Linker length and composition influence the flexibility of Oct-1 DNA binding

Hans C.van Leeuwen, Marijke J.Strating, Marije Rensen, Wouter de Laat¹ and Peter C.van der Vliet²

Laboratory for Physiological Chemistry, Utrecht University, Stratum, PO Box 80042, 3508 TA Utrecht and ¹Department of Cell Biology and Genetics, Erasmus University, 3000 DR Rotterdam, The Netherlands

²Corresponding author

POU domain transcription factors have two separate helix–turn–helix DNA-binding subdomains, the POU homeodomain (POUhd) and the POU-specific domain (POUs). Each subdomain recognizes a specific subsite of 4 or 5 bp in the octamer recognition sequence. The Oct-1 POU subdomains are connected by a 23 amino acid unstructured linker region. To investigate the requirements for the linker and its role in DNA recognition, we constructed POU domains in which the subdomains are connected with linkers varying in length between 2 and 37 amino acids. Binding to the natural octamer site required a minimal linker length of between 10 and 14 amino acids. A POU domain with an eight amino acid linker, however, had a high affinity for a site in which the POU recognition sequence was inverted. Computer modelling shows that inversion of the POU subdomain shortens the distance between the subdomains sufficiently to enable an eight amino acid linker to bridge the distance. DNase I footprinting as well as mutation of the POU binding site confirms the inverted orientation of the POU domain. Switching of the POU and POUhd subdomains and separation by 3 bp leads to a large distance which could only be bridged effectively by a long 37 amino acid linker. In addition to linker length, mutation of a conserved glutamate residue in the linker affected binding. As shown by surface plasmon resonance measurements, this was caused by a decrease in the on-rate. Our data indicate that there are both length and sequence requirements in the linker region which allow flexibility leading to selective binding to differently spaced and oriented subsites.

Keywords: DNA binding/linker region/octamer site/POU proteins

Introduction

High sequence specificity of *cis*-acting proteins is essential for correct target site selection. Several strategies are utilized to achieve accurate binding. Most of these require two or more proteins which either hetero- or homodimerize or bind DNA cooperatively. Increased specificity is also achieved when independent DNA-binding domains are joined via covalent linkage. Connecting two or more

domains can create a novel DNA recognition protein with combined specificity and higher affinity. Binding of one subdomain tethers the other, thereby creating a high local concentration. Because of this chelating effect, a stable DNA-binding domain is formed. Examples of such connected DNA-binding modules are zinc finger proteins (Pavletich and Pabo, 1993), the *myb* gene family (Ogata *et al.*, 1994), Cut repeat homeoproteins (Andres *et al.*, 1994) and POU domain proteins (Herr and Cleary, 1995). In the latter family of transcription factors, a homeodomain (POUhd) is covalently attached at its amino-terminus to another helix–turn–helix DNA-binding domain, the POU-specific domain (POUs). The isolated POUhd makes specific contacts to the sequence (A/T)AAT and the POU subdomain recognizes TATGC (Verrijzer *et al.*, 1992a). The recognition sequence of the bipartite Oct-1 POU protein consists of the consecutive joining of the two separate subsites to form the optimal octamer sequence, TATGC(A/T)AAT.

Extensive structural studies of the isolated subdomains and the intact Oct-1 POU domain, employing both NMR spectroscopy and X-ray diffraction, have established the folding topology of this DNA-binding domain (Assa-Munt *et al.*, 1993; Dekker *et al.*, 1993; Klemm *et al.*, 1994; Cox *et al.*, 1995). In Oct-1, both subdomains are coupled by a stretch of 23 amino acids which does not have a defined structure in the Oct-1 co-crystal (Klemm *et al.*, 1994). Proteolysis experiments show that this linker sequence is accessible to proteases both when bound to DNA (Aurora and Herr, 1992) and in solution (Botfield *et al.*, 1992), suggesting that this region is a disordered and possibly flexible part of the protein. This agrees well with the lack of sequence conservation and length observed for the >40 POU domain family members where the linker length varies from only 15 amino acids in Pit-1 (Ingraham *et al.*, 1988) to as many as 57 amino acids in the *Caenorhabditis elegans Ceh-18* gene product (Greenstein *et al.*, 1994).

Almost all residues which make DNA contacts in the Oct-1 co-crystal structure are conserved. Nevertheless, the optimal binding site differs considerably amongst the various POU domain family members (Aurora and Herr, 1992; Verrijzer *et al.*, 1992c). This may indicate that the linker sequence plays a role in site selection. Earlier experiments showed that when linkers are exchanged between Pit-1 and Oct-1, the DNA binding specificity of the POU domains is influenced but only in particular POU domain contexts (Aurora and Herr, 1992). For the Brn-2 POU protein, it was shown that the orientation of POU relative to the POUhd can be inverted for optimal binding to a site in which the POU recognition sequence is inverted (Li *et al.*, 1993).

Here we show that the linker length and composition can influence both binding specificity and affinity, independently of the DNA-binding subdomains. An important

determinant of the configuration and orientation of the bound protein appeared to be the distance to be bridged between the two subdomains.

Results

To investigate the requirements of the Oct-1 linker in POU DNA binding we made a set of deletions by introducing a second *EcoRI* site next to the unique *EcoRI* site in the linker sequence and removed the *EcoRI*–*EcoRI* fragment. Longer linker constructs were obtained by cloning double-stranded oligonucleotides into the *EcoRI* site. Proteins were bacterially expressed with a histidine tag at the amino-terminus. The various POU proteins were purified by successive application of anion exchange, nickel-NTA affinity and cation exchange chromatography (see Materials and methods). Purification was verified by SDS–PAGE and Coomassie Blue staining (Figure 1A).

The minimal linker length for binding to the octamer sequence is between 10 and 14 amino acids

Specific binding of the various mutant proteins to the optimal octamer site TATGCAAAT, was tested in a band-shift assay (Figure 1C). Equilibrium dissociation constants were calculated from the amount of proteins used to obtain 50% DNA binding (Figure 1B). Deletion of the middle eight amino acids, leaving a 15 amino acid linker, had no effect on the binding affinity, but reducing the linker length further to nine or eight amino acids resulted in a 3- to 4-fold lower affinity. This indicates that the minimal linker length for optimal binding to the octamer lies between 10 and 14 amino acids.

