An Invariant Asparagine in the POU-Specific Homeodomain Regulates the Specificity of the Oct-2 POU Motif†

Martyn C. Botfield,‡,§ Agnes Jancso,‡,∥ and Michael A. Weiss*,‡,⊥,#

Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02115, and Department of Medicine, Massachusetts General Hospital, Boston, Massachusetts 02114

Received January 24, 1994; Revised Manuscript Received April 14, 1994®

ABSTRACT: The homeodomain defines a family of transcription factors broadly involved in the regulation of gene expression. DNA recognition, as observed in three representative complexes (Engrailed, Antennapedia, and MATα2), is mediated in the major groove by a helix–turn–helix (HTH) element and in the minor groove by an N-terminal arm. The three complexes share similar overall features, but they also exhibit significant differences in DNA interactions. Because these differences may distinguish the biological activities of different classes of homeodomains, we have investigated the contribution of the Oct-2 POU-specific homeodomain (POU_{HD}) to the specificity of the bipartite POU motif. Comparative studies of variant protein–DNA complexes demonstrate the following. (i) Mutations in an invariant residue in the POU_{HD} HTH (N347; residue 10 of the putative recognition α-helix) reduce octamer binding with the relaxation of specificity at one position (5′-ATGCAAAT). The inferred HTH side chain–base interaction, although not observed in the solution structure of the Antennapedia complex, is in accord with homologous contacts in the Engrailed and MATα2 cocrystal structures. (ii) Comparison of the DNA-binding properties of POU and POU_{HD} demonstrates that POU_S and POU_{HD} independently regulate specificity at opposite ends of the DNA site (5′-TATGCAAAT). Both domains contact the two central bases (5′-TATGCAAAT) where coordinate binding of POU_S in the major groove overrides the intrinsic specificity of POU_{HD} in the minor groove. (iii) The differential sensitivity of POU and POU_{HD} to 2′-deoxyinosine substitutions (a minor-groove modification) suggests that POU_S binding repositions the POU_{HD} N-terminal "arm". Such repositioning is likely to be mediated indirectly through the DNA structure. The existence of alternative POU_{HD} arm binding sites would be in accord with the differences between minor-groove binding sites observed in Engrailed and MATα2 cocrystal structures.

The helix–turn–helix (HTH[1]), first described in prokaryotic gene-regulatory proteins (Anderson et al., 1981; McKay & Steitz, 1981; Ohlendorf et al., 1982; Pabo & Lewis, 1982; Steitz et al., 1982; Sauer et al., 1982), provides a conserved motif of protein–DNA recognition (Pabo & Sauer, 1984, 1992). Eukaryotic HTH elements were first recognized in sequences of homeodomains (Laughon & Scott, 1984; Shepherd et al., 1984), which define a family of proteins broadly involved in the regulation of gene expression (Scott et al., 1989). The structures of three homeodomain–DNA complexes (Engrailed, Antennapedia, and MATα2) have been determined by X-ray crystallography (Kissinger et al., 1990; Wolberger et al., 1991) or NMR spectroscopy (Otting et al., 1990). These structures demonstrate contacts in the major groove by HTH and in the minor groove by an N-terminal arm. The orientation of the homeodomain HTH in the major groove differs from that of prokaryotic DNA-binding proteins (Anderson et al., 1985, 1987; Jordan & Pabo, 1988; Aggarwal et al., 1988; Wolberger et al., 1988; Mondragon & Harrison, 1991). A model of the Engrailed–DNA complex is shown in Figure 1A.

The bipartite POU motif defines a subgroup of homeodomain-containing transcription factors (Bodner et al., 1988; Ingraham et al., 1988; Sturm et al., 1988; Clerc et al., 1988; Ko et al., 1988; Schneidereit et al., 1988; Finney et al., 1988; Herr et al., 1988). The motif consists of an N-terminal POU-specific subdomain (POU_S) and a C-terminal variant homeodomain (POU_{HD}). The structure of POU_S is similar to those of the DNA-binding domains of bacteriophage repressor and Cro proteins (Assa-Munt et al., 1993; Dekker et al., 1993), whereas the structure of POU_{HD} is similar to those of canonical homeodomains in secondary (Morita et al., 1993) and tertiary structures (Sivaraja et al., 1994). The POU motif thus contains two distinct HTH elements, which are proposed to bind in successive DNA major grooves (Verrijzer et al., 1992; Assa-Munt et al., 1993; Dekker et al., 1993; Jancso et al., 1994; Botfield & Weiss, 1994). The present study focuses on the role of POU_{HD} in POU–DNA recognition. A model POU domain is obtained from human Oct-2, a B-cell-specific transcription factor also expressed in the developing central nervous system (Clerc et al., 1988; Ko et al., 1988; Müller-Immerglück et al., 1988; Scheidereit et al., 1988). Oct-2

† This work was supported in part by grants from the NIH (GM45290) and the Council for Tobacco Research to M.A.W. and by a Cancer Research Institute Postdoctoral Fellowship to M.C.B.

* Author to whom correspondence should be addressed.

‡ Harvard Medical School.

§ Present address: ARIAD Pharmaceuticals, Inc., Cambridge, MA 02139.

∥ Present address: Department of Biology, Brandeis University, Waltham, MA 02254.

⊥ Massachusetts General Hospital.

Present address: Center for Molecular Oncology and Departments of Biochemistry & Molecular Biology and Chemistry, The University of Chicago, Chicago, IL 60637-5419.

® Abstract published in *Advance ACS Abstracts,* June 1, 1994.

[1] Abbreviations: GRA, gel-retardation assay; GST, glutathione-*S*-transferase of *Schistosoma japonica* (amino acids 1–226); HTH, helix–turn–helix; NMR, nuclear magnetic resonance spectroscopy; PAGE, polyacrylamide gel electrophoresis; POU, POU-specific subdomain, intervening linker region, and homeodomain of Oct-2 (amino acids 194–359); POU_{HD}, homeodomain of Oct-2 (amino acids 295–359); POU_S, POU-specific subdomain of Oct-2 (amino acids 194–270); SDS, sodium dodecyl sulfate; U, 2′-deoxyuridine. Amino acids are designated by their standard single-letter codes; bp, base pair.

0006-2960/94/0433-8113$04.50/0

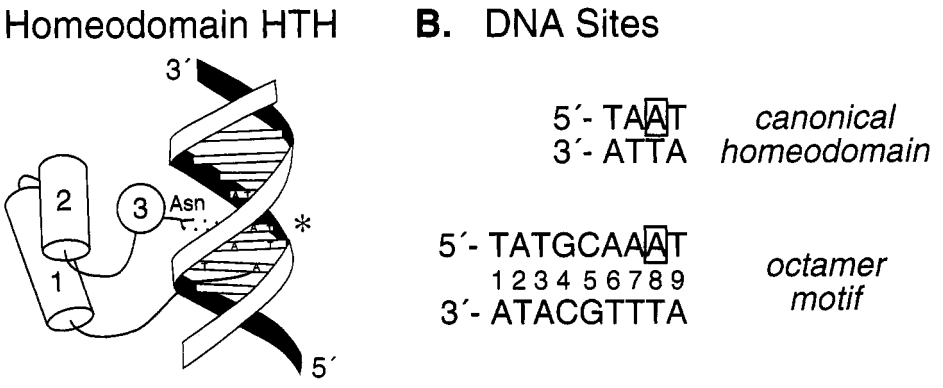

FIGURE 1: (A) Cylinder model of the Engrailed homeodomain (Kissinger et al., 1990). α-Helices 2 and 3 compose the HTH. The side chain of N51 is shown; in the cocrystal structure the carboxamide forms bidentate hydrogen bonds with an adenine (dotted lines; see also Figure 5). (B, upper panel) Core TAAT recognition sequence of a canonical homeodomain; the boxed adenine is contacted by N51 in the Engrailed cocrystal structure. (Lower panel) Octamer motif (base pairs 2–9). The proposed alignment is based on shared asparagine contact (boxed) as inferred from relaxed specificity mutant N347K (Figure 7; see part II of Results).

belongs to a subfamily of POU proteins, whose members (designated Oct-1, Oct-2, Oct-3, etc.) recognize an octanucleotide sequence (5′-ATGCAAAT-3′) conserved in vertebrate promoter and enhancer elements.

