account, at least partially, for stimulation of DNA replication (5). Direction of these proteins to their target site is highly dependent on the specificity of the intact POU domain. The covalent linkage of the POUs domain with the homeodomain increases the specificity ~200-fold (23). Residue Arg49 is involved in contacting three bases in the POUs recognition site, therefore the loss of specificity due to mutation to alanine explains why mutant R49A does not enhance replication. Similar accurate targetting of proteins by the POU domain also occurs in co-recruitment of the herpesvirus transactivator protein VP16 (34–37).

ACKNOWLEDGEMENTS

We would like to thank Dr Klemm for making coordinates for the Oct-1 POU DNA complex available before publication. We are grateful to Dr E.Holthuizen for synthesizing oligonucleotides and Roel Schiphof for providing culture media. Thanks are also due to Job Dekker, Frank Holstege and Bas Werten for useful discussions and 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).

REFERENCES

1 Rosenfeld,M.G. (1991) *Genes Dev.*, **.5**, 897–907.
2 Verrijzer,C.P. and van der Vliet,P.C. (1993) *Biochim. Biophys. Acta*, **1173**, 1–21.

3 Wegner,M., Drolet,D.W. and Rosenfeld,M.G. (1993) *Curr. Opin. Cell Biol.*, **5**, 488–498.
4 Wright,P.E. (1994) *Curr. Opin. Struct. Biol.*, **4**, 22–27.
5 Coenjaerts,F.E.J., van Oosterhout,J.A.W.M. and van der Vliet,P.C. (1994) *EMBO J.*, **21**, 5401–5409.
6 Chuprina,V.P., Rullmann,J.A.C., Lamerichs,R.M.J.N., Boom,J.H., Boelens,R. and Kaptein,R. (1993) *J. Mol. Biol.*, **234**, 446–462.
7 Gehring,W.J., Affolter,M. and BYrglin,T. (1994) *Annu. Rev. Biochem.* . **63**, 487–526.
8 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.
9 Sturm,R.A., Das,G. and Herr,W. (1988) *Genes Dev.*, **2**, 1582–1599.
10 Verrijzer,C.P., van Oosterhout,J.A.W.M., van Weperen,W.W. and van der Vliet,P.C. (1992) *EMBO J.*, **10**, 3007–3014.
11 Jancso,A., Botfield,M.C., Sowers,L.C. and Weiss,M.A. (1994) *Proc. Natl. Acad. Sci. USA*, **91**, 3887–3891.
12 Klemm,J.D., Rould,M.A., Aurora,R., Herr,W. and Pabo,C.O. (1994) *Cell*, **77**, 21–32.
13 Assa-Munt,N., Mortishire-Smith,R.J., Aurora,R., Herr,W. and Wright,P.E. (1993) *Cell*, **73**, 193–205.
14 Dekker,N., Cox,M., Boelens,R., Verrijzer,C.P., van der Vliet,P.C. and Kaptein,R. (1993) *Nature*, **362**, 852–855.
15 Morita,E.H., Shirakawa,M., Hayashi,F., Imagawa,M. and Kyogoko,Y. (1993) *FEBS Lett.*, **321**, 107–110.
16 Sivaraja,M., Botfield,M.C., Mueller,M., Jansco,A. and Weiss,M.A. (1994) *Biochemistry*, **33**, 9845–9855.
17 Suzuki,M. (1993) *EMBO J.*, **12**, 3221–3226.
18 Studier,W.F., Rosenberg,A.H., Dunn,J.J. and Dubendorf,J.W. (1990) *Methods Enzymol.*, **185**, 60–89.
19 Verrijzer,C.P., Kal,A.J. and van der Vliet,P.C. (1990) *Genes. Dev.*, **4**, 1964–1974.
20 Polluck,R. and Treisman,R. (1990) *Nucleic Acid. Res.*, **21**, 6197–6204.
21 Coenjaerts,F.E.J. (1995) *Methods Enzymol.*, in press.
22 Jordan,S.R. and Pabo,C.O. (1988) *Science*, **242**, 893–899.
23 Verrijzer,C.P., Alkema,M.J., van Weeperen,W.W., van Leeuwen,H.C., Strating,M.J.J. and van der Vliet,P.C. (1992) *EMBO J.*, **11**, 4993–5002.
24 Beamer,L.J. and Pabo,C.O. (1992) *J. Mol. Biol.*, **227**, 177–196.
25 Verrijzer,C.P., Kal,A.J. and van der Vliet,P.C. (1990) *EMBO J.*, **9**, 1883.
26 Aggarwal,A.K., Rodgers,D.W., Drottar,M., Ptashne,M. and Harrison,S.C. (1988) *Science*, **242**, 899–907.
27 Wharton,R.P. and Ptashne,M. (1987) *Nature*, **326**, 888–891.
28 Weikert,M.J. and Adhya,S. (1992) *J. Biol. Chem.*, **267**, 15869–15874.
29 Li,P., He,X., Gerrero,M.R., Mok,M., Aggarwal,A. and Rosenfeld,M.G. (1993) *Genes Dev.*, **7**, 2483–2496.
30 Hochschild,A. and Ptashne,M. (1986) *Cell*, **44**, 925–933.
31 Suzuki,M. (1994) *Curr. Biol.*, **2**, 317–326.
32 Botfield,M.C. and Weiss,M.A. (1994) *Biochemistry*, **33**, 2349–2355.
33 Bork,P. and Doolittle,R. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 8990–8994.
34 Hayes,S. and O'Hare,P. (1993) *J. Virol.*, **67**, 852–862.
35 Lai,J.S., Cleary,M.A. and Herr,W. (1992) *Genes Dev.*, **6**, 2058–2065.
36 Pomerantz,J.L., Kristie,T.M. and Sharp,P.A. (1992) *Genes Dev.*, **6**, 2047–2057.
37 Walker,S., Hayes,S. and O'Hare,P. (1994) *Cell*, **79**, 841–852.
38 Kraulis,P. (1991) *J. Appl. Crystallogr.*, **24**, 946–950.
39 Schumacher,M.A., Yell Choi,K., Zalkin,H. and Brennan,R.G. (1994) *Science*, **266**, 763–770