In addition, there are also compositional requirements since the 16 and 19 amino acids linkers, which obviously are longer than the 15 amino acids linker, have lower binding affinities (discussed below).

The 4-fold lower affinity of the eight amino acid linker protein indicates that POU's still contributes to DNA binding, since deletion of the whole POU's subdomain results in a 600-fold lower affinity (Verrijzer *et al.*, 1992a). When the linker region is almost completely deleted, leaving only two amino acids, DNA binding can no longer be detected (Figure 1C).

Extending the linker to 28 or 37 amino acids did not result in higher affinity for the octamer site.

Separation of the POU's and POUhd recognition sites cannot be compensated for by a longer linker

Although no protein–protein interactions exist between POU's and POUhd in the co-crystal structure (Klemm *et al.*, 1994), separation of both DNA-binding sites by introducing one or two C:G base pairs between the subsites (TATGCCAAAT or TATGCCCAAAT) resulted in lower affinity for the wild-type POU domain (Figure 2 and Klemm and Pabo, 1996). This loss in affinity might be caused by restrictions imposed by the length of the linker and, therefore, we tested various lengths (Table I, Figure 2). The binding characteristics of the 15, 23 or 37 amino acid linker proteins were almost identical with all three differently spaced sites (Table I). Thus the lower affinity cannot be compensated for by an extended linker. The eight amino acid linker showed a 3- to 4-fold reduced

Fig. 1. Oct-1 POU domain linker length influences DNA binding. (A) SDS gel of the purified His-tagged Oct-1 POU domain proteins with different linker lengths. The linker length in amino acids is indicated above the lanes. A total of 0.5 μ g of each protein was used. The gel was stained with Coomassie Blue. M = marker lane. (B) Amino acid sequence of linker deletion/insertion constructs and their relative dissociation constants for the octamer site (TATGCAAAT) compared with the wild-type protein ($K_d = 1.4$ nM set at 1.0). – = no binding detected ($K_d > 1$ μ M). All values are the averages of three measurements. Inserted amino acids are shown in bold. (C) Gel retardation experiment with 10 nM of the various proteins performed on an octamer-containing probe.

affinity with all three binding sites, indicating that the distance constraint on all three sequences is comparable.

Modelling of the spaced subsites

To obtain information on the actual restrictions imposed by the linker, we modelled the subdomains on the separated sites using the coordinates of the co-crystal structure of

Fig. 2. Spacing of the POU and POUhd sites decreases DNA binding and is not restored by longer linker lengths. Labelled oligonucleotides were incubated with increasing amounts of proteins and assayed by gel retardation. The percentage of free DNA is plotted against amounts of protein. Probes contained the consensus site (left panel), the +1 site (middle panel) or the +2 site (right panel). ●, eight amino acid linker; □, 15 amino acid linker; ▲, 23 amino acid linker; △, 37 amino acid linker.

Table I. Comparison of relative binding affinities of four different linker length Oct-1 POU proteins for differently spaced and oriented sites

		Relative dissociation constants			
		8 aa	15 aa	23 aa	37 aa
Normal orientation					
0	CGGCTATG CAAAT CACG	4	1	1	1
+1	CGGCTATG CCTAAT CAC	80	25	21	21
+2	CGGCTATG CCCTAAT CA	375	140	140	140
Inverted orientation					
0	CGGCGCATTAATCACGG	190	50	40	40
+1	GGCGCATCTAATCACGG	—	65	45	45
+2	GCGCATCCTAATCACGG	13	20	28	35
+2 m1	GCGTATCCTAATCACGG	300	25	16	30
+2 optimal	GTGCATACTAATCACGG	1	1	3	3
Switched sites					
+3	CGGCTAATCCTATGCGC	140	47	23	19
+3 m2	CGGCTAATCCTATCGGC	nt	nt	nt	75
+3 inverted	CGGCTAATCCCGCATAG	—	295	115	19
+3 inv. m3	CGGCTAATCCCGATAG	nt	nt	nt	95

Equilibrium dissociation constants were measured from the amount of protein necessary to obtain 50% DNA binding. Affinities are presented as relative K_d s compared with the K_d of the wild-type protein for the octamer site ($K_d = 1.4$ nM). Bold letters indicate bases presumed to be contacted by the POU domain. 0/+1/+2/+3 indicate the number of C:G base pair(s) separating the POU and POUhd. — = no binding detected ($K_d > 1$ μ M). All values are averages of at least two measurements, and the deviation was no more than 15%.

the complete Oct-1 POU domain on the octamer site (Klemm *et al.*, 1994). Employing the program InsightII (Biosym Technologies), both subdomains connected to their recognition sites were defined as separate objects and disconnected. After insertion of one or two C:G base pairs, the objects were reconnected (Figure 3A). From these three modelled structures, we estimated the minimal connecting distance in space between the carboxy-terminus of POU to the amino-terminus of POUhd (Figure 3A). In all cases, this straight line runs through part of the protein. The actual path of the linker will therefore be longer. For the contiguous octamer site, the straight distance is 28 Å and the shortest distance over the surface of POU is 32 Å. Assuming a length spanned by one

amino acid of 3 Å, the 23 amino acid wild-type linker could easily bridge this distance but the eight amino acid linker could not. Because POU binding was detected for the eight amino acid linker protein (Figures 1 and 2), some conformational change must allow POU binding. It is known that the POU protein bends the DNA slightly (Verrijzer *et al.*, 1992b; Klemm *et al.*, 1994). One could envisage that an increased DNA bend could bring these ends closer together. In a circular permutation assay, however, we could not detect any difference between the eight and 37 amino acid linker in DNA bending (Figure 4).

Upon insertion of two C:G base pairs (site referred to as +2), the straight distance in the modelled complex did not increase much (31 Å), whereas upon insertion of one

Fig. 3. Modelled arrangements of the Oct-1 POU subdomains bound to differently spaced and oriented sites. The POU homeodomain (Hd) and POU-specific domain (Sp) are indicated. The sequence and orientation of the domains are indicated above the structure (arrows according to Herr and Cleary, 1995). The distance in Ångstroms between the carboxy-terminus of POU (Glu75, indicated with C) and the amino-terminus of POUhd (Arg102, indicated N) is indicated. Models were generated with the program InsightII (Biosym Technologies) and depicted with the program Molscript (Kraulis, 1991). DNA contacts are according to Klemm *et al.* (1994). (A) Arrangements on ATGC n TAAT with spacing $n = 0, 1$ or 2 C:G base pair(s). (B) Inverted arrangements on GCAT n TAAT with spacing $n = 0, 1$ or 2 C:G base pair(s). (C) Switched arrangements on TAAT n ATGC with spacing $n = 0$ or 3 C:G base pair(s). (D) Switched, inverted arrangements on TAAT n GCAAT with spacing $n = 0$ or 3 C:G base pair(s).