We have previously described the overexpression of Oct-2 POU and POU_{HD} fragments as thrombin-cleavable fusion proteins in *Escherichia coli*, their purification following thrombin digestion, and the initial characterization of their domain structures and DNA-binding properties (Botfield et al., 1992). Here we investigate the contribution of Oct-2 POU_{HD} to the DNA-binding properties of the intact POU motif. Alignment of POU_{HD} HTH against a consensus octamer site is established by study of a mutant POU domain with relaxed sequence specificity (Ebright, 1985, 1986; Hochschild & Ptashne, 1986). The mutation occurs at an invariant site in the putative POU_{HD} recognition α-helix (position 10); this site (N347 in Oct-2) is homologous to N51 in Engrailed (shown in Figure 1A). The inferred side chain–base interaction is shown to regulate the specificity of the POU domain at one position of the octamer (boxed in Figure 1B). Comparison of the sequence specificities of the POU domain and the POU_{HD} fragment at other positions suggests a functional interaction between domains: in the center of the octamer site, the intrinsic specificity of POU_{HD} in the minor groove is overridden by POU_S contacts in the major groove. Our results are discussed in relation to cocrystal and NMR-derived structures of homeodomain–DNA complexes (Kissinger et al., 1990; Otting et al., 1990; Wolberger et al., 1991) and the recent NMR structure of an Oct-1 POU_S fragment (Assa-Munt et al., 1993; Dekker et al., 1993).

MATERIALS AND METHODS

Synthetic Methods. Oligonucleotides containing standard nucleosides and the 2′-deoxyinosine nucleoside were purchased from Oligos, Etc. (Ridgefield, CT). Oligonucleotides containing N^6-methyl-2′-deoxyadenosine were kindly provided by Dr. L. C. Sowers (City of Hope Medical Center, Duarte, CA).

Site-Directed Mutagenesis. The Oct-2 POU domain (nucleotides 646–1143) was cloned into phage M13mp18 as described (Botfield et al., 1992). Oligonucleotide-directed mutagenesis was accomplished by the phosphorothioate method as described by the vendor (Amersham Corp., Inc., Arlington Heights, IL). Substitutions were identified by single-stranded–DNA sequencing; variant POU coding regions were recloned by PCR into pGEX-2T and resequenced to exclude PCR errors (Botfield et al., 1994).

Protein Expression and Purification. The native Oct-2 POU domain and isolated homeodomain (POU_{HD}) were expressed in pGEX-2T (Pharmacia, Inc.), purified, cleaved with thrombin, and fractionated as described (Botfield et al., 1992). Mutant POU domains were isolated as described for the wild type; purity was >90% as assessed by SDS–PAGE.

DNA-Binding Assays. Gel-retardation assays (GRA; Fried & Crothers, 1981) were performed at 4 °C under low-salt conditions as described (Botfield et al., 1992, 1994). Shifted and unshifted counts were quantified using a PhosphorImager (Molecular Dynamics, Inc., Sunnyvale, CA). All GRA experiments were repeated at least twice. The low affinity of the POU_{HD} fragment for 12-base-pair sites limits the quantitative determination of dissociation constants by GRA.

RESULTS

Overview of Experimental Design. A model of bipartite octamer recognition by the POU motif has been proposed by homology between POU_{HD} and canonical homeodomains (Verrijzer & Van der Vliet, 1993) and by analogy between POU_S and the DNA-binding domain of λ repressor (Assa-Munt et al., 1993; Dekker et al., 1993). This model is predicated on the following alignment of DNA recognition sites:

```
5'              T A A T T A        canonical
3'              A T T A A T        homeodomain

   0 1 2 3 4 5 6 7 8 9 10 11
5' G T A T G C A A A T G G         octamer
3' G A T A C G T T T A C C
   0 1 2 3 4 5 6 7 8 9 10 11

5'   T A T C A C G G C             λ operator
3'   A T A G T G C C G             half-site
```

Such alignment predicts that, in a POU–DNA complex, POU_{HD} regulates the specificity of the 3′ base pairs (5′-TATGCAAAT), whereas POU_S regulates the specificity of the 5′ base pairs (5′-TATGCAAAT); the central base pairs (5′-TATGCAAAT) presumably would be jointly recognized. The model further predicts that the POU_{HD} HTH affects DNA recognition in the 3′ major groove by contacts homologous to those of canonical homeodomains; because the POU_{HD} N-terminal arm is part of a linker between domains, its contacts in the central minor groove are not specified. This model is supported by the recent crystal structure of a specific complex between the Oct-1 POU domain and a related DNA site (Klemm et al., 1994; see Added in Proof).

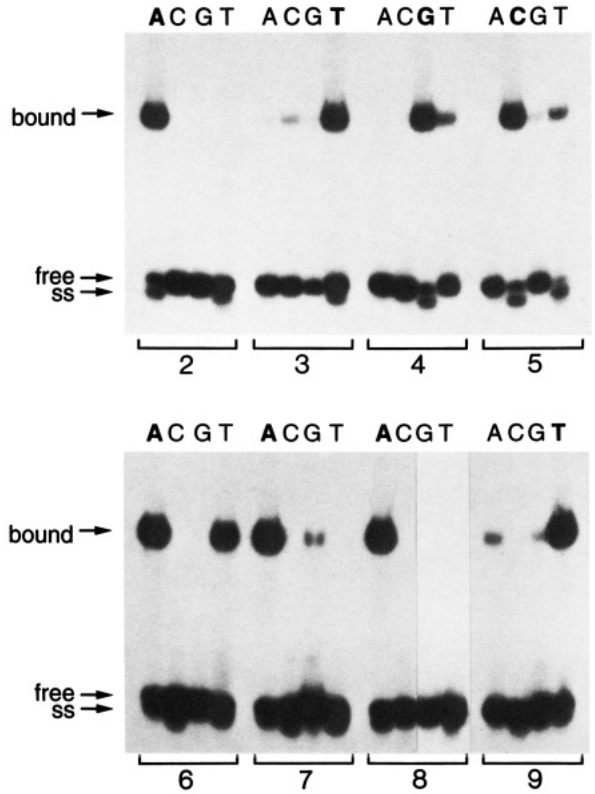

FIGURE 2: Sequence specificity of the Oct-2 POU domain at positions 2–9 (the classical octamer; 5′-GTATGCAAATGG). In each set of four, the preferred base (upper strand) is show in boldface type. The protein concentration was 50 nM. Sequence specificity at position 1 is weak (data not shown; see Figure 3A). The figure is a composite of two gels. The experiment was repeated in quadruplicate.