Fig. 4. Linker length does not influence DNA bending on an octamer site. The eight and 37 amino acid linker POU proteins were bound to three labelled circular permutation fragments containing an octamer site either at the 5' end (A), in the centre (D) or at the 3' end (H) of the fragments.

C:G (+1) base pair in the complex this distance is slightly shorter (25 Å, Figure 3A). The insertions led to a rotation of POU_s towards POU_{hd}, thus bringing the subdomains closer together. Since 15 amino acids are sufficient to bridge the 'wild-type' distance between the subdomains this explains why we did not observe differences in binding affinity between the 15 and 37 amino acid linker proteins.

The positional rotation of POU_s towards the POU_{hd} might also explain the lower affinity of the POU proteins for the spaced sites. When bound to the +1 and +2 sites, POU_s residues come into close proximity with the POU_{hd} minor groove residues which will most likely lead to steric hindrance.

Binding to inverted POU_s sites

To see if the linker length could influence the flexibility and freedom of movement of the individual subdomains, we inverted the POU_s recognition sequence and varied the intervening nucleotides, GCAT-*n*-TAAT (*n* = 0, 1 or 2).

Model building (Figure 3B) showed that inversion of the POU_s-DNA complex without introducing nucleotides hardly changed the minimal distance connecting both subdomains when compared with the normal octamer orientation (26 versus 28 Å). Further separation decreased the connecting distance to 18 (+1) and 19 Å (+2). No obvious interference of POU_s and POU_{hd} residues occurs. The +2 inverted site represents the only configuration in which the linker does not have to pass over the POU_s surface.

The binding affinity of wild-type POU protein for the three inverted sites varies slightly and is 28- to 45-fold

lower compared with the octamer site (Table I). In these cases, the distance between the two subdomains can be bridged by the 23 amino acid linker. The binding characteristics of the 15 and 37 amino acid linker proteins are comparable with the 23 amino acid protein (Table I). For the eight amino acid linker a different pattern was seen (Table I). Weak binding (190-fold lower) was observed to the GCATTAAT (*n* = 0) site and no binding to the +1 site was detected. However, a surprisingly high affinity was observed for the +2 site despite the almost equal minimal connecting distance (18 versus 19 Å). This difference could be explained by the fact that the straight linker line in the +1 complex runs through the protein, suggesting steric interference, while on the +2 site the linker does not encounter any part of the protein (Figure 3B).

To investigate whether POU_s is really inverted, we used two approaches: mutagenesis and DNase I footprinting. An essential contact of the POU_s subdomain is the third G:C base pair which is contacted by two amino acid residues in the recognition helix (Klemm *et al.*, 1994). We mutated this base pair in the inverted +2 site to an A:T base pair, leading to GTATccTAAT (+2 m1, Table I). Changing this base pair should lead to a lower affinity in case the POU_s subdomain prefers to bind to the inverted (GCAT) sequence. However, if POU_s binds in the normal orientation sequence, this should lead to a higher affinity since a T at the -1 position in the octamer orientation is a preferred contact via a hydrophobic T-methyl interaction (Botfield *et al.*, 1994; van Leeuwen *et al.*, 1995). Binding of the eight amino acid linker to this mutated site appeared to be strongly reduced (Table I), indicating that POU_s indeed binds in the inverted orientation. Binding of the longer linker length proteins which bind preferentially in the normal orientation is hardly affected by this mutation. The 23 amino acid linker protein even binds slightly better due to the extra T -1 contact.

Another consequence of inverted POU_s binding is that extension of this site to TGCATA should lead to a higher affinity site since these bases are additional contacts in the octamer site (Verrijzer *et al.*, 1992a; Klemm *et al.*, 1994). Indeed, this site became more tightly bound by the eight amino acid linker (+2 optimal, Table I) and was even 3-fold higher than the wild-type protein. From these mutation studies, we conclude that the POU_s subdomain can bind in the inverted orientation on a reversed site.

DNase I footprinting confirmed the inverted POU_s orientation. We compared the octamer sequence with the inverted +2 optimal site since both are high affinity sites. If POU_s is inverted this should produce a larger protected region on the inverted +2 optimal site. This is indeed what we observe (Figure 5). On the octamer site the region protected by the eight amino acid protein is 3 bp shorter than on the inverted +2, indicating that POU_s binds in an inverted orientation rather than tolerating the imperfect ATAC in the octamer arrangement. The same result was obtained for the wild-type protein, showing that this also binds in the inverted orientation to the +2 optimal site. The homeodomain border was not visible on these footprints because the smallest DNA fragments do not precipitate efficiently. When the bottom strand was footprinted, the homeodomain was positioned normally over the TAAT sequence both on the octamer and on the inverted site (data not shown).

Fig. 5. DNase I footprint analysis shows inverted orientation of the POU subdomain. Footprints were performed on the octamer site (A) and on the reversed +2 optimal POU site (B) with increasing amounts (0.25, 0.5, 1 and 2 μ M) of the wild-type protein (lanes 3–6) and the eight amino acid linker protein (lanes 8–11). Lanes 2 and 7, no protein; lane 1, G + A sequencing ladder. Note that some aspecific cleavage of T bases has occurred. The protected regions are indicated with a black bar. The 3' sequence is not visible because of insufficient precipitation of small DNA fragments.

Switching POU and POUhd recognition sequences

The observed flexibility of the POU subdomain seems not to be restricted to its ability to invert. The POU subdomain can also be positioned on the other side of the POUhd, leading to a POUhd–POU order rather than POU–POUhd. Such a POUhd–POU configuration was suggested for the Oct-1–VP16 complex (Cleary and Herr, 1995) and was observed recently for the *Drosophila* Drifter (DFR) protein (Certel *et al.*, 1996).

To study whether this was possible for Oct-1, we first switched the POU–DNA complex and the POUhd complex in the computer models. Contiguous binding to TAAT-ATGC, as suggested for DFR, or TAAT-GCAT, seems very unlikely in view of the large overlap of POU and POUhd in the major groove (Figure 3C and D). Further spacing of the elements, however, prevented overlapping protein contacts and, therefore, binding to these sites seems plausible.

We tested the three nucleotide spaced sites because no overlapping DNA contacts are made in these configurations (Figure 3C and D). Binding of the wild-type protein to the switched +3 site with a 34 Å connecting distance between POUhd and POU shows a 23-fold lower affinity. An almost equal affinity is observed for the 37 amino acid linker protein (Figure 6A, Table I). The 15 amino acid linker which has wild-type affinity on all sites tested so far now shows a 2-fold lower affinity compared with wild-type. The eight amino acid linker has a much lower

Fig. 6. Binding to the switched sites reveals a correlation between linker length and affinity. The various POU linker length proteins were bound to the sites 5'-TAATcccGCAT-3' (A) and 5'-TAATcccATGC-3' (B). Lanes 1, 6, 11 and 16 contain the proteins at 250 nM. Concentrations were decreased 2.5-fold in four subsequent lanes of each set.

CLASS2 POU FAMILY LINKERS

POUs	Linker	POUhd
EDAD STVAKSGGGVFNINTMTSTLSSTP ESILG RRRK KR		dPOU28 <i>drosophila</i>
DDAD RTIQATGG-VFDPAALQATVS-TP EIIIG - RRRK KR		dPOU19 <i>drosophila</i>
DDAD -----TTVSNP-ALLGPHN-SP EGIN - RRRK KR		SpOCT <i>sea urchin</i>
NDAE --NLSSDSSLSSPSALNSP---G IEGLS - RRRK KR		Oct-1 <i>human</i>
NDAE --NLSSDSTLSSPSALNSP-GQ IEGVN - RRRK KR		Oct-1 <i>c.elegans</i>
NDAE --NLSSDSALCSPSALNSP-GLGV EGLN - RRRK KR		Oct-1 <i>pig</i>
NDAE --NLSSDSTASSPSDLNSP-GLGA EGLN - RRRK KR		Oct-1 <i>mouse/rat</i>
NDAE --NITSDSTLTNQSVLNSP-GHGM EGLN - RRRK KR		Oct-1 <i>xenopus</i>
NDAE --TMSVDSSLPSPNQLSSPS-LGF DGLPG RRRK KR		Oct-2 <i>human/mouse/rat</i>
NDAE --SSPSDPSAALPV---RPH--SQ EYLR - RGRGDR		Rov <i>rat</i>
NDAE --SSPADPSASTPS--SYPT--LS EVFG - RRRK KR		Skn-1A <i>rat/mouse</i>
NDAE --SAPSD-TALTPSA-NY--AQL SEMF FG - RRRK KR		Nrl16 <i>xenopus</i>

Fig. 7. Alignment of POU class 2 linker sequences with low gap penalty and few amino acids flanking the linker region. Conserved residues are shown in bold. The arrow indicates the conserved Glu95 which was mutated to Lys (E95K, see text).

affinity, presumably due to the long distance to be bridged (Table I). The contribution of POU's binding on this distant site was confirmed by mutating two POU's contacts (+3 m2, Table I), which resulted in a lower affinity of the 37 amino acid linker protein.

The switched, inverted +3 site TAATcccGCAT has a minimal spanning distance of 44 Å, but the actual linker path is probably longer since the straight line runs through the protein (Figure 3D). At this large connecting distance, a different pattern is observed in which the longer the linker length, the higher is the DNA binding affinity (Figure 6B and Table I). The wild-type protein binds with low affinity to this site, but extension to 37 amino acids allows more effective binding.

Since the linker length in the POU class 2 family (to which Oct-1 belongs) varies from 19 to 29 amino acids (Figure 7), it seems plausible that different spacing preferences are possible within the members of this group.

A glutamate residue in the linker is required for optimal binding

In addition to the linker length, the linker composition is also important for site-specific binding. As shown in Figure 1B, deletion of eight amino acids in the middle part of the linker does not impair binding to the octamer, whereas deletion of seven amino acids of the linker adjacent to the homeodomain resulted in a 5-fold lower affinity (Figure 1B). A smaller deletion in the same region, leaving as many as 19 amino acids resulted in a 2.5-fold weaker binding. Alignment of all known class 2 POU linkers (Figure 7) shows low homology throughout divergent species. A glutamate residue is almost completely conserved and is removed in the 16 and 19 amino acid proteins. To study the role of this glutamate residue in DNA binding, we mutated it to a lysine (E95K). This led to a 2.5-fold reduced affinity (data not shown).

We also tested both the wild-type and the E95K protein in an IBIS surface plasmon resonance biosensor which allows real time analysis of protein–DNA complex formation. The cuvet setup was a 5'-biotinylated octamer-binding site bound to a streptavidin sensor chip (Pharmacia). Association was measured at three protein concentrations. Curve fitting of the association phase resulted in

Fig. 8. Linker mutant E95K has a lower on-rate as shown by surface plasmon resonance measurement. Kinetic parameters were determined using a surface plasmon resonance IBIS instrument (Intersens instruments). The slope of the plotted line ($-k_s = k_a \cdot C + k_d$) is equal to the association rate constant and the y-intercept is equal to the dissociation rate constant. Equilibrium dissociation constants ($K_d = k_d/k_a$) are indicated in the box. Because standard surface interactions are measured at 150 mM NaCl without ficoll, the K_d is lower than measured under bandshift conditions.

the determination of the k_s values [$-k_s = k_a \cdot C + k_d$ (O'Shannessy *et al.*, 1993)]. Plotting of $-k_s$ values versus the concentration (Figure 8) showed that the association constant for E95K is 3-fold lower compared with wild-type, while the dissociation constant is almost equal.

These measurements show that the lower affinity of the linker E95K mutant is caused by a lower on-rate of the protein and that there is little difference in off-rate.

Discussion

In this study, we show that both the length and the composition of the linker region connecting the two subdomains of Oct-1 can influence the specificity and affinity of POU domain DNA binding. We tested the affinity of various linker length POU proteins on recogni-

tion sites which have differently arranged and spaced subsites. Using computer-built models, we estimated the length that the linker has to bridge on these variably spaced and oriented subsites.