Here, we test these predictions by biochemical studies of variant POU–DNA complexes. Our results are presented in three parts. In part I, the contribution of POU_{HD} to sequence specificity is investigated by comparison of the DNA-binding properties of Oct-2 POU and POU_{HD} fragments; differences in specificity are ascribed to POU_S. In part II, the role of an invariant residue in the POU_{HD} HTH recognition α-helix (N347) is investigated by the analysis of a mutant domain with relaxed sequence specificity (N347K). In part III, DNA analogues are used as complementary probes of the protein–DNA interaction (McLaughlin et al., 1987).

(I) Sequence Specificity of POU and POU_{HD}. The following 12-bp DNA site is used as a baseline for analysis of specificity:

```
    0 1 2 3 4 5 6 7 8 9 10 11
5'  G T A T G C A A A T G  G
3'  G A T A C G T T T A C  C
    0 1 2 3 4 5 6 7 8 9 10 11
```

Base pairs 2–9 (in boldface type) represent a consensus octamer; base pair 1 is the first position identified by random binding-site selection (Verrijzer et al., 1992). GC base pairs at each end (positions 0, 10, and 11) were included in this model site to enhance DNA stability and protein affinity. In the presence of 0.1 μg/mL poly(dI–dC) as a nonspecific competitor, the Oct-2 POU domain binds to this oligonucleotide with an apparent K_d of 60 nM at 4 °C; the isolated homeodomain binds with an apparent K_d of >10 μM (Botfied et al., 1992).

Twenty-seven variant duplex oligonucleotides were prepared that contain all possible single-site transitions or transversions at positions 1–9. The POU-binding properties of the variant sites were tested individually by GRA (Figure 2). The native Oct-2 POU domain binds more strongly to the consensus site than to any of the 27 variants. However, the extent of discrimination varies significantly from position to position (Figure 3A). The wild-type base pair is uniquely specified at positions 2 and 8; at these sites any alternative reduces POU binding by at least 25-fold. At positions 1, 4, 5, 6, and 7, a single alternative is tolerated to within a factor of 5. Recognition is least specific at position 1. Sequence preferences at position 1 are similar to those of phage λ repressor (Sarai & Takeda, 1989), in accord with a predicted analogy (Assa-Munt et al., 1993; Dekker et al., 1993).

The variant 12-mer DNA sites were also used to characterize the specificity of Oct-2 POU_{HD} (Figures 3B and 4). The results differ from those obtained with the intact POU domain. Loss of sequence specificity is observed at 5′ positions (5′-G-TATGCAAATGG-3′), which presumably are contacted by POU_S and not by POU_{HD}. In addition, altered specificity is observed in the center of the binding site (positions 5 and 6; 5′-GTATGCAAATGG-3′). The isolated homeodomain binds more strongly to T at positions 5 and 6 than to the wild-type C and A (asterisks in Figure 3B). These differences occur at positions expected to be contacted in the minor groove by the POU_{HD} N-terminal arm (Kissinger et al., 1990; Otting et al., 1990; Wolberger et al., 1991). We speculate that differences in the specificity of POU and POU_{HD} at these positions reflect the effects of coordinate POU_S binding in the major groove (Assa-Munt et al., 1993; Dekker et al., 1993), as will be investigated further in part III using 2′-deoxyinosine analogues. The A→T transversion at position 6 (asterisks in Figures 3B and 4) recreates a consensus TAAT homeodomain target site (Figure 1B). Preferential binding of POU_{HD} to this site is consistent with the proposed alignment of the POU_{HD} subsite in the octamer.

(II) A Mutant Oct-2 POU Domain with Relaxed Sequence Specificity. The above studies establish that POU–DNA recognition is most stringent at position 8 (5′-GTATGCAA-ATGG-3′). Indeed, no specific binding is detectable following any sequence change (Figure 3A), indicating that the wild-type dissociation constant is perturbed by at least 10^3 in each case. Moreover, cocrystal structures of canonical homeodomains (Kissinger et al., 1990; Wolberger et al., 1991) predict that such recognition is effected by bidentate hydrogen bonds between adenine N^7 (donor) and 6-NH_2 (acceptor) and the side-chain carboxamide of an invariant asparagine in the HTH recognition α-helix. This interaction, shown in Figure 5, is specific for adenine (Seeman et al., 1976). The homologous contact is not observed, however, in the solution structure of the Antennapedia homeodomain–DNA complex, despite an analogous overall alignment of its HTH and TAAT binding site (Otting et al., 1990). The NMR evidence indicates that the carboxamide side chain is instead disordered adjacent to a solvent cavity in the protein–DNA interface (Billeter et al., 1993).

Because of the apparent disagreement between the predictions of crystal and NMR structures, we have focused on the role of the homologous asparagine (N347) in the Oct-2 POU_{HD} HTH. In a previous study, we showed that each of the 19 possible substitutions at Oct-2 position 347 reduces octamer recognition by at least 50-fold (Botfield et al., 1994), i.e., only the wild-type N347 confers high-affinity octamer binding. Here we extend this observation by screening the 19 mutant POU domains for altered or relaxed DNA specificity as follows. (i) At each octamer position, variant DNA sites with significantly reduced affinity for the native POU domain were pooled to define a set of eight "challenge sites". For example, the challenge site at position 8 contains the sequences 5′-