The linker length influences DNA recognition

The minimal linker length required for optimal binding to the octamer site lies between 10 and 14 amino acids. This fits well with the shortest natural linkers (15 amino acids) found in the POU family. Such a linker could span 45 Å, allowing some flexibility since the smallest measured distance connecting POU and POUhd in the Oct-1 co-crystal structure is 32 Å.

A large deletion resulting in an eight amino acid linker resulted in a 4-fold lower affinity for the octamer site because this linker is too small to bridge the distance between the subdomains. This eight amino acid linker protein, however, has a high affinity for a site in which the optimal POU recognition site is inverted and separated from the POUhd site by 2 bp, tGCATacTAAT. Model building shows that on this site the distance to be bridged between POU and POUhd is only 19 Å, explaining why the eight amino acids now can bind efficiently. This affinity is even 3-fold higher than the longer wild-type protein, possibly due to the higher flexibility of the wild-type protein in solution.

While DNA binding of a short linker can be restored on a site which brings the two subdomains closer together, a large linker can (partially) restore binding to two subsites which are distant. The site TAATcccGCAT creates a connecting distance over the surface of at least 50 Å. This site is bound six times more efficiently by a 37 amino acid linker protein than by the wild-type protein.

These data show that the linker length can play an active role in site selection. The sites efficiently bound by the eight and 37 amino acid proteins do not resemble the octamer-binding site and therefore would not be detected easily in a computer search for possible target sites.

The linker composition influences binding affinity

Two observations show that the linker composition influences binding to DNA. The 16 amino acid linker which has nearly the same length as the 15 amino acid linker protein, but is different in amino acid sequence, has a 5-fold lower affinity for the octamer site. Secondly, a linker mutant E95K has a 3-fold lower on-rate but off-rates comparable with the octamer site. Apparently, DNA docking is influenced by the linker but not its stability on DNA. This seems to exclude a direct DNA contact by the glutamate residue. Another observation which seems to rule out DNA contacts by the linker is that the linker region attached to POUhd by itself does not increase the affinity (data not shown). An explanation for the observed lower affinity of the E95K protein could be that there is a structural constraint in the linker region which determines the flexibility of the overall POU structure in solution. Introducing a positively charged residue could affect this flexibility.

Earlier experiments have shown that swapping Oct-1 and Pit-1 linkers only influences specificity depending on the POU DNA-binding domain context (Aurora and Herr, 1992). This context dependency could indicate that the linker makes protein-protein contacts on the POU surface.

The linker glutamate could form a salt bridge with a lysine or arginine residue on the POU surface. This salt bridge would be disrupted in the E95K mutant. We found two possible candidates on the POU surface: Lys36 and Lys69. However, changing these to Glu in the linker E95K mutant context, and thereby possibly restoring a reversed salt bridge, did not restore affinity to wild-type levels (data not shown).

A comparable E→K mutation in the linker region was found in a random mutagenesis screen of the Pit-1 DNA-binding domain fused to the GCN4 transactivation domain with *LacZ* as an indicator gene, but the precise effect of this mutation in this screen is unknown (Liang *et al.*, 1995).

The octamer site remains the optimal binding site

Insertions of up to 2 bp between the consecutive POU- and POUhd-binding sites do not change the connecting distance much. Nevertheless, binding to these sites is reduced. Several explanations are possible. First, computer modelling shows the possible interference of POU residues and part of the POUhd in the centre. Another explanation could be that preferential DNA contacts of POU and POUhd to the same base pair(s) are lost upon spacing (Klemm and Pabo, 1996) or that new overlaps are disadvantageous. Klemm and Pabo (1996) showed that the isolated POU and POUhd bind cooperatively even in the absence of the linker. Since no protein-protein contacts were detected between the domains in the crystal structure, they suggest that overlapping DNA contacts near the centre of the octamer site may mediate this cooperativity and explain why the non-spaced octamer site is the preferred site. Several of such joined contacts have indeed been observed. For example, the fifth A/T base pair of the octamer site (ATGCA~~A~~AAT) seems to be contacted by the POU subdomain, via a major groove contact (Gstaiger *et al.*, 1996), and by the POUhd in the minor groove (Klemm *et al.*, 1994). However, when we mutated this POU residue (Leu55) or the POUhd contacts (Lys103 and Arg105), this did not influence its binding pattern to spaced sites (data not shown). Also, mutation of a POU residue which makes a phosphate backbone contact to the fifth A:T base pair (Asn59) did not change the spacing preference. Thus, no evidence has been obtained so far that a single residue with overlapping contacts is responsible for keeping the two POU subdomains in the octamer arrangement.

Finally, it could be that an overall DNA configuration causes the preference for the contiguous octamer site. Verrijzer *et al.* (1992b) showed by biochemical means that the POU domain bends the DNA slightly, and a DNA bend was also seen in the co-crystal structure (Klemm *et al.*, 1994). Indirect evidence for structural changes in the DNA comes from the DNase I footprints displaying hypersensitive sites, higher up in the gel, when POU binds in the inverted orientation (Figure 5). Such a change in DNA bending angle was observed when the tail connecting the MAT α 2 and MAT α 1 homeodomains was extended (Jin *et al.*, 1995; Li *et al.*, 1995). The normal spacing of 6 bp between the α 2 and α 1 half-sites can be increased to 7 bp when three glycine residues are inserted within this linker.

Differences in binding flexibility might influence gene activation

Differences in sequence specificity between the many members of the POU domain family can be achieved by preferential binding to particular spaced and orientated subsites (Li *et al.*, 1993; Certel *et al.*, 1996). As a consequence, interactions with other proteins might be influenced.

An example of an interacting protein which requires a particular recognition site is the herpes simplex virus co-activator VP16 which associates with the Oct-1 homeodomain (Lai *et al.*, 1992; Pomerantz *et al.*, 1992). This multiprotein–DNA complex is formed on the sequence TAATGAGATAC but not on the optimal octamer site (Walker *et al.*, 1994). Oct-1 by itself binds weakly to this site but is stabilized by its association with VP-16. POU binding to this sequence has been suggested 3' of the POUhd TAAT sequence, contacting either ATAC or the opposite strand TATCT (Cleary and Herr, 1995). The first arrangement is comparable with the site TAATcctATGC which we tested but is more stably bound by Oct-1 due to the optimal POU sequence (Verrijzer *et al.*, 1990). In contrast to VP16, a B-cell-specific transcriptional co-activator of Oct-1 and Oct-2, variously termed Bob1, OBF-1 or OCA-B (Gstaiger *et al.*, 1995; Luo and Roeder, 1995; Strubin *et al.*, 1995), only forms complexes with Oct-1 on a contiguous octamer arrangement but not on the TAATGARAT sequence (Gstaiger *et al.*, 1996). Thus, these two factors require different recognition sequences of Oct-1. Co-factors, rather than being dependent on the sequence, can also dictate the sequence arrangement. One could envisage that factors bound to the surface of the POU domain influence the path the linker takes and thus the arrangement of the subdomains.

Most POU domain transcription factors are expressed in different cell types where they are implicated in developmental regulation through specific activation of their target genes (Schöler, 1991). It is evident that different preferences in orientation and spacing of the subsites may influence gene activation either directly or via interacting proteins.

In many POU family members the coding sequence for the two subdomains, POU and POUhd, are separated by an intron in the genome. This could be a potential determinant of linker length if differential splicing of the intron occurs. The Oct-1 linker intron is out-of-frame with the coding sequence (Sturm *et al.*, 1993), thus only alternative splicing of this intron could generate a different functional protein. To our knowledge, no such Oct-1 mRNAs have been reported. In the case of Oct-2, lack of removal of this particular intron (93 bp) might indeed lead to lengthening of the linker (an extra 31 amino acids) since this intron does not contain a stop codon (Matsuo *et al.*, 1994). However, such a protein has not been observed so far.

Using models to analyse putative conformations

The Oct-1 co-crystal structure enabled us to rearrange the two subdomains on differently spaced and oriented sites and subsequently measure either the straight distance or estimate the distance over the surface that the linker has to bridge in order to connect the two subdomains. These models cannot, by definition, take into account changes

in the protein or DNA conformation and this may limit our interpretation of the data. However, such modelling can be effective as shown here and in a comparable computer modelling strategy, where Pomerantz and co-workers reasoned that a four amino acid linker connecting a fusion between the Zif268 zinc finger motif and the Oct-1 homeodomain only allowed one orientation in which the carboxy-terminal end of the zinc finger was within 8.8 Å of the connecting amino-terminal arm of the homeodomain (Pomerantz *et al.*, 1995).

Combining two DNA-binding domains with a flexible linker into a single structure has provided the cell with a new site-specific DNA-binding protein. Further divergence of the linker length and composition has created even more complexity in sequence recognition. Different arrangements of the subdomains directed by linker composition may be accompanied by other protein–DNA and protein–protein interactions in various POU family members, leading to further stabilization. Such a fine tuning can only be unravelled if other POU–DNA complex structures on differently spaced sites are solved.

Materials and methods

Construction of linker mutants

The construction of POU linker mutants was facilitated by the presence of a unique *EcoRI* site in the linker region. Using oligonucleotide-directed *in vitro* mutagenesis (Promega), nucleotides were added and/or changed to introduce a second *EcoRI* site. The *EcoRI*–*EcoRI* fragment subsequently was removed and the ends religated, thereby creating small deletions. The mutations are given below: the newly introduced nucleotides are in bold, the *EcoRI* sites are underlined. The numbers indicate the amino acid linker length after removal of the *EcoRI* fragment. 9, gaattcctctcatctgattcgtccctctccagcccaagtgcctgaattcccaggaattgagggctt-gag; 15, gaacctctcatctgattcgaattcctctccagcccaagtgcctgaattcccaggaattgagggctt-gag; 23, gaacctctcatctgattcgtccctctccagcccaagtgcctgaattcccaggaattgagggctt-gag; 19, gaacctctcatctgattcgtccctctccagcccaagtgcct-gaattcccaggaattgagggctt-gag; 16, gaacctctcatctgattcgtccctctccagcccaagtgcctgaattcccaggaattgagggcttgaattc.

Ligation of the nine amino acid mutant *EcoRI* site to the 16 amino acid mutant *EcoRI* site resulted in the two amino acid linker construct. Combination of the 15 amino acid mutant *EcoRI* site with the 16 amino acid mutant *EcoRI* site resulted in the eight amino acid linker construct.

The 28 amino acid linker was created by hybridization of two oligonucleotides, 5'AATTCTACCGCTCC3' and 5'AATTGGAGGCGGTAG3', which were cloned into the *EcoRI* site of the wild-type linker. The 37 amino acid linker was created by introducing the 27 bp *EcoRI*–*EcoRI* fragment which was removed from the 15 amino acid linker mutant in the unique *EcoRI* site of the 28 amino acid linker construct. For amino acid sequences of the deletion constructs, see Figure 1B.

Expression and purification of wild-type and mutant POU proteins

Oct-1 POU constructs were cloned in pET15b (Novagen) as described in van Leeuwen *et al.* (1995). Proteins were expressed in *Escherichia coli* BL21(DE3) pLysS cells using the T7 expression system (Studier *et al.*, 1990). The His-tagged fusion proteins were purified on DEAE–Sephacel (Pharmacia), Ni²⁺–nitrilo-tri-acetic acid (Qiagen) columns essentially as described by the manufacturer and an SP Sepharose (Pharmacia) column as described in van Leeuwen *et al.* (1995). Purification was checked on a 15% SDS–polyacrylamide gel.

DNA-binding studies

The probes used for bandshift assays were double-stranded oligonucleotides, end-labelled with T4 polynucleotide kinase and purified by preparative polyacrylamide gel electrophoresis. The sequences are as indicated in Table I. DNA concentrations were determined by absorption at 260 nm. The concentration of input DNA was 0.5 nM. Binding reactions were carried out for 60 min on ice in 20 ml of binding buffer [20 mM HEPES–KOH pH 7.5, 1 mM EDTA, 1 mM dithiothreitol (DTT),

0.025% NP-40, 4% Ficoll, 100 mM NaCl]. Free DNA and protein–DNA complexes were separated on a 15% polyacrylamide gel (37.5:1) run in $0.5\times$ TBE at 4°C for 4 h at 100 V. Dried gels were exposed and quantified by phosphorimaging (Molecular Dynamics). The equilibrium dissociation constant (K_d) was calculated at half saturation from the equation $K_d = P_t/D_b$ (D_b = DNA bound) (Verrijzer *et al.*, 1992a). The total protein concentration (P_t) was calculated using a deduced M_r of 20 kDa for the wild-type His₆-POU protein. Appropriate corrections were made for insertions and deletions.