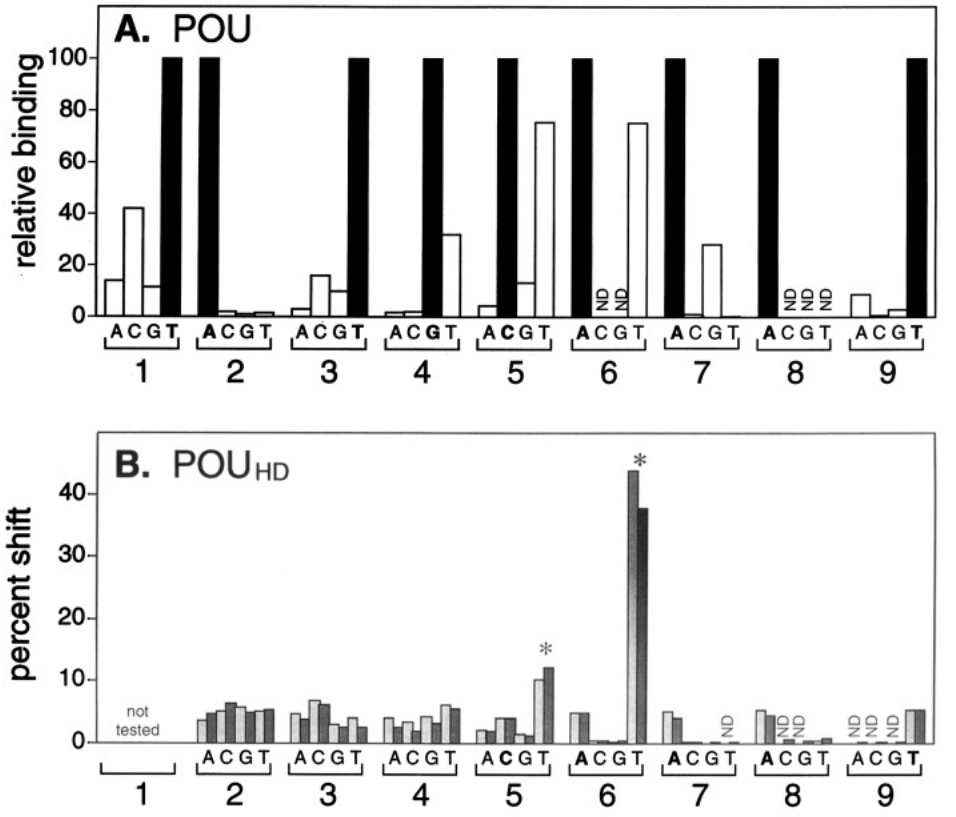

FIGURE 3: Quantitation of gel-retardation assays. (A) Effect of octamer base-pair changes on binding of the Oct-2 POU domain (see Figure 2). The vertical scale shows the relative binding as a % of the wild-type shift under conditions in which the native domain shifts approximately 50% of the native 12-mer site. (B) Effect of base-pair changes on binding of the POU_{HD} fragment (see Figure 4). The vertical scale shows the observed % shift in each case. Results are shown in duplicate. In each panel, the preferred base (upper strand) is show in boldface type. ND indicates that no complex was detectable.

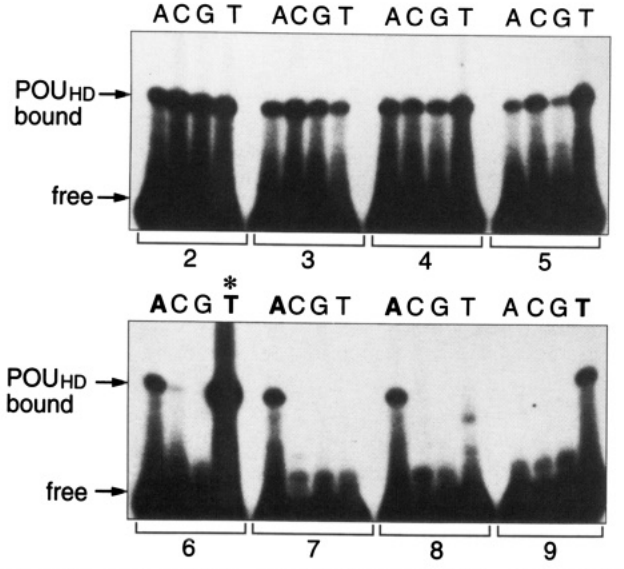

FIGURE 4: Sequence specificity of the Oct-2 POU_{HD} fragment at positions 2–9 (5′-GTATGCAAATGG). The asterisk indicates enhanced binding to non-native site 5′-GTATGCTAATGG, which recapitulates the consensus homeodomain target TAAT (Figure 1B). In each set of four, the preferred base (upper strand) is show in boldface type. The protein concentration was 1 μM.

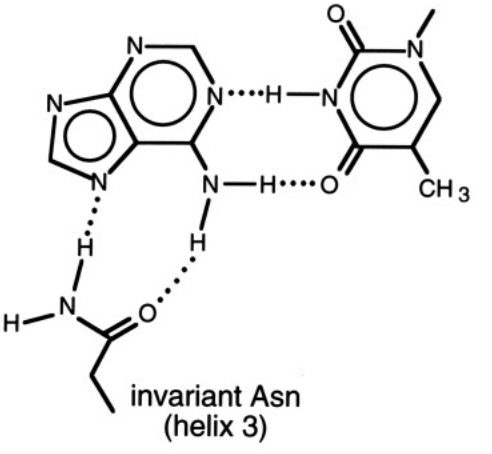

FIGURE 5: Bidentate hydrogen bonds between the N51 carboxamide in the Engrailed homeodomain and the adenine N^7 and 6-NH_2 functions in the major groove (Kissinger et al., 1990). This interaction is also observed in the MATα2 complex (Wolberger et al., 1991), but not in the NMR structure of the Antennapedia–DNA complex (Otting et al., 1990; Billeter et al., 1993).

GTATGCAACTGG, 5′-GTATGCAATTGG, and 5′-GTATGCAAGTGG. (ii) The residual affinities of the native POU domain for these challenge sites were determined by GRA. These affinities are as predicted; the individual results are summarized in Figure 3A. (iii) The affinities of the mutant POU domains for the challenge sites were screened by GRA. Since the mutant domains each exhibit >50-fold-reduced octamer binding, and since the challenge sites are also of low affinity, little or no detectable complex formation would be expected under the conditions used. Substitutions conferring altered or relaxed sequence specificity are thus identifiable by anomalous binding to one or more challenge sites, as illustrated in Figure 6.

Of the 19 substitutions, three were thus identified: N347K, N347S, and N347G. In each case, the site of relaxed or altered specificity was position 8 only. The N347K mutant POU domain was investigated in more detail. Of the three variant sites at position 8, only one (5′-GTATGCAAGTGG-3′)

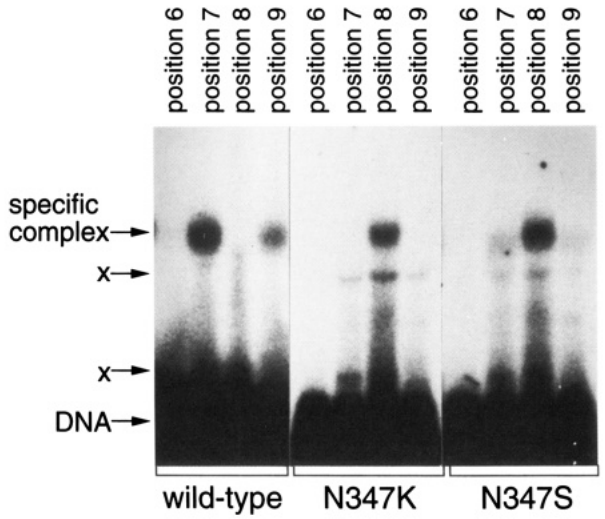

FIGURE 6: Screening for altered or relaxed specificity was accomplished by testing wild-type and mutant POU domains against a mixture of oligonucleotides, excluding high-affinity sequences (Figure 3). The screening of positions 6–9 is illustrated as follows: position 6, 5′-GGTATGCCAATGG and 5′-GGTATGCGAATGG; position 7, 5′-GGTATGCATATGG, 5′-GGTATGCACATGG, and 5′-GGTATGCAGATGG; position 8, 5′-GGTATGCAATTGG, 5′-GGTATGCAACTGG, and 5′-GGTATGCAAGTGG; and position 9, 5′-GGTATGCAAACGG, 5′-GGTATGCAAAAGG, and 5′-GGTATGCAAAGGG. An X indicates contaminating DNA-binding activity in the partially purified extract.