Circular permutation assay

DNA fragments were generated by digestion of plasmid pBend2Ad4 (Verrijzer *et al.*, 1992b) with restriction enzymes *Mlu*I (A), *Eco*RV (D) and *Bam*HI (H). Fragments were dephosphorylated and end-labelled with [γ -³²P]ATP and polynucleotide kinase. Binding conditions were as described above. Free DNA and protein–DNA complexes were separated on an 8% polyacrylamide gel (37.5:1) run in $0.5\times$ TBE at 4°C for 16 h at 5 V/cm.

DNase I footprinting

The DNA used in the footprinting assays was a 40 bp *Eco*RI–*Xba*I fragment from the plasmids WT38, 45.02 and 55.07 described before (van Leeuwen *et al.*, 1995). The *Eco*RI site was end-labelled by a partial fill-in reaction with Klenow polymerase and [α -³²P]dATP. Footprint reactions were performed as described previously (Verrijzer *et al.*, 1990). Amounts of protein are indicated in the figure legends.

Surface plasmon resonance measurements

Real time analyses of POU–DNA interactions have been investigated using an optical IBIS biosensor device (Intersense Instruments, Amersfoort, The Netherlands) based on surface plasmon resonance signals which are related to the number of molecules bound to the sensor surface. 5'-Biotinylated DNA-binding sites were immobilized to a streptavidin sensor chip (Pharmacia) at $\sim 1\times 10^{-13}$ mol/mm². Association rates were measured at various protein concentrations in 60 ml of phosphate-buffered saline/0.02% NP-40. Biphasic curve fitting (O'Shannessy *et al.*, 1993) of the binding curves reveals the k_s value ($-k_s = k_a \times C + k_d$). When $-k_s$ values are plotted versus the protein concentrations, the slope equals the association rate constant (k_a) and the y-intercept equals the dissociation rate constant (k_d).

Modelling

Models were constructed by excising the POU subdomains plus their respective recognition sequences, 5'ATGC3' and 5'AAAT3', as separate objects from the original co-crystal structure (ID = 1oct; (Klemm *et al.*, 1994). The POU_s complex was reconnected with the POU_hd in the two possible orientations and, in some cases, 1–3 B-DNA C:G base pair(s) were first connected to the appropriate end of the POU_s recognition sequence. Construction of these models was performed with the program InsightII (Biosym Technologies) on an Indigo XZ machine (Silicon Graphics) and depicted with the program Molscript (Kraulis, 1991).

Acknowledgements

We would like to thank Job Dekker, Frank Holstege, Marian Walhout and Bas Werten for stimulating discussions and Marc Timmers 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).

References

Andres, V., Chiara, M.D. and Mahdavi, V. (1994) A new bipartite DNA-binding domain: cooperative interaction between the cut repeat and homeo domain of the cut homeo proteins. *Genes Dev.*, **8**, 245–257.
 Assa-Munt, N., Mortishire-Smith, R.J., Aurora, R., Herr, W. and Wright, P.E. (1993) The solution structure of the Oct-1 POU-specific domain reveals a striking similarity to the bacteriophage λ repressor DNA binding domain. *Cell*, **73**, 193–205.
 Aurora, R. and Herr, W. (1992) Segments of the POU domain influence one another's DNA-binding specificity. *Mol. Cell. Biol.*, **12**, 455–467.
 Botfield, M.C., Jansco, A. and Weiss, M.A. (1992) Biochemical characterization of the Oct-2 POU domain with implications for bipartite DNA recognition. *Biochemistry*, **31**, 5841–5848.