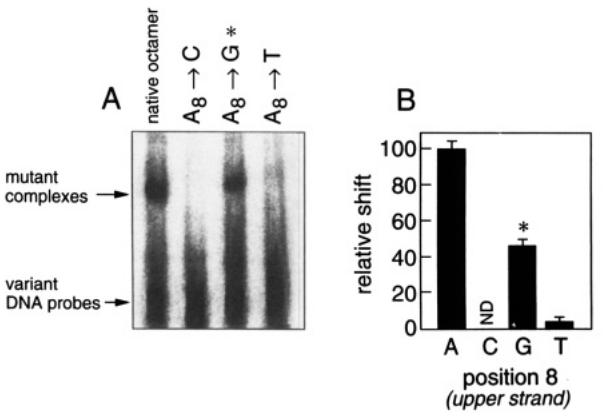

FIGURE 7: (A) GRA data testing the relative affinities of the N347K mutant POU domain for variant sites 5′-GTATGCAANTGG, where N is A, C, G, or T. (B) Quantitation of GRA using a PhosphorImager (Molecular Dynamics, Inc., Sunnyvale, CA). The asterisks in each panel indicate new (relaxed) specificity for guanine at position 8.

exhibits significant binding to the mutant protein (asterisk in Figure 7). The N347K substitution thus confers a new specificity for G at position 8. Because preferential binding to the wild-type site is maintained, specificity is relaxed rather than altered (Ebright, 1985, 1986; Hochschild & Ptashne, 1985). Variant DNA sites were also tested individually at positions 7 and 9 but revealed no changes in specificity (data not shown), verifying that the relaxation of specificity is selective for position 8. The affinity of the N347K domain for the native octamer site is approximately 100-fold weaker than that of the native POU domain, as estimated by GRA titration (Figure 8). The autoradiogram also exhibits a "tail" of complexes of intermediate mobilities, indicating the dissociation of unstable complexes during electrophoresis. Although the relaxed specificity of the N347S and N347G mutants may be rationalized as "loss of contact" substitutions (Ebright, 1991), the structural basis for the DNA-binding properties of the N347K analogue is not clear. Nevertheless,

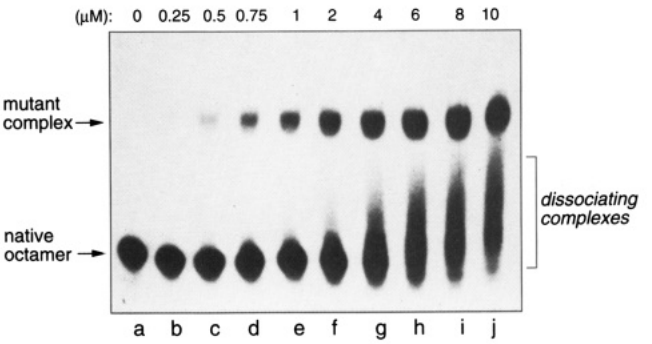

FIGURE 8: Binding of the HPLC-purified N347K mutant POU domain to the wild-type octamer site (5′-GTATGCAAATGG) at successive protein concentrations (indicated at top of figure). The estimated dissociation constant (6 μM) is 100-fold greater (weaker binding) than that of the native domain (60 nM) for the 12-mer site. Instability of the mutant complex leads to a broad tail of dissociating complexes (lanes h–j).

correlation between mutations at protein position 347 and selective relaxation of specificity at octamer position 8 demonstrate that the wild-type asparagine regulates DNA specificity at this site. Such regulation is consistent with (but does not prove) the formation of an adenine–carboxamide contact, as seen in the cocrystal structures of canonical homeodomains (Figure 5). This contact has recently been observed in an Oct-1 cocrystal structure (Klemm et al., 1994; see Added in Proof).

(III) DNA Analogues. The POU–octamer interaction was investigated further using DNA analogues N^6-methyladenosine (a major-groove modification; Figure 9B) and 2′-deoxyinosine nucleoside (a minor-groove modification; Figure 10C). The results of systematic adenosine modification are shown in Figure 9A. In accord with models of the POU–DNA complex, only two sites (boxed in Figure 9C) exhibit >100-fold interference with protein binding: positions 2 (in the POU_S-binding site) and 8 (in the POU_{HD}-binding site). Such interferences are in accord with the critical roles of these two 6-NH_2 groups as predicted by model-building (Verrijzer & Van der Vliet, 1993; Assa-Munt et al., 1993; Dekker et al., 1993). Elsewhere, we have shown that the complementary thymidines ($T_{2'}$ and $T_{8'}$) are unlikely to be contacted, as uridine substitutions are well tolerated at these sites (Jancso et al., 1994). The results of DNA modification thus further support the hypothesis that POU_{HD} specificity at position 2 is effected by an adenine-specific contact. At positions 6 and 7, substitution of N^6-methyladenosine reduces POU binding by approximately 50% (dashed circles in Figure 9C). Because sites of partial interference may reflect nonlocal changes in DNA structure or transmitted perturbations in protein–DNA contacts, structural interpretation is not possible.

Inosine and guanine are distinguished by the absence (I) or presence (G) of a 2-NH_2 group (Figure 10C). Since POU_S and POU_{HD} HTH motifs are expected to bind in the major groove, 2′-deoxyinosine nucleoside substitutions are of particular interest as probes of minor-groove recognition by the POU_{HD} N-terminal arm. This arm is in fact part of a flexible linker between POU_S and POU_{HD} and would be expected to bind near the center of the octamer site (5′-TATGCAAAT-3′). The role of the N-terminal arm is of particular interest in light of the differential specificity of POU and POU_{HD} at positions 5 and 6 (part I; see above). Moreover, structural variation in arm binding sites is observed between Engrailed and MATα2 cocrystal structures (Kissinger et al.,

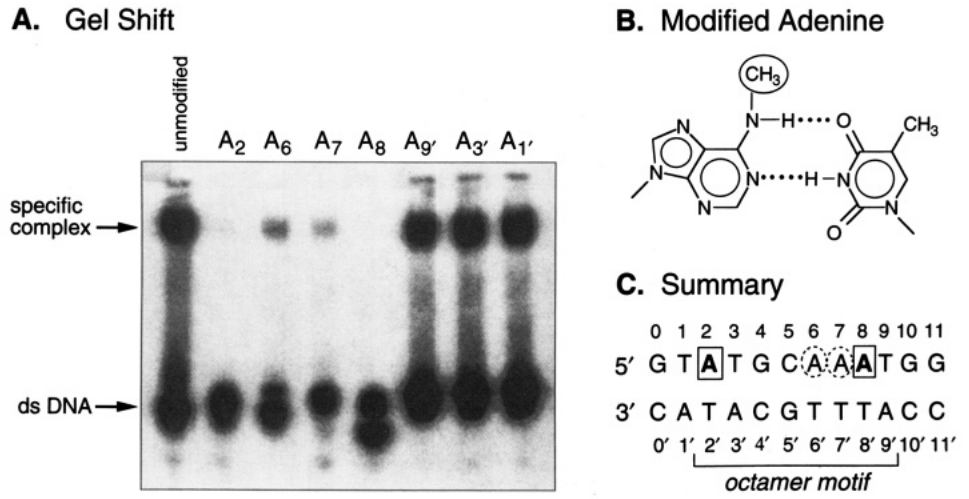

FIGURE 9: DNA-binding study of N^6-methyladenosine analogues. (A) GRA demonstrates the absence of detectable complex formation when A_8 is modified; almost complete interference is observed at position 2 in the POU_S-binding site (boxed in panel C). Positions 6 and 7 (circled in panel C) exhibit partial interference in accord with previous footprinting studies (Verrijzer et al., 1990). The protein concentration was 60 nM. (B) Structure of the modified base pair; the N^6-methyl group is circled. (C) Summary of interference results and the base numbering scheme.