Botfield, M.C., Jansco, A. and Weiss, M.A. (1994) An invariant asparagine in the POU-specific homeodomain regulates the specificity of the Oct-2 POU motif. *Biochemistry*, **33**, 8113–8121.
 Certel, K., Anderson, M.G., Shrigley, R. and Johnson, W.A. (1996) Distinct variant DNA-binding sites determine cell-specific autoregulated expression of the *Drosophila* POU domain transcription factor Drifter in midline glia or trachea. *Mol. Cell. Biol.*, **16**, 1813–1823.
 Cleary, M.A. and Herr, W. (1995) Mechanisms for flexibility in DNA sequence recognition and VP16-induced complex formation by the Oct-1 POU domain. *Mol. Cell. Biol.*, **15**, 2090–2100.
 Cox, M., van Tilborg, P.J., de Laat, W., Boelens, R., van Leeuwen, H.C., van der Vliet, P.C. and Kaptein, R. (1995) Solution structure of the Oct-1 POU homeodomain determined by NMR and restrained molecular dynamics. *J. Biomol. NMR*, **6**, 23–32.
 Dekker, N., Cox, M., Boelens, R., Verrijzer, C.P., van der Vliet, P.C. and Kaptein, R. (1993) Solution structure of the POU-specific DNA-binding domain of Oct-1. *Nature*, **362**, 852–855.
 Greenstein, D., Hird, S., Plaster, R.H., Andachi, Y., Kohara, Y., Wang, B., Finney, M. and Ruvkun, G. (1994) Targeted mutations in the *Caenorhabditis elegans* POU homeo box gene *ceh-18* cause defects in oocyte cell cycle arrest, gonad migration, and epidermal differentiation. *Genes Dev.*, **8**, 1935–1948.
 Gstaiger, M., Knoepfel, L., Georgiev, O., Schaffner, W. and Hovens, C.M. (1995) A B-cell coactivator of octamer-binding transcription factors. *Nature*, **373**, 360–362.
 Gstaiger, M., Georgiev, O., van Leeuwen, H., van der Vliet, P. and Schaffner, W. (1996) The B cell coactivator Bob1 shows DNA sequence-dependent complex formation with Oct-1/Oct-2 factors, leading to differential promoter activation. *EMBO J.*, **15**, 2781–2790.
 Herr, W. and Cleary, M.A. (1995) The POU domain: versatility in transcriptional regulation by a flexible two-in-one DNA-binding domain. *Genes Dev.*, **9**, 1679–1693.
 Ingraham, H.A., Chen, R.P., Mangalam, H.J., Elsholtz, H.P., Flynn, S.E., Lin, C.R., Simmons, D.M., Swanson, L. and Rosenfeld, M.G. (1988) A tissue-specific transcription factor containing a homeodomain specifies a pituitary phenotype. *Cell*, **55**, 519–529.
 Jin, Y., Mead, J., Li, T., Wolberger, C. and Vershon, A.K. (1995) Altered DNA recognition and bending by insertions in the alpha 2 tail of the yeast a1/alpha 2 homeodomain heterodimer. *Science*, **270**, 290–293.
 Klemm, J.D. and Pabo, C.O. (1996) Oct-1 POU domain–DNA interactions: cooperative binding of isolated subdomains and effects of covalent linkage. *Genes Dev.*, **10**, 27–36.
 Klemm, J.D., Rould, M.A., Aurora, R., Herr, W. and Pabo, C.O. (1994) Crystal structure of the Oct-1 POU domain bound to an octamer site: DNA recognition with tethered DNA-binding modules. *Cell*, **77**, 21–32.
 Kraulis, P.J. (1991) Molscript: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.*, **24**, 946–950.
 Lai, J.S., Cleary, M.A. and Herr, W. (1992) A single amino acid exchange transfers VP16-induced positive control from the Oct-1 to the Oct-2 homeo domain. *Genes Dev.*, **6**, 2058–2065.
 Li, P., He, X., Guerrero, M.R., Mok, M., Aggarwal, A. and Rosenfeld, M.G. (1993) Spacing and orientation of bipartite DNA-binding motifs as potential functional determinants for POU domain factors. *Genes Dev.*, **7**, 2483–2496.
 Li, T., Stark, M.R., Johnson, A.D. and Wolberger, C. (1995) Crystal structure of the MATA1/MAT alpha 2 homeodomain heterodimer bound to DNA. *Science*, **270**, 262–269.
 Liang, J., Moye-Rowley, S. and Maurer, R.A. (1995) *In vivo* mutational analysis of the DNA binding domain of the tissue specific transcription factor, Pit-1. *J. Biol. Chem.*, **270**, 25520–25525.
 Luo, Y. and Roeder, R.G. (1995) Cloning, functional characterization, and mechanism of action of the B-cell-specific transcriptional coactivator OCA-B. *Mol. Cell. Biol.*, **15**, 4115–4124.
 Matsuo, K., Clay, O., Künzler, P., Georgiev, O., Urbánek, P. and Schaffner, W. (1994) Short intron interrupting the Oct-2 POU domain may prevent recombination between POU family genes without interfering with potential POU domain 'shuffling' in evolution. *Biol. Chem. Hoppe-Seyler*, **375**, 675–683.
 Ogata, K., Morikawa, S., Nakamura, H., Sekikawa, A., Inoue, T., Kanai, H., Sarai, A., Ishii, S. and Nishimura, Y. (1994) Solution structure of a specific DNA complex of the Myb DNA-binding domain with cooperative recognition helices. *Cell*, **79**, 639–648.

- O'Shannessy,D.J., Brigham-Burke,M., Soneson,K.K., Hensley,P. and Brooks,I. (1993) Determination of rate and equilibrium binding constants for macromolecular interactions using surface plasmon resonance: use of nonlinear least square analysis methods. *Anal. Biochem.*, **212**, 457–468.
- Pavletich,N.P. and Pabo,C.O. (1993) Crystal structure of a five-finger GLI–DNA complex: new perspectives on zinc fingers. *Science*, **261**, 1701–7.
- Pomerantz,J.L., Kristie,T.M. and Sharp,P.A. (1992) Recognition of the surface of a homeo domain protein. *Genes Dev.*, **6**, 2047–2057.
- Pomerantz,J.L., Sharp,P.A. and Pabo,C.O. (1995) Structure-based design of transcription factors. *Science*, **267**, 93–96.
- Schöler,H.R. (1991) Octamania: the POU factors in murine development. *Trends Genet.*, **7**, 323–329.
- Strubin,M., Newell,J.W. and Matthias,P. (1995) OBF-1, a novel B cell-specific coactivator that stimulates immunoglobulin promoter activity through association with octamer-binding proteins. *Cell*, **80**, 497–506.
- Studier,W.F., Rosenberg,A.H., Dunn,J.J. and Dubendorf,J.W. (1990) Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol.*, **185**, 60–89.
- Sturm,R.A., Cassady,J.L., Das,G., Romo,A. and Evans,G.A. (1993) Chromosomal structure and expression of the human OTF1 locus encoding the Oct-1 protein. *Genomics*, **16**, 333–341.
- van Leeuwen,H.C., Strating,M.J., Cox,M., Kaptein,R. and van der Vliet,P.C. (1995) Mutation of the Oct-1 POU-specific recognition helix leads to altered DNA binding and influences enhancement of adenovirus DNA replication. *Nucleic Acids Res.*, **23**, 3189–3197.
- Verrijzer,C.P., Kal,A.J. and Van der Vliet,P.C. (1990) The Oct-1 homeo domain contacts only part of the octamer sequence and full Oct-1 DNA-binding activity requires the POU-specific domain. *Genes Dev.*, **4**, 1964–1974.
- Verrijzer,C.P., Alkema,M.J., van Weeperen,W.W., van Leeuwen,H.C., Strating,M.J.J. and van der Vliet,P.C. (1992a) The DNA-binding specificity of the bipartite POU domain and its subdomains. *EMBO J.*, **11**, 4993–5002.
- Verrijzer,C.P., van Oosterhout,J.A.W.M., van Weperen,W.W. and van der Vliet,P.C. (1992b) POU proteins bend DNA via the POU-specific domain. *EMBO J.*, **10**, 3007–3014.
- Verrijzer,P.C., Strating,M.J.J., Mul,Y.M. and van der Vliet,P.C. (1992c) POU domain transcription factors from different subclasses stimulate adenovirus DNA replication. *Nucleic Acids Res.*, **20**, 6369–6375.
- Walker,S., Hayes,S. and O'Hare,P. (1994) Site-specific conformational alteration of the Oct-1 POU domain–DNA complex as the basis for differential recognition by Vmw65 (VP16). *Cell*, **79**, 841–852.

Received on August 14, 1996; revised on November 8, 1996