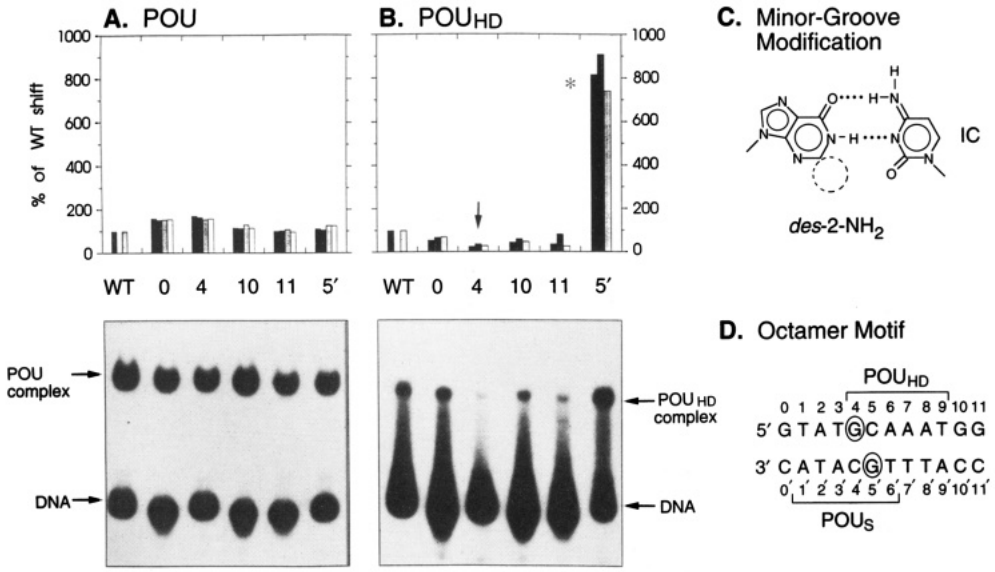

FIGURE 10: DNA-binding studies of 2′-dexoyinosine nucleoside analogues: (A) binding of the POU domain and (B) binding of POU_{HD}. The lower panels show GRA data; upper panels summarize the results of three separate experiments. WT indicates wild-type octamer DNA; numbers designate positions of inosine modification (see panel D). The arrow and asterisk in panel B (upper panel) indicate positions of differential sensitivity between POU and POU_{HD} (positions 4 and 5′). (C) Structure of the inosine/cytidine (IC) base pair, which is distinguished from GC by the absence of the 2-NH_2 group in the minor groove (dashed circle). (D) Summary of the numbering scheme and positions of overlapping POU_{HD}- and POU_S-binding sites in the octamer motif. Positions of differential sensitivity to inosine are circled (G_4 and $G_{5'}$).

1990; Wolberger et al., 1991). Such variation alters the pattern of side chain–DNA contacts, demonstrating that this flexible region can recognize DNA in different ways.

The effects of 2′-deoxyinosine nucleoside substitution were determined separately for POU and POU_{HD} (Figure 10A,B, respectively). No significant effects of the 2′-deoxyinosine nucleoside on POU binding are observed. In contrast, inosine at position 5′ (lower strand; Figure 10D) enhances POU_{HD} binding (asterisk in Figure 10B). This enhancement is in accord with preferential binding of the isolated POU_{HD} fragment to the variant 5′-TATGTAAAT-3′ site (part I; asterisk in Figure 3), as the complementary adenine at position 5′ would (like inosine) lack the 2-NH_2 group. Inosine at position 4 (upper strand) also reduces POU_{HD} binding somewhat (arrow in Figure 10B). The differential effects of inosine on POU and POU_{HD} further suggest an interaction between the binding of POU_S in the major groove and the binding of the POU_{HD} N-terminal arm in the minor groove. Because the intervening linker is flexible (Botfiel et al., 1992) and not well conserved (Verrijzer & Van der Vliet, 1993), this interaction is more likely to reflect an indirect effect of DNA structure than a direct POU_S–arm interaction (see Added in Proof).

DISCUSSION

The POU motif defines a class of transcription factors of central importance in metazoan development [for review, see Verrijzer and Van der Vliet (1993)]. Bipartite DNA recognition is accomplished by distinct HTH scaffolds, POU_S and POU_{HD}, which differ in structure and presumed mode of DNA binding (Assa-Munt et al., 1993; Dekker et al., 1993). Whereas the structure of POU_{HD} is similar to those of canonical homeodomains (Morita et al., 1993; Sivaraja et al., 1994), the structure of POU_S is similar to that of the DNA-binding domain of phage λ repressor (Assa-Munt et al., 1993; Dekker et al., 1993). In the present study, we have investigated the

contribution of the Oct-2 POU_{HD} subdomain to the affinity and specificity of the intact POU motif. Evidence that the mechanism of POU_S binding is analogous to that of λ repressor is presented elsewhere (Botfield & Weiss, 1994; Jancso et al., 1994). A crystal structure of a homologous Oct-1–DNA complex has recently been described (Klemm et al., 1994; see Added in Proof).

In this paper, the baseline specificity of the Oct-2 POU domain has been investigated by a systematic study of variant DNA sites. The parent oligonucleotide is a 12-mer duplex 5′-GTATGCAAATGG-3′, which contains a consensus octamer site (underlined). Each possible transition and transversion have been synthesized and tested individually. Because the parent site is identical to the core recognition element, identified by random binding-site selection using an Oct-1 POU fragment (Verrijzer et al., 1992), the variant DNA sites explore sequence space in the immediate neighborhood of the global optimum. A limitation of such a local search is the possible exclusion of sites with multiple sequence changes of compensatory effect. Nevertheless, the results presented in part I indicate that the octamer motif consists of an exclusive POU_S-binding subsite (5′-GTATGCAAATGG-3′), an exclusive POU_{HD}-binding subsite (5′-GTATGCAAATGG-3′), and a region of central overlap (5′-GTATGCAAATGG-3′). The region of overlap presumably corresponds to coordinate binding of POU_S in the major groove and the POU_{HD} N-terminal arm in the minor groove; corecognition may also be effected by POU_S and POU_{HD} contacts to opposing phosphates of the same base pairs (Botfield & Weiss, 1994). Within these three subsites, the stringency of specificity differs from position to position. Recognition is most stringent at positions 2 and 8, which are predicted to be contacted by analogous adenine–carboxamide interactions (Figure 5) from invariant residues in POU_S and POU_{HD}, respectively. These adenines are distinguished as sites of interference following N^6-methylation (Figure 9). The predicted asparagine–adenine contact at position 8 is consistent with the observation that mutations in this invariant site in the POU_{HD} recognition α-helix are associated with the selective relaxation of specificity.

An alignment of prokaryotic and eukaryotic HTH binding sites is provided in Figure 11. The upper two sequences are the binding sites for the Engrailed and MATα2 homeodomains used in cocrystallization (Kissinger et al., 1990; Wolberger et al., 1991). In each case, major-groove contacts by the HTH recognition α-helix are boxed, and minor-groove contacts by the N-terminal arm are circled. Despite the similar overall orientation of the two protein–DNA complexes, differences are seen in the pattern of base-specific contacts, especially in the extent of the arm binding site. The central panel shows the octamer sequence and possible POU_{HD} contacts inferred from biochemical footprinting methods (Kristie & Sharp, 1990;Verrijzer et al., 1990; Verrijzer & van der Vliet, 1993) and studies of DNA analogues (Jancso et al., 1994). Positions 10 and 11, not tested by these methods, may also contribute to POU_{HD} binding, as suggested by sequence preferences at these positions detected by random binding-site selection (Verrijzer et al., 1992). The extension of POU_{HD} contacts outside the classical octamer (dashed line in central panel of Figure 11) would be in accord with the Engrailed and MATα2 cocrystal structures and with observations that 12-mer, 14-mer, and 18-mer sites exhibit progressively higher affinities for the Oct-2 POU domain (A. Jancso and M. A. Weiss, unpublished results). These downstream base pairs may also contribute additional phosphate contacts (Botfield & Weiss, 1994). The lower two sequences in Figure 11 compare base-specific contacts by the operator-binding domain of the λ repressor (Jordan & Pabo, 1988) with POU_S contacts inferred from studies of DNA analogues (Jancso et al., 1994). The consistency of these results supports the hypothesis that POU_S and phage HTH domains bind DNA by a common mechanism (Assa-Munt et al., 1993; Dekker et al., 1993).

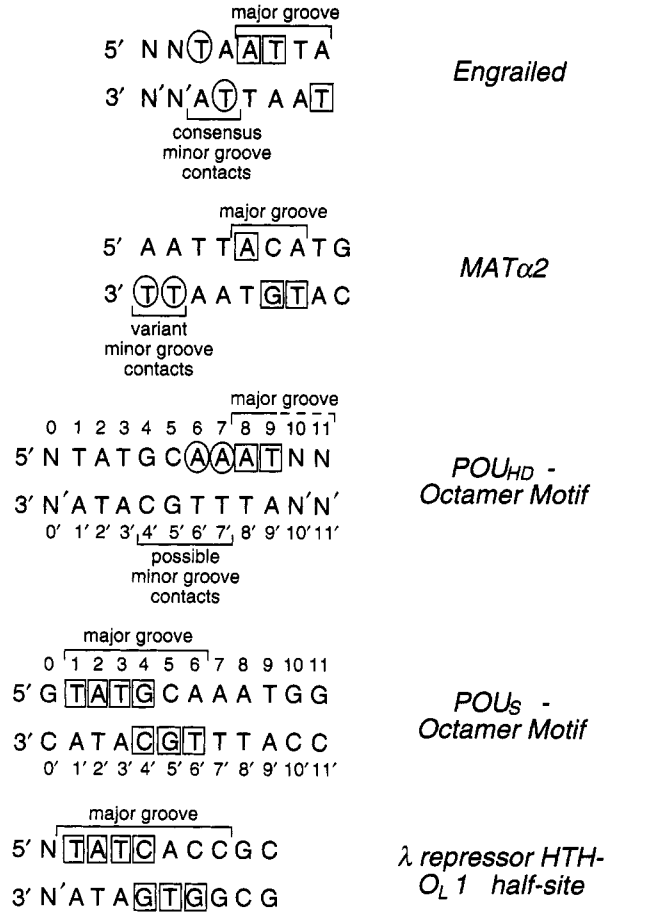

FIGURE 11: Proposed alignment of DNA-binding sites for canonical homeodomains (upper panels; Engrailed and MATα2), POU_{HD} and POU_S (central panels), and bacteriophage λ repressor (bottom panel). Major-groove contacts are boxed, and minor-groove contacts are circled. Engrailed, MATα2, and λ repressor contacts are obtained from the respective cocrystal structures (Kissinger et al., 1990; Wolberger et al., 1991; Jordan & Pabo, 1988). POU_{HD} and POU_S contacts are inferred from footprinting and studies of DNA analogues, as described in this and previous studies (LeBowitz et al., 1988; Kristie & Sharp, 1990; Verrijzer et al., 1990, 1992; Jancso et al., 1994). The dashed line in the POU_{HD} panel leaves open the possibility of additional major-groove contacts in accord with the Engrailed structure and random binding-site selection (Verrijzer et al., 1992).

In summary, the POU motif contacts two distinct HTH motifs, POU_S and POU_{HD}, which each contribute base-specific contacts and are jointly required for DNA recognition (LeBowitz et al., 1988; Kristie & Sharp, 1990; Verrijzer et al., 1990, 1992). In the present study, the contribution of POU_{HD} to the DNA-binding properties of the intact POU motif is investigated by analysis of sequence specificity, site-directed mutagenesis, and DNA modification. Our results strongly suggest that in the POU–DNA complex the orientation of the POU_{HD} HTH is similar to those of canonical homeodomains. In particular, an invariant asparagine in the recognition α-helix (position 10) appears to make an analogous side chain-base contact. By contrast, the POU_{HD} N-terminal arm, which is part of a flexible linker between POU_{HD} and POU_S, is likely to be repositioned in a POU complex from its

intrinsic specificity in an isolated homeodomain complex. The POU–octamer complex provides an intriguing example of overlapping DNA target sites. Coordinate recognition of its central base pairs in the major (POU_S) and minor (POU_{HD}) DNA grooves provides a model for the higher-order assembly of multiple transcription factors at eukaryotic promoters.

ADDED IN PROOF

The crystal structure of a specific complex between a homologous Oct-1 POU fragment and related DNA site (5′-GTATGCAAATAAGG) has recently been determined by Pabo and colleagues (Klemm et al., 1994). The structure demonstrates the predicted analogy between POU_S and phage λ repressor (Assa-Munt et al., 1993; Dekker et al., 1993) and verifies that POU_{HD} binds the CAAAT subsite by a mechanism similar to that of Engrailed (Kissinger et al., 1990). The linker region is not visualized, which is consistent with its flexibility in solution (Botfield et al., 1992). The conclusions of the present study are in accord with the cocrystal structure. The mechanism by which coordinate POU_S binding overrides the intrinsic specificity of POU_{HD} in the minor groove remains unclear, however, but may involve overlapping phosphate contacts near the center of the site (Klemm et al., 1994). Such coupling reflects broader issues of cooperativity in protein–DNA recognition.

ACKNOWLEDGMENT

We thank Prof. P. A. Sharp and J. LeBowitz (MIT, Cambridge, MA) for Oct-2 cDNA and advice; Dr. L. C. Sowers (City of Hope Medical Canter, Duarte, CA) for oligonucleotide synthesis; J. Englert for technical assistance; and A. Hochschild, C. Richardson, H. Singh, K. Struhl, and S. Tabor for helpful discussions.

REFERENCES

Aggarwal, A. K., Rogers, D. W., Drottar, M., Ptashne, M., & Harrison, S. C. (1988) *Science 242*, 899–907.

Anderson, W. F., Ohlendorf, D. H., Takeda, Y., & Matthews, B. W. (1981) *Nature 298*, 754–758.

Anderson, J. E., Ptashne, M., & Harrison, S. C. (1985) *Nature 316*, 596–601.

Anderson, J. E., Ptashne, M., & Harrison, S. C. (1987) *Nature 326*, 846–852.

Assa-Munt, N., Mortishire-Smith, R., Aurora, R., Herr, W., & Wright, P. E. (1993) *Cell 73*, 193–205.

Billeter, M., Qian, Y. Q., Otting, G. Muller, M., Gehring, W., & Wuthrich, K. (1993) *J. Mol. Biol. 234*, 1084–1093.

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

Botfield, M. C., & Weiss, M. A. (1994) *Biochemistry 33*, 2349–2355.

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

Botfield, M. C., Jancso, A., & Weiss, M. A. (1994) *Biochemistry 33*, 6177–6185.

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

Dekker, N., Cox, M., Boelens, R., Verrijer, C. P., van der Vliet, P. C., & Kaptein, R. (1993) *Nature 362*, 852–855.

Ebright, R. H. (1985) *J. Biomol. Struct. Dyn. 3*, 281–297.

Ebright, R. H. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 303–307.

Ebright, R. H. (1991) *Methods Enzymol. 208*, 620–640.

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

Fried, M., & Crothers, D. M. (1981) *Nucleic Acids Res. 9*, 6505–6525.

Harrison, S. C., & Aggarwal, A. K. (1990) *Annu. Rev. Biochem. 59*, 533–569.

Herr, W., Sturm, R. A., Clerc, R. G., Corcoran, L. M., Baltimore, D., Sharp, P. A., Ingraham, H. A., Rosenfeld, M. G., Finney, M., Ruvkun, G., & Horvitz, H. R. (1988) *Genes Dev. 2*, 1513–1516.

Hochschild, A., & Ptashne, M. (1986) *Cell 44*, 925–933.

Ingraham, H. A., Chen, R., Mangalam, H. J., Elsholtz, H. P., Flynn, S. E., Lin, C. R., Simmons, D. M., Swanson, L., & Rosenfeld, M. G. (1988) *Cell 55*, 519–529.

Jancso, A., Botfield, M. C., & Weiss, M. A. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 3887–3891.

Jordan, S. R., & Pabo, C. O. (1988) *Science 242*, 893–899.

Kissinger, C. R., Liu, B., Martin-Blanco, E., Kornberg, T. B., & Pabo, C. O. (1990) *Cell 63*, 579–590.

Klemm, J. D., Rould, M. A., Aurora, R., Herr, W., & Pabo, C. O. (1994) *Cell 77*, 21–32.

Ko, H.-S., Fast, P., McBride, W., & Staudt, L. M. (1988) *Cell 55*, 135–144.

Kristie, T. M., & Sharp, P. A. (1990) *Genes Dev. 4*, 2383–2396.

Laughon, A., & Scott, M. P. (1984) *Nature 310*, 25–31.

LeBowitz, J. H., Kobayashi, T., Staudt, L., Baltimore, D., & Sharp, P. A. (1988) *Genes Dev. 2*, 1227–1237.

McKay, D. B., & Steiz, T. A. (1981) *Nature 290*, 744–749.

McLaughlin, L. W., Benseler, F., Graeser, E., Piel, N., & Scholtissek, S. (1987) *Biochemistry 26*, 7238–7245.

Mondragon, A., & Harrison, S. C. (1991) *J. Mol. Biol. 219*, 321–334.

Morita, E. H., Shirakawa, M., Hayashi, F., Imagawa, M., & Kyogoku, Y. (1993) *FEBS Lett. 321*, 107–110.

Müller-Immerglück, M. M., Schaffner, W., & Matthias, P. (1990) *EMBO J. 9*, 1625–1634.

Ohlendorf, D. H., Anderson, W. F., Fisher, R. G., Takeda, Y., & Matthews, B. W. (1982) *Nature 298*, 718–723.

Otting, G., Qian, Y.-Q., Billeter, M., Müller, M., Affolter, M., Gehring, W., & Wüthrich, K. (1990) *EMBO J. 9*, 3085–3092.

Pabo, C. O., & Lewis, M. (1982) *Nature 298*, 443–447.

Pabo, C. O., & Sauer, R. T. (1984) *Annu. Rev. Biochem. 53*, 293–321.

Pabo, C. O., & Sauer, R. T. (1992) *Annu. Rev. Biochem. 61*, 1053–1095.

Pabo, C. O., Sauer, R. T., Sturtevant, J. M., & Ptashne, M. (1979) *Proc. Natl. Acad. Sci. U.S.A. 59*, 1259–1263

Pabo, C. O., Aggarwal, A. K., Jordon, S. R., Beamer, L. J., Obeysekare, U. R., & Harrison, S. C. (1990) *Science 247*, 1210–1213.

Pfaffle, R. W., DiMattia, G. E., Parks, J. S., Brown, M. R., Wit, J. M., Jansen, M., Van der Nat, H., Van den Brande, J. L., Rosenfeld, M. G., & Ingraham, H. A. (1992) *Science 257*, 1210–1213.

Sarai, A., & Takeda, Y. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 6513–6517.

Sauer, R. T., Yocum, R. R., Doolittle, R. F., Lewis, M., & Pabo, C. O. (1982) *Nature 298*, 447–451.

Scheidereit, C., Cromlish, J. A., Gerster, T., Kawakami, K., Balmaceda, C.-G., Currie, R. A., & Roeder, R. G. (1988) *Nature 336*, 551–557.

Schultz, S. C., Shields, G. C., & Steitz, T. A. (1991) *Science 253*, 1001–1007.

Scott, M. P., Tamkun, J., & Hartzell, G. W. (1989) *Biochim. Biophys. Acta 989*, 25–48.

Seeman, N. C., Rosenberg, J. M., & Rich, A. (1976) *Proc. Natl. Acad. Sci. U.S.A. 73*, 804–808.

Shepherd, J. C. W., McGinnis, W., Carrasco, A. E., DeRobertis, E. M., & Gerhing, W. J. (1984) *Nature 310*, 70–71.

Siebenlist, U., & Gilbert, W. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 122–126.

Singh, H., Sen, R., Baltimore, D., & Sharp, P. A. (1986) *Nature 319*, 154–158.

Sivaraja, M., Botfield, M. C., Mueller, M., Jancso, A., & Weiss, M. A. (1994) *Biochemistry* (in press).

Steitz, T. A., Ohlendorf, D. H., McKay, D. B., Anderson, W. F., & Matthews, B. W. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 3097–3100.

Sturm, R. A., & Herr, W. (1988) *Nature 336*, 601–604.

Verrijzer, C. P., & Van der Vliet, P. C. (1993) *Biochim. Biophys. Acta 1173*, 1–21.

Verrijzer, C. P., Kal, A. J., & van der Vliet, P. C. (1990) *Genes Dev. 4*, 1964–1974.

Verrijzer, C. P., Alkema, M. J., Van Weperen, W. W., Van Leeuwen, H. C., Strating, M. J. J., & Van der Vliet, P. C. (1992) *EMBO J . 11*, 4993–5003.

Wolberger, C., Dong, Y. C., Ptashne, M., & Harrison, S. C. (1988) *Nature 335*, 789–795.

Wolberger, C., Vershon, A. K., Liu, B., Johnson, A. D., & Pabo, C. O. (1991) *Cell 67*, 517–